INTELLI 2019

The Eighth International Conference on Intelligent Systems and Applications

ISBN: 978-1-61208-723-8

June 30 – July 4, 2019

Rome, Italy

INTELLI 2019 Editors

Gil Gonçalves, University of Porto, Portugal

Leo van Moergestel, HU Utrecht University of Applied Sciences, the Netherlands

INTELLI 2019

Foreword

The Eighth International Conference on Intelligent Systems and Applications (INTELLI 2019), held between June 30 – July 4, 2019 - Rome, Italy, was an inaugural event on advances towards fundamental, as well as practical and experimental aspects of intelligent and applications.

The information surrounding us is not only overwhelming but also subject to limitations of systems and applications, including specialized devices. The diversity of systems and the spectrum of situations make it almost impossible for an end-user to handle the complexity of the challenges. Embedding intelligence in systems and applications seems to be a reasonable way to move some complex tasks form user duty. However, this approach requires fundamental changes in designing the systems and applications, in designing their interfaces and requires using specific cognitive and collaborative mechanisms. Intelligence became a key paradigm and its specific use takes various forms according to the technology or the domain a system or an application belongs to.

We take here the opportunity to warmly thank all the members of the INTELLI 2019 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to INTELLI 2019. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the INTELLI 2019 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that INTELLI 2019 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of intelligent systems and applications.

We are convinced that the participants found the event useful and communications very open. We also hope that Rome provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

INTELLI 2019 Chairs

INTELLI Steering Committee

Lars Braubach, Hochschule Bremen, Germany
Leo van Moergestel, HU University of Applied Sciences Utrecht, Netherlands
Maiga Chang, Athabasca University, Canada
Sérgio Gorender, Federal University of Bahia (UFBA) & Federal University of the South of Bahia

INTELLI 2019

Committee

INTELLI Steering Committee

Lars Braubach, Hochschule Bremen, Germany
Leo van Moergestel, HU University of Applied Sciences Utrecht, Netherlands
Maiga Chang, Athabasca University, Canada
Sérgio Gorender, Federal University of Bahia (UFBA) & Federal University of the South of Bahia (UFSB), Brazil
Chin-Chen Chang, Feng Chia University, Taiwan
Stefano Berretti, University of Florence, Italy
Antonio Martin, Universidad de Sevilla, Spain

INTELLI Industry/Research Advisory Committee

David Greenhalgh, University of Strathclyde, Glasgow, UK
Carsten Behn, Technische Universität Ilmenau, Germany
Paolo Spagnolo, National Research Council, Italy
Luca Santinelli, ONERA Toulouse, France
Floriana Gargiulo, Gemass - CNRS | University of Paris Sorbonne, Paris, France

INTELLI 2019 Technical Program Committee

Azizi Ab Aziz, Universiti Utara Malaysia, Malaysia
Jair Minoro Abe, Paulista University / Institute of Advanced Studies – University of São Paulo, Brazil
Witold Abramowicz, Poznan University of Economics and Business, Poland
Giovanni Acampora, University of Naples Federico II, Italy
Ari Aharari, SOJO University, Japan
Bilal Ahmad, University of Warwick, UK
Zaher Al Aghbari, University of Sharjah, UAE
Gábor Alberti, University of Pécs, Hungary
Raul Alcaraz Martinez, University of Castilla-La Mancha, Spain
Ana Almeida, GECAD-ISEP-PPorto, Portugal
Rachid Anane, Coventry University, UK
Sílvia Araújo, University of Minho, Portugal
Davide Bacciu, Università di Pisa, Italy
Suzanne Barber, The University of Texas at Austin, USA
Mohammadamin Barekatain, Technical University of Munich, Germany
Kamel Barkaoui, Cedric-Cnam, France
Senén Barro Ameneiro, University of Santiago de Compostela, Spain
Ana Isabel Barros, TNO, Netherlands
Carmelo J. A. Bastos-Filho, University of Pernambuco, Brazil

Emmanuelle Grislin, LAMIH-CNRS UMR 8201 | Université de Valenciennes, France
Jerzy Grzymala-Busse, University of Kansas, USA
Sung Ho Ha, Kyungpook National University, South Korea
Maki K. Habib, The American University in Cairo, Egypt
Ibrahim A. Hameed, NTNU, Norway
Jessica Heesen, University of Tübingen, Germany
Wladyslaw Homenda, Warsaw University of Technology, Poland
Katsuhiro Honda, Osaka Prefecture University, Japan
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Wei-Chiang Hong, School of Education Intelligent Technology - Jiangsu Normal University, China
Martin Horauer, University of Applied Sciences - Technikum Wien, Austria
Christopher-Eyk Hrabia, Technische Universität Berlin | DAI-Labor, Germany
Matin Hosseini, University of Louisiana at Lafayette, USA
Michael Hübner, Ruhr-Universität Bochum, Germany
Chih-Cheng Hung, Kennesaw State University - Marietta Campus, USA
Sardar Jaf, University of Durham, UK
Mahdi Jalili, RMIT University, Melbourne, Australia
Juergen Jasperneite, Fraunhofer IOSB-INA, Germany
Richard Jiang, Northumbria University, UK
Maria João Ferreira, Universidade Portucalense, Portugal
Janusz Kacprzyk, Systems Research Institute - Polish Academy of Sciences, Poland
Epaminondas Kapetanios, University of Westminster, London, UK
Nikos Karacapilidis, University of Patras, Greece
Mehdi Kargar, Ryerson University, Canada
Fakhri Karray, University of Waterloo, Canada
Alexey M. Kashevnik, SPIIRAS, Russia
Fouad Khelifi, Northumbria University at Newcastle, UK
Shubhalaxmi Kher, Arkansas State University, USA
Hyunju Kim, Wheaton College, USA
Christina Kluever, University of Duisburg-Essen, Germany
Stefanos Kollias, University of Lincoln, UK
Ah-Lian Kor, Leeds Beckett University, UK
Sotiris Kotsiantis, University of Patras, Greece
Tobias Küster, DAI-Labor/Technische Universität Berlin, Germany
Ruggero Donida Labati, Universita' degli Studi di Milano, Italy
Asma Ladj, LS2N laboratory, Nantes, France
Ramoni Lasisi, Virginia Military Institute, USA
María Elena Lárraga Ramírez, Instituto de Ingeniería | Universidad Nacional Autónoma de México, Mexico
Antonio LaTorre, Universidad Politécnica de Madrid, Spain
Egons Lavendelis, Riga Technical University, Latvia
Frédéric Le Mouël, Univ. Lyon / INSA Lyon, France
George Lekeas, City Universty - London, UK
Carlos Leon de Mora, University of Seville, Spain
Chanjuan Liu, Dalian University of Technology, China
Niels Lohse, Loughborough University, UK
Volker Lohweg, Ostwestfalen-Lippe University of Applied Sciences, Germany
Daniela López De Luise, CI2S Labs, Argentina

Table of Contents

Implementation of a Self-Enforcing Network to Identify Determinants of the WiFi Quality on German Highspeed Trains

Jonathan Berrisch
Institute for Business and Economic Studies
University of Duisburg-Essen
Essen, Germany
email: jonathan.berrisch@vwl.uni-due.de

Timo Rammert
Institute for Business and Economic Studies
University of Duisburg-Essen
Essen, Germany
email: timo.rammert@vwl.uni-due.de

Christina Klüver
Institute for Computer Science and Business Information
University of Duisburg-Essen
Essen, Germany
email: christina.kluever@uni-due.de

***Abstract*—In this paper, we demonstrate how to analyze the WiFi data of the German highspeed trains called InterCityExpress (ICE) on the basis of a neural network. To achieve this, we apply a Self-Enforcing Network with cue validity factors to underline the importance of selected features. It is shown that the quality of the WiFi connection, in terms of the rate of downloads and the latency, can be grouped and explained by just a few determinants. We will show where the network coverage is especially good or bad and suggest ways to improve this quality to enhance the comfort of traveling on the highspeed trains and therefore to possible expand the profits of the operating company.**

***Keywords*–Self-Enforcing Network (SEN); self-organized learning; cue validity factor; Intelligent data analysis; Industry 4.0 data analysis.**

I. INTRODUCTION

In recent years, the demand for mobile Internet access grew at a rapid pace. The reasons for this development are diverse. There is ongoing research on how this change affects people in various ways (e.g., [1][2]). Furthermore, there is a growing literature trying to identify the factors which drive the ongoing rise in demand [3][4]. While user habits change, new challenges, and opportunities for many businesses arise.

Technical challenges to manage wireless networks in general [5][6] and in highspeed trains in particular [7][8], are discussed in numerous recent publications, using very different algorithms like k-means, deep- and reinforcement learning, support vector machines, optimization algorithms, Decision Trees, Naive Bayes, to name only a few [5][9][10][11][12][13]. Despite the analysis of technical improvements, the possibility to get other information such as the numbers of train travelers using mobile phone data ([14] for an overview), or how passengers use travel time [15][16] are also of interest.

In an increasingly competitive transportation industry, the existence of free mobile Internet access will be a crucial factor to attract new customers [17]. That is why the biggest German railway company "Deutsche Bahn AG" announced to offer a free WiFi system to all travelers on their highspeed trains in the course of their quality improvement program "Zukunft Bahn" [18]. The main goal of this program is to achieve higher customer satisfaction and hence to expand their profits.

This paper contributes to this sphere by analyzing the supply side of mobile Internet access. To be more specific, we analyze the user experience when using the WiFi network that is provided on German highspeed trains (ICEs) in different regions of Germany.

In particular, the analysis of GPS based data with traditional statistical methods is challenging. While, for example, regression approaches are good at describing continuous linear and nonlinear relationships between variables and are computationally simple, they are not suited for clustering data. Algorithms like k-means clustering are better suited but still require quite strong assumptions like knowledge about the exact number of clusters in the data. Therefore, we utilized a self-organized learning neural network to analyze a dataset containing locations specific WiFi data [19]. The usage of this self-organized learning neural network enabled us to analyze the data with only a minimum of prior assumptions needed and without the need of prior variable selection.

The remainder of this paper is structured as follows. Section II describes the Self-Enforcing Network (SEN). In section III, the technology for providing the WiFi on the trains, the data used and methodology are explained. In section IV, the key results are presented focusing on factors affecting the network coverage; section V concludes.

II. THE SELF-ENFORCING NETWORK (SEN)

SEN is a self-organized learning neural network, developed by the Research Group "Computer-Based Analysis of Social Complexity" (CoBASC). Only the functionalities that are relevant to this analysis are briefly presented. More detailed descriptions of the SEN are found in, e.g., [20][21]. The data, consisting of attributes and objects, are represented in a "semantical matrix" where the rows represent the objects o, and the columns represent the attributes a. The values in the matrix w_{ao} represent the degree of affiliation of an attribute to an object. In this case, the semantical matrix contains the preprocessed real data imported from .csv-files (see below).

The training of the network is done by transforming the (min-max normalized) values of the semantical matrix into the weight matrix of the network according to the learning rule. The most specific for SEN is, that the weight values are not generated at random (as usually in neural networks),

 ISBN: 978-1-61208-723-8

meaning that the weight matrix displays the real data. In addition, a “cue validity factor” (cvf) [22][20] is introduced to exclude, to dampen, or to increase the importance of selected attributes. The whole learning rule is then defined as follows (see Equation 1) with w being the assigned weight and c being a constant defined as $0 \leq c \leq 1$:

$$w(t+1) = w(t) + \Delta w, \text{and} \\ \Delta w = c * w_{ao} * cvf_a \tag{1}$$

As in any neural network, the activation functions play an important role. In SEN several activation functions are at disposal [23]; in this study we have used the logarithmic-linear activation function (LLF), which is defined as follows (see Equation 2) with w_{ij} being the value of the (i, j)th Element of the weight matrix and a_i being the corresponding values of the semantical matrix:

$$a_j = \sum \begin{cases} \log_3(a_i + 1) * w_{ij}, & \text{if } a_i \geq 0. \\ \log_3(|a_i + 1|) * -w_{ij}, & \text{else.} \end{cases} \tag{2}$$

The logarithmic-linear activation function is well suited for our purpose because the dataset includes many extreme observations and also differs much across the different variables. This logarithmic-linear function ensures that those extreme values do not receive weights that would be too high otherwise and thus outweigh the other (smaller) values.

III. Data Description and Analysis

To understand how the WiFi on a German Highspeed Train (ICE) works, the technical background will be discussed here shortly.

Figure 1 gives an overview of the installed hardware on the ICEs to provide the local WiFi. The base of this system is the combined infrastructure of the three big telecommunication companies Telekom, Telefónica, and Vodafone, resulting in a more stable system and better network coverage [18]. The signals (the data) coming from cell towers of all three providers are received by the antenna mounted on top of the train. Then the collected stream of data is processed by the router to merge the different signals and is sent to different access points across the train, from which the travelers can access the mobile Internet by connecting to the WiFi Network. To be recognizable, the Service Set Identifier (SSID) of this Network is always set to “WiFionICE”.

A. Data

The dataset called “WiFi on ICE” which is analyzed in this paper is provided by the Deutsche Bahn AG [19]. This dataset consists of about 23.5 million observations with 15 variables. Table I sums up all variables which we have used in our analysis. The sid stores a unique ID of each router used in the trains. Together with the GPS variables, it is possible to match observations to individual train connections. Furthermore, the rate of downloads and the latency are essential factors that influence the user experience. High rates of download are good because the content will be downloaded faster, e.g., a website can be displayed fast, or a movie can be played at higher image quality. High latency values, on the other hand, are undesirable. The latency values describe the time which passes until an initial response from the server is received.

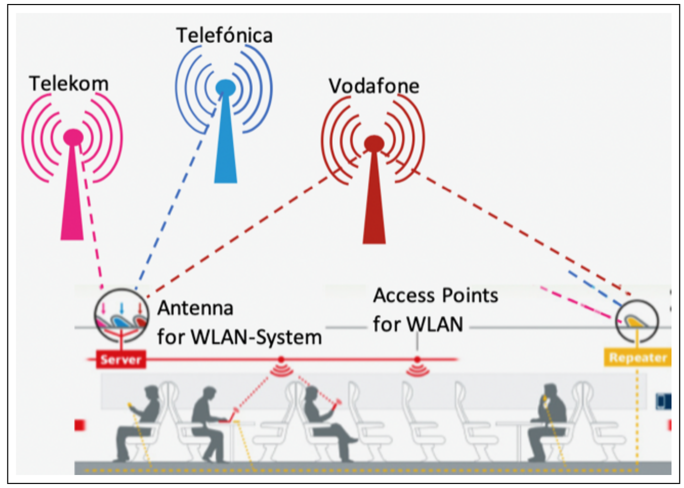

Figure 1. WiFi system on ICE trains [18].

TABLE I. SELECTED VARIABLES.

Variable	*Description*
sid	ID of the X6-router
gps_breite	Latitude
gps_laenge	Longitude
pax_auth	No. of authenticated devices in the local WiFi
pax_total	No. of total devices in the local WiFi
tprx	Rate of downloads (in bytes/s)
tptx	Rate of uploads (in bytes/s)
link_ping	Latency (in ms)
gps_v	Train Speed (in m/s)
gps_richtung	Direction (in degrees)

Additionally, we included the rate of uploads and data on the number of devices connected to the network to check the consistency of our results and to interpret the latter.

B. Preprocessing the Data

To get a first impression of the dataset, we utilized R-Studio. R-Studio is an integrated development environment which allows to comfortably program in the statistical programming language R [24].
Because of the huge number of observations looking at them individually was not feasible. Therefore, we looked at first at the number of missing values. In 2 variables, namely *gps_v* and *gps_richtung*, the amount of missing observations exceeds about 10%. In consequence, we omitted all observations that have missing values. This step was necessary to interpret the results later on. Furthermore, we transformed the variable concerning the velocity of the train by multiplying with the factor 3.6 to obtain velocities in kilometers per hour instead of meters per second, which was the original unit of measurement. We excluded observations with an altitude lower than -80 meters and or speed of over 350 kilometers per hour. This is because velocities much larger than this value cannot be reached by an ICE train. Therefore those values have to be seen as results of measurement errors. Lastly, we created a unique identifier which functions as a label. This allows us to easily search for an individual observation of interest when needed. Finally, we are left with 10.1 million observations. Those observations are used to analyze two different train connections across

 ISBN: 978-1-61208-723-8

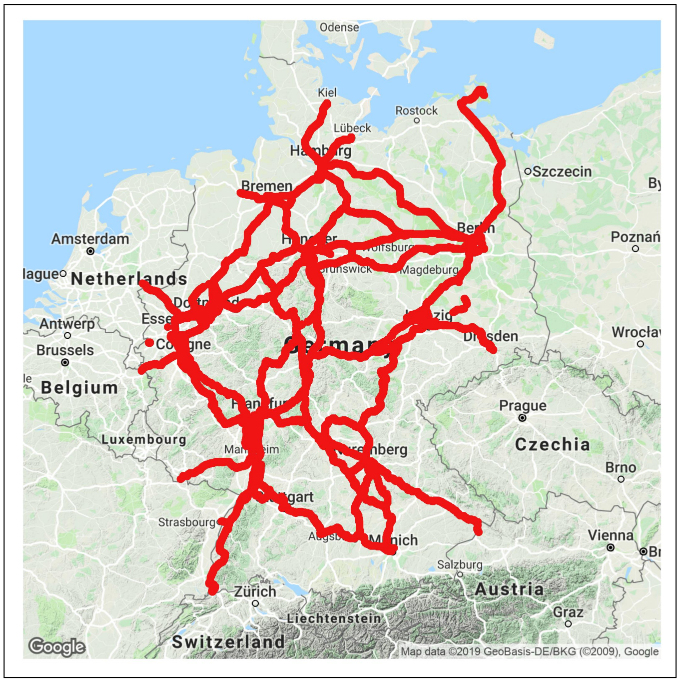

Figure 2. Visualisation of the used dataset.

Germany. We chose connections to cover a good amount of the German railway network. The selection of individual train connections was possible by using the *sid* variable. The dataset after preprocessing, which is used in the subsequent analysis, is visualized in Figure 2.

C. Clustering the Data

For clustering the data with regard to the rate of downloads and the latency, the selected data is reduced once more to $1,000$ observations per chosen connection. Those observations are not chosen at random but regularly over the entire observed period of time (e.g., one observation per minute).

Those in total $2,000$ observations in 8 variables are then labeled with their relative latency, and rate of download and afterward imported into Self-Enforcing Network Second Edition (SEN.SE) the tool for applying the neural network and visualizing the results. The first step then was letting the Self-Enforcing Network (SEN) structure the given data without further information or any adjustments. This procedure resulted in the following visualization (see Figure 3).

There are clearly some clusters, and there is some structure in the data. This is good news because it is evidence that the data is not uniformly distributed, i.e., there is information that can be extracted by further analyzing the dataset.

To steer the clustering towards the variables of interest, the available settings and parameters are set differently. As already mentioned, the Cue Validity Factor (CVF) is a measure for to which extent attributes are associated with a particular category. Therefore one can adjust the importance of an attribute by setting the value of its CVF. If it is set to a high value (i.e., CVF = 1 or possibly even higher), the attribute is highly relevant. If it is set to zero (i.e., CVF = 0), the

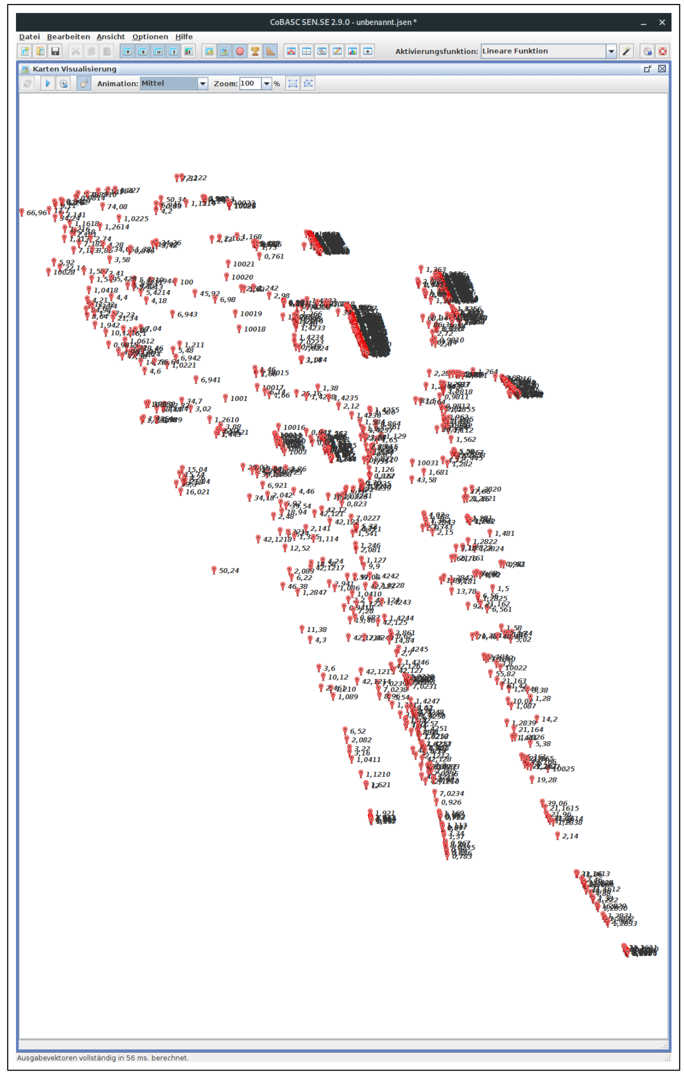

Figure 3. First results of clustering with respect to the normalized latency values.

attribute is not important at all and therefore not considered in the following clustering process.

The value of the CVF for the attributes we want to explain (i.e., the rate of downloads and the latency) is set to the highest value, i.e., 1. The values of the CVF for the other attributes is lowered stepwise until a clear grouping of the objects is reached. A good starting point for setting the CVFs is the correlations between the variables we want to explain and the other covariates. A (in absolute) high correlation indicates that the specific CVF should be set relatively high at the beginning of the process of lowering the values of the CVF for the attributes. The correlations between all the variables in the dataset are visualized in Figure 4. The resulting values of the CVFs are given in Table II. Those adjustments on the SEN then result in the following clustering of the data (see Figure 5). One can see that the data now is ordered in a very different way. Observations with high normalized latency values (red labels) are sorted in the bottom right corner whereas smaller values (i.e., observations with a smaller normalized latency) are sorted in the top left corner (blue labels). The respective

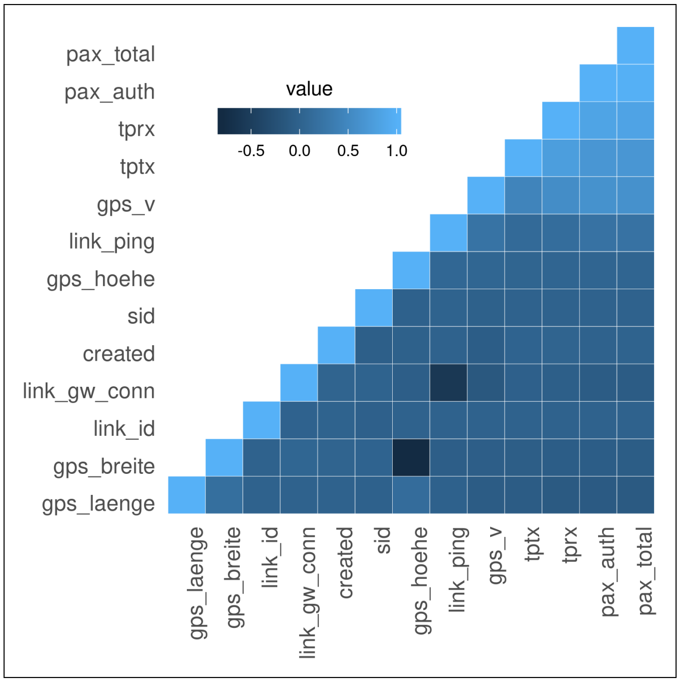

Figure 4. Correlations between the variables.

TABLE II. VALUES OF THE CVFS.

Variable	*Value of CVF to explain latency values*	*Value of CVF to explain rate of downloads*
sid	0	0
gps_breite	0.4	0.3
gps_laenge	0.4	0.3
pax_auth	0.2	0.2
pax_total	0	0
tprx	0	1
tptx	0	0
link_ping	1	0

latency values are displayed right next to the corresponding red markers. This clustering, on the one hand, allows us to select regions of observations with very high or low latency or download rates to further analyze them utilizing R. On the other hand one obtains a good impression how the observations are distributed according to their latency values or download rates.

IV. Results

This section presents the key results obtained from the analysis of two different routes across Germany.

A. Berlin - Cologne

ICE line 10, the connection between Berlin in East Germany and Cologne in West Germany, links two important areas of high population density. As the red points in Figure 6 reveal, the rates of downloads are high, especially at the train stations in the major cities. Furthermore, Figure 6 shows that the line segment between Cologne and Hanover is characterized by high rates of downloads whereas the segment between Hanover and Berlin consists of mostly low rates of downloads (indicated by the blue points) with only a few exceptions. Those high rates can be caused by either many active devices in the WiFi or by a bad connection to the Internet.

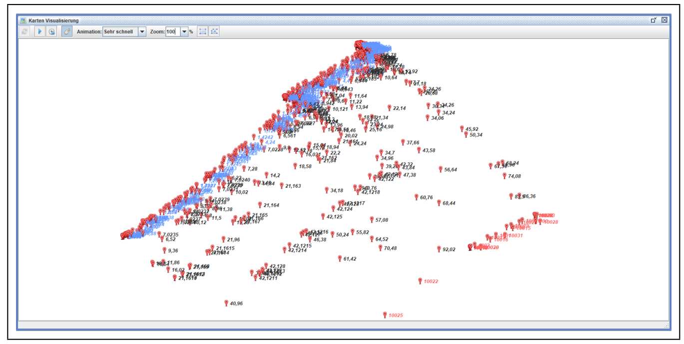

Figure 5. Final clustering with respect to the normalized latency values. Blue labels indicate low values whereas red indicates high latency values.

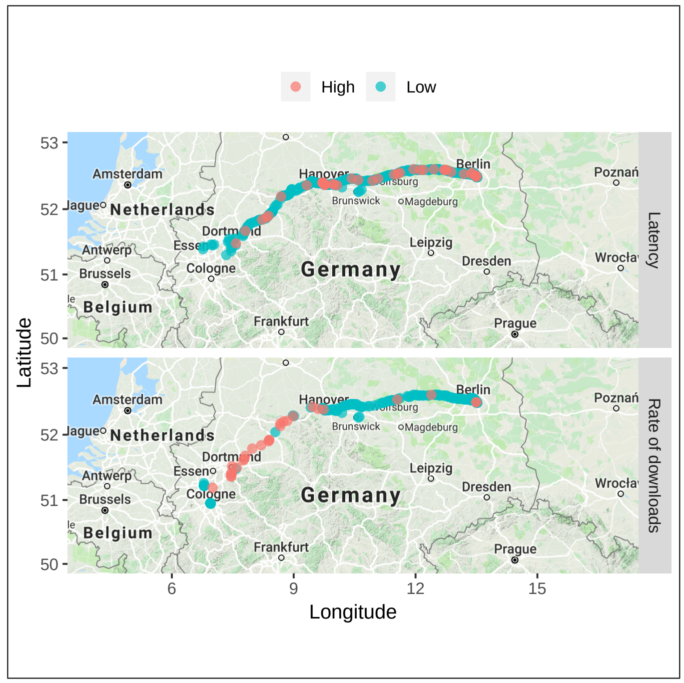

Figure 6. Selected latency values and rates of downloads between Berlin and Cologne.

The latter hypothesis can be evaluated with regard to the latency. Looking at Figure 6 again, it becomes clear, that the latency is far more volatile over the entire length of the connection than the rate of downloads. In particular, between Hanover and Berlin, the latency is constantly relative high explaining the in general lower rates of downloads in this segment. This could possibly be explained by the fact that this region is not highly populated. Therefore the infrastructure of mobile communication may not be that far developed (like for example a larger distance between the cell towers) as for example in the Ruhr area with a much higher population density.

B. Binz - Munich

Figure 7 illustrates the route profile of ICE line 26 connecting Binz in northern Germany and Munich in southern

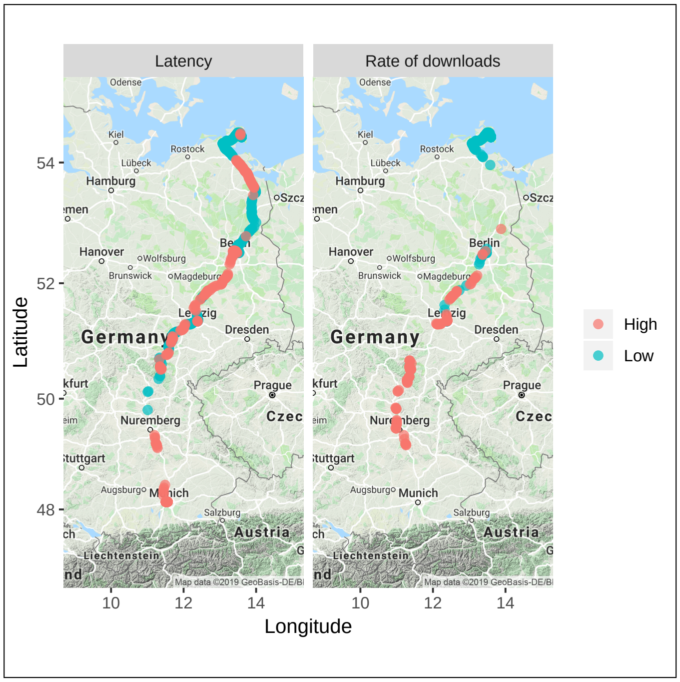

Figure 7. Selected latency values and rates of downloads between Binz and Munich.

Germany. High latency values and high rates of downloads are indicated in red, whereas low ones are blue. It can be seen that on the one hand near Binz, a small coastal town on an island in the Baltic Sea, the rates of downloads are typically low. This comes as no surprise because the population density of this region is quite low. This is also reflected in the usage data, i.e., there are very few devices authenticated to the network. On the other hand, there are quite a few observations with high download rates between Berlin and Nuremberg.

Looking at the latency values, one can notice the low values near Binz. Hence the supply is very good while there is no great demand for it. On the contrary, there is high demand between Berlin and Munich, on average, there are more than 100 devices connected, but the latency is very high most of the time. The latter especially holds between Berlin and Leipzig. Between Leipzig and Nuremberg, the values of the latency are very volatile, which means that the connection to the WiFi might be unstable. Other interesting observations are the high latency values at Munich; they contrast with most observations in other big cities. Unfortunately, the download rates are average at Munich, so it is difficult to assess any further conclusions. At least there is some room for improvement around Munich.

V. Conclusion

In the preceding chapters, we showed how to analyze the WiFi attributes of two selected line-sections of the German railroad network using a SEN. To visualize our results, we used the statistical programming language R. High relevance of our analysis is attributed to the GPS data itself, the download rate, latency and the authenticated devices in the network. While well known statistical approaches suffered fulfilling our requirements, the SEN enabled us to cluster the data without extensive prior model specification.

Our results show a good quality of the WiFi in Northern Germany near Binz. Rather bad quality can be expected when traveling between Berlin and Munich. The identification of areas where the supply of mobile Internet access is good or bad was possible.

We assumed that the user experience using the WiFi mainly depends on the latency and the download rate. To prove this, one could analyze the correlation between data on the user experience and our results. Unfortunately, we were not able to gather data on the user experience, so this remains an exciting area of subsequent research. Furthermore, subsequent research could utilize our methodology to analyze which factors determine the connection between trains and cell towers by including additional data like micro deployment data of the antennas (e.g., the distance between antennas, signal strength, load, etc.). Unfortunately, micro-deployment data is rarely publicly available. This would allow exploiting further relationships between the quality of the WiFi on the ICEs and the mobile network in Germany. Furthermore, one could explore the relationships between train delays and network usage.

While our results cannot be generalized because they are specific for Germany, the generalization of our methodology should be possible. Our analysis mainly relies on the assumption that WiFi access is positively correlated to the customer satisfaction of railway companies. If this assumption holds, and proper data is available, one can conduct the same analysis for other countries.

References

[1] P. Holicza and E. Kadëna, “Smart and Secure? Millennials on Mobile Devices,” Interdisciplinary Description of Complex Systems, vol. 16, no. 3-A, 2018, pp. 376–383, URL: https://ideas.repec.org/a/zna/indecs/v16y2018i3-ap376-383.html [retrieved: 2019-05-14].

[2] J. Martins, C. Costa, T. Oliveira, R. Gonçalves, and F. Branco, “How smartphone advertising influences consumers’ purchase intention,” Journal of Business Research, vol. 94, jan 2019, pp. 378–387, URL: https://linkinghub.elsevier.com/retrieve/pii/S0148296317305507 [retrieved: 2019-05-14].

[3] M. C. Enache, “E-commerce Trends,” ”Dunarea de Jos”, no. 2, 2018, pp. 67–71.

[4] S. Abel, L. Mutandwa, and P. L. Roux, “A Review of Determinants of Financial Inclusion,” International Journal of Economics and Financial Issues, vol. 8, no. 3, 2018, pp. 1–8, URL: https://ideas.repec.org/a/eco/journ1/2018-03-1.html [retrieved: 2019-05-14].

[5] Y. Fu, S. Wang, C.-X. Wang, X. Hong, and S. McLaughlin, “Artificial Intelligence to Manage Network Traffic of 5G Wireless Networks,” IEEE Network, vol. 32, no. 6, nov 2018, pp. 58–64, URL: https://ieeexplore.ieee.org/document/8553655/ [retrieved: 2019-05-14].

[6] H. Zhang and L. Dai, “Mobility Prediction: A Survey on State-of-the-Art Schemes and Future Applications,” IEEE Access, vol. 7, 2019, pp. 802–822, URL: https://ieeexplore.ieee.org/document/8570749/ [retrieved: 2019-05-14]. [Online]. Available: https://ieeexplore.ieee.org/document/8570749/

[7] L. Li et al., “A measurement study on multi-path TCP with multiple cellular carriers on high speed rails,” in Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication - SIGCOMM ’18, 2018, pp. 161–175.

[8] M. Munjal and N. P. Singh, “Group mobility by cooperative communication for high speed railway,” Wireless Networks, jan 2019, pp. 1–10, URL: http://link.springer.com/10.1007/s11276-018-01923-2 [retrieved: 2019-05-14].

[9] R. Cheng, Y. Song, D. Chen, and X. Ma, “Intelligent Positioning Approach for High Speed Trains Based on Ant Colony Optimization and Machine Learning Algorithms,” IEEE Transactions on Intelligent

Transportation Systems, 2018, pp. 1–10, URL: https://ieeexplore.ieee.org/document/8527682/ [retrieved: 2019-05-14].

[10] Y. Bi, J. Zhang, Q. Zhu, W. Zhang, L. Tian, and P. Zhang, "A Novel Non-Stationary High-Speed Train (HST) Channel Modeling and Simulation Method," IEEE Transactions on Vehicular Technology, vol. 68, no. 1, jan 2019, pp. 82–92, URL: https://ieeexplore.ieee.org/document/8542786/ [retrieved: 2019-05-14].

[11] C. Wu, Y. Wang, and Z. Yin, "Realizing Railway Cognitive Radio: A Reinforcement Base-Station Multi-Agent Model," pp. 1–16, 2018, URL: https://ieeexplore.ieee.org/document/8424487/ [retrieved: 2019-05-14].

[12] Q. Mao, F. Hu, and Q. Hao, "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey," pp. 2595–2621, 2018, URL: https://ieeexplore.ieee.org/document/8382166/ [retrieved: 2019-05-14].

[13] K. Sabanci, E. Yigit, D. Ustun, A. Toktas, and M. F. Aslan, "WiFi Based Indoor Localization: Application and Comparison of Machine Learning Algorithms," in 2018 XXIIIrd International Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory (DIPED). IEEE, sep 2018, pp. 246–251, URL: https://ieeexplore.ieee.org/document/8543125/ [retrieved: 2019-05-14].

[14] A. Ø. Sørensen, N. O. E. Olsson, M. M. Akhtar, and H. Bull-Berg, "Approaches, technologies and importance of analysis of the number of train travellers," Urban, Planning and Transport Research, vol. 7, no. 1, jan 2019, pp. 1–18, URL: https://www.tandfonline.com/doi/full/10.1080/21650020.2019.1566022 [retrieved: 2019-05-14].

[15] J. Tang, F. Zhen, J. Cao, and P. L. Mokhtarian, "How do passengers use travel time? A case study of Shanghai Nanjing high speed rail," Transportation, vol. 45, no. 2, mar 2018, pp. 451–477, URL: http://link.springer.com/10.1007/s11116-017-9824-9 [retrieved: 2019-05-14].

[16] B. Wang and B. P. Loo, "Travel time use and its impact on high-speed-railway passengers' travel satisfaction in the e-society," pp. 1–13, may 2018, URL: https://www.tandfonline.com/doi/full/10.1080/15568318.2018.1459968 [retrieved: 2019-05-14].

[17] U. Kluge, A. Paul, M. Urban, and H. Ureta, "Assessment of Passenger Requirements Along the Door-to-Door Travel Chain." Springer, Cham, 2019, pp. 255–276, URL: http://link.springer.com/10.1007/978-3-319-99756-8{_}17 [retrieved: 2019-05-14].

[18] Deutsche Bahn AG, "Zukunft Bahn: Statusbericht Nordrhein-Westfalen 2016," DB AG, Tech. Rep., 2018, URL: https://www.deutschebahn.com/resource/blob/1296202/c1423d8966c791c726550fe9e98ecf20/20161129-Praeesentation-Zukunft-Bahn-NRW-data.pdf [retrieved: 2019-05-14].

[19] Deutsche Bahn, "Dataset: wifi-on-ice," 2019, URL: https://data.deutschebahn.com/dataset/wifi-on-ice [retrieved: 2019-05-14].

[20] C. Klüver, "Steering clustering of medical data in a Self-Enforcing Network (SEN) with a cue validity factor," in 2016 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, dec 2016, pp. 1–8, URL: http://ieeexplore.ieee.org/document/7849883/ [retrieved: 2019-05-14].

[21] C. Klüver, J. Klüver, and D. Zinkhan, "A self-enforcing neural network as decision support system for air traffic control based on probabilistic weather forecasts," in Proceedings of the International Joint Conference on Neural Networks, vol. 2017-May. IEEE, may 2017, pp. 729–736, URL: http://ieeexplore.ieee.org/document/7965924/ [retrieved: 2019-05-14]. [Online]. Available: http://ieeexplore.ieee.org/document/7965924/

[22] E. Rosch and C. B. Mervis, "Family resemblances: Studies in the internal structure of categories," Cognitive Psychology, vol. 7, no. 4, oct 1975, pp. 573–605, URL: https://www.sciencedirect.com/science/article/pii/0010028575900249 [retrieved: 2019-05-14].

[23] C. Klüver, "Self-Enforcing Networks (SEN) for the development of (medical) diagnosis systems," in Proceedings of the International Joint Conference on Neural Networks, vol. 2016-Octob. IEEE, jul 2016, pp. 503–510, URL: http://ieeexplore.ieee.org/document/7727241/ [retrieved: 2019-05-14].

[24] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019, URL: https://www.R-project.org/ [retrieved: 2019-05-14].

 ISBN: 978-1-61208-723-8

A Linear Matrix Inequality Based Strategy For Maximum Amplitude Analysis In Discrete Time Systems

Graziano Chesi
Department of Electrical and Electronic Engineering
The University of Hong Kong
Email: chesi@eee.hku.hk

Tiantian Shen
Department of Electronic Information Engineering
Hunan Normal University
Email: titiansaint@foxmail.com

***Abstract*—A typical problem in engineering consists of quantifying the maximum amplitude of the signals of discrete-time linear time-invariant systems. This paper suggests a strategy based on the solution of a convex optimization problem, in particular, a linear matrix inequality feasibility problem. The proposed strategy is motivated by the fact that the existing methods are generally conservative.**

Keywords: Discrete-Time; Maximum Amplitude; LMI.

I. INTRODUCTION

Discrete-time linear time-invariant systems represent a wide class of control systems of interest [1]–[3]. For instance, they can model a worktable motion control system or a space station orientation control system where digital control is implemented through the use of digital-to-analog and analog-to-digital converters [4]. A typical problem in these systems consists of quantifying the maximum amplitude of their signals. Indeed, this problem is met in order to establish that physical quantities do not exceed their allowable limits, a necessary step for ensuring safety.

This paper suggests a strategy for dealing with this problem. This strategy exploits polynomials and the Gram matrix method [5] [6], and requires the solution of a Linear Matrix Inequality (LMI) feasibility problem, which belongs to the class of convex optimization.

The paper is organized as follows. Section II discusses some related works. Section III introduces the problem formulation. Section IV presents a motivating example. Section V described the proposed strategy. Section VI reports the conclusions.

II. RELATED WORKS

The problem of quantifying the maximum amplitude of the signals of linear time-invariant systems has been studied in the literature. In particular, LMI methods have been sought because they can be tested through convex optimization and because they may be exploited in the design of feedback controllers. A pioneering work is [7], which considers the canonical problem of quantifying the peak of the impulse response, and shows how upper bounds can be established via LMIs based on set invariance of quadratic Lyapunov functions. Another pioneering work is [8], which describes how this methodology can be used in the design of feedback controllers ensuring desired upper bounds on the peak of the impulse response. The reader is also referred to recent works such as [9] [10] where this methodology is exploited in optimal control and in model predictive control.

However, the methodology exploited in these related works is generally conservative and may provide upper bounds that are rather far from the sought quantities as it will be shown in Section IV. The idea proposed in this paper aims at providing an LMI-based strategy where conservatism is eliminated. This paper extends our previous work [11] where only continuous-time systems are considered.

III. PROBLEM FORMULATION

The notation is as follows. The set of integer numbers is denoted by $\mathbb{Z}$. The set of real numbers is denoted by $\mathbb{R}$. The transpose of a matrix A is denoted by A'. The infinity norm of a matrix A is denoted by $\|A\|_\infty$, i.e., $\|A\|_\infty = \max_{i,j} |a_{i,j}|$ where $a_{i,j}$ is the entry of A on the (i,j)-th position.

We consider the discrete-time linear time-invariant system

$$\left\{ \begin{array}{rcl} x(t+1) & = & Ax(t) + Bu(t) \\ y(t) & = & Cx(t) \end{array} \right. \tag{1}$$

where $t \in \mathbb{Z}$ is the time, $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input, $y(t) \in \mathbb{R}^p$ is the output, and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{p\times n}$ are given matrices. For the system (1), we consider the problem in canonical form described hereafter. Let us start by introducing the following definition.

Definition 1: The impulse response of the system (1) with respect to the i-th input channel is the function $Y_i(t)$, defined as the solution $y(t)$ for initial condition $x(0) = 0$ and input $u(t) = \delta(t)E_i$, where E_i is the i-th column of the $m \times m$ identity matrix, and $\delta(t)$ is the impulse defined by

$$\delta(t) = \left\{ \begin{array}{ll} 1 & \text{if } t = 0 \\ 0 & \text{else.} \end{array} \right. \tag{2}$$

The problem addressed in this paper is as follows.

Problem 1: Given $c \in (0,\infty)$, establish whether c is an upper bound of the maximum amplitude of the impulse response of the system (1) with respect to all input channels, i.e.,

$$\|Y_i(t)\|_\infty < c \quad \forall t \geq 0 \quad \forall i = 1,\ldots,m. \tag{3}$$

IV. Motivating Example

In order to motivate the strategy proposed in this paper, let us consider the simple discrete-time linear time-invariant system described by

$$\begin{cases} x(t+1) &= \begin{pmatrix} 0 & 1 \\ -0.3 & 1 \end{pmatrix} x(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t) \\ y(t) &= \begin{pmatrix} 1 & 0 \end{pmatrix} x(t). \end{cases}$$

This is a second-order system (i.e., $x(t) \in \mathbb{R}^2$), with scalar input and scalar output (i.e., $u(t), y(t) \in \mathbb{R}$). The impulse response of this system, as defined in Definition 1, is shown in Figure 1.

Let us use the LMI methods [7] [8] based on set invariance of quadratic Lyapunov functions for establishing upper bounds of the impulse response of this system. These methods are formulated as LMI feasibility problems, and provide the upper bound 1.307 shown in Figure 1. It is interesting to observe from the figure that, in spite of the simplicity of the system under consideration, this upper bound is conservative.

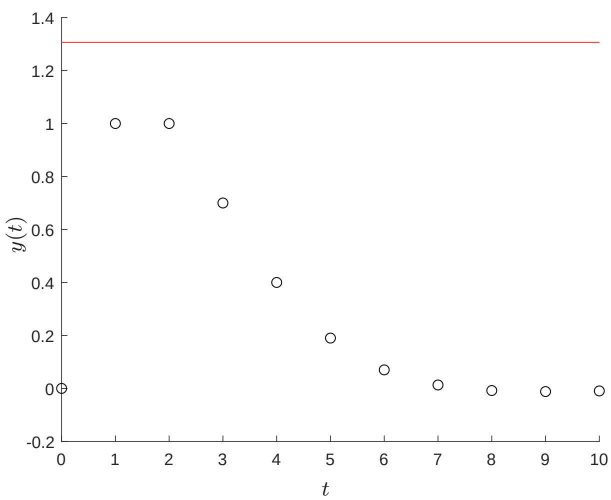

Fig. 1. Black circles: impulse response. Red line: upper bound provided by [7], [8].

V. Proposed Strategy

In this section, we describe the strategy proposed for dealing with Problem 1. The first idea is to describe the trajectories of the system (1) that correspond to its impulse responses through the sublevel set of a polynomial. Specifically, we denote this sublevel set as

$$\mathcal{V} = \{x \in \mathbb{R}^n : \ v(x) \leq 1\} \tag{4}$$

where $v(x)$ is a polynomial. The polynomial $v(x)$ should be determined under two main conditions. The first condition is that the trajectories of the system (1) that correspond to its impulse responses are included in the sublevel set $\mathcal{V}$. The second condition is that the sublevel set $\mathcal{V}$ is included in the set of states for which the maximum amplitude among the entries of the corresponding output is smaller than the desired upper bound c. By denoting with $\mathcal{W}$ the latter set, we have

$$\mathcal{W} = \{x \in \mathbb{R}^n : \ \|Cx\|_\infty < c\}. \tag{5}$$

In order to impose the first condition, a possibility is to require that $v(x)$ does not increase with the time starting from a point on the trajectories of the system (1) that correspond to its impulse responses. This can be expressed as follows. Let us consider the i-th input channel of the system (1), and let B_i be the i-th column of B. Let σ be a nonnegative integer. Then,

$$\begin{cases} c &> \left\|CA^k B_i\right\|_\infty \quad \forall k = 0, \ldots, \sigma - 1 \\ 1 &\geq v(A^\sigma B_i) \\ v(x) &\geq v(Ax). \end{cases} \tag{6}$$

In order to impose the second condition, let us observe that the inclusion of the sublevel set $\mathcal{V}$ in the set $\mathcal{W}$ can be expressed by the condition

$$v(x) > 1 \quad \forall x : \ \|Cx\|_\infty \geq c. \tag{7}$$

This condition could be imposed through the introduction of polynomial multipliers, which are typically exploited in order to convert local properties (such as, local positivity) into global properties (such as, global positivity).

Lastly, in order to search for a polynomial $v(x)$ that satisfies the conditions previously mentioned, one could exploit the Gram matrix method [5] [6]. Indeed, let us consider a polynomial $p(x)$ of degree not greater than $2d$, where d is a nonnegative integer. Then, $p(x)$ can be expressed through the Gram matrix method as

$$p(x) = b(x)' \left(P + L(\alpha)\right) b(x) \tag{8}$$

where $b(x)$ is a vector of monomials in x of degree not greater than d, P is a symmetric matrix, $L(\alpha)$ is a linear parameterization of the linear set

$$\mathcal{L} = \left\{\tilde{L} = \tilde{L}' : \ b(x)'\tilde{L}b(x) = 0\right\}, \tag{9}$$

and α is a free vector. The existence of α satisfying the LMI

$$P + L(\alpha) \geq 0 \tag{10}$$

is equivalent to the property that the polynomial $p(x)$ is a sum of squares of polynomials. Since this property implies that $p(x)$ is nonnegative, one could exploit the Gram matrix method to turn the search for a polynomial $v(x)$ into an LMI feasibility problem.

VI. Conclusions

This paper has suggested an LMI-based strategy for quantifying the maximum amplitude of the signals of discrete-time linear time-invariant systems. Future work will investigate the realization of this strategy and its use in the design of feedback controllers.

ACKNOWLEDGEMENTS

The authors would like to thank the Reviewers for their useful comments. This paper is supported by the National Natural Science Foundation of China (Grant No. 61803152).

REFERENCES

[1] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory*. New York: McMillan, 1992.
[2] P. J. Antsaklis and A. N. Michel, *Linear Systems*. Cambridge University Press, 2006.
[3] L. Qiu and K. Zhou, *An Introduction to Feedback Control*. Prentice Hall, 2010.
[4] R. C. Dorf and R. H. Bishop, *Modern Control Systems*. Prentice Hall, 2016.
[5] B. Reznick, “Some concrete aspects of Hilbert’s 17th problem,” *Contemporary Mathematics*, vol. 253, pp. 251–272, 2000.
[6] G. Chesi, “LMI techniques for optimization over polynomials in control: a survey,” *IEEE Transactions on Automatic Control*, vol. 55, no. 11, pp. 2500–2510, 2010.
[7] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.
[8] C. W. Scherer, P. Gahinet, and M. Chilali, “Multiobjective output-feedback control via LMI optimization,” *IEEE Transactions on Automatic Control*, vol. 42, no. 7, pp. 896–911, 1997.
[9] A. White, G. Zhu, and J. Choi, “Optimal LPV control with hard constraints,” *International Journal of Control, Automation and Systems*, vol. 14, no. 1, pp. 148–162, 2016.
[10] W. Jiang, H. Wang, J. Lu, W. Qin, and G. Cai, “Synthesis of mixed objective output feedback robust model predictive control,” *Asian Journal of Control (in press)*, 2017.
[11] G. Chesi and T. Shen, “On the computation of the peak of the impulse response of LTI systems,” in *International Conference on Information Science and System*, Tokyo, Japan, 2019.

A Comparison of Global and Saturated Probabilistic Approximations Using Characteristic Sets in Mining Incomplete Data

Patrick G. Clark
Department of Electrical Engineering
and Computer Science,
University of Kansas
Lawrence, KS, USA
Email: patrick.g.clark@gmail.com

Jerzy W. Grzymala-Busse
Department of Electrical Engineering
and Computer Science,
University of Kansas,
Lawrence, KS, USA
Department of Expert Systems
and Artificial Intelligence,
University of Information
Technology and Management,
Rzeszow, Poland
Email: jerzy@ku.edu

Teresa Mroczek and Rafal Niemiec
Department of Expert Systems
and Artificial Intelligence,
University of Information
Technology and Management,
Rzeszow, Poland
Email: tmroczek@wsiz.rzeszow.pl
and rniemiec@wsiz.rzeszow.pl

***Abstract*—Data mining systems form granules of information from data sets. Methods used to construct these granules can significantly impact the overall accuracy of the resulting model. In this paper, we study incomplete data sets with two interpretations of missing attribute values, lost values and "do not care" conditions, to determine the best method between two approaches and achieve the highest accuracy. For such incomplete data sets, we apply data mining based on two probabilistic approximations, global and saturated. The main objective of our paper is to compare both approaches in terms of an error rate, evaluated by ten-fold cross validation. Saturated probabilistic approximations are closer to the concept than global probabilistic approximations, so the corresponding error rate should be smaller. Using a 5% level of significance, our main result shows that there are differences between both approaches. However, in general neither is better for all data sets and thus, both approaches should be tried for each data set with the best selected for rule induction.**

Keywords–Data mining; rough set theory; probabilistic approximations; MLEM2 rule induction algorithm; lost values; "do not care" conditions.

I. INTRODUCTION

In this paper, we use two interpretations of a missing attribute value: lost values and "do not care" conditions. Lost values indicate that the original values were erased, and as a result we should use only existing, specified attribute values for rule induction. A lost value is denoted by "?". "Do not care" conditions mean that the missing attribute value may be replaced by any specified attribute value. A "do not care" condition is denoted by "*".

We use for data mining probabilistic approximations, a generalization of the idea of lower and upper approximations known in rough set theory. A probabilistic approximation is associated with a parameter (probability) α, if α = 1, a probabilistic approximation is reduced to the lower approximation; if α is a small positive number, e.g., 0.001, the probabilistic approximation becomes the upper approximation. Usually, probabilistic approximations are applied to completely specified data sets [1]–[9]. Such approximations were generalized to incomplete data sets in [10].

Characteristic sets were introduced in [11] for incomplete data sets with any interpretation of missing attribute values. One of the methods being used in this paper, global probabilistic approximations, were introduced in [12]. This prior work expanded the ideas of characteristic sets and also studied their performance as a data mining tool. To further improve our methodology, we introduce a new idea of saturated probabilistic approximations. Our main objective is to compare both approaches in terms of an error rate, evaluated by ten-fold cross validation. The Modified Learning from Examples Module, version 2 (MLEM2) [13] was used for rule induction.

This paper starts with a discussion on incomplete data in Section II where we define attribute-value blocks and characteristic sets. In Section III, we present two types of probabilistic approximations, global and saturated. Section IV contains the details of our experiments. Finally, conclusions are presented in Section V.

II. INCOMPLETE DATA

We assume that the input data sets are presented in the form of a decision table. An example of a decision table is shown in Table I. Rows of the decision table represent cases, while columns are labeled by variables. The set of all cases will be denoted by U. In Table I, U = {1, 2, 3, 4, 5, 6, 7, 8}. Independent variables are called attributes and a dependent variable is called a decision and is denoted by d. The set of all attributes will be denoted by A. In Table I, A = {*Temperature*, *Headache*, *Cough*}. The value for a case x and an attribute a will be denoted by $a(x)$. For example, *Temperature*(1) = *high*.

The set X of all cases defined by the same value of the decision d is called a *concept*. For example, a concept associated with the value *yes* of the decision *Flu* is the set {1, 2, 3, 4}.

For a variable a and its value v, (a, v) is called a variable-value pair. A *block* of (a, v), denoted by $[(a, v)]$, is the set $\{x \in U \mid a(x) = v\}$ [14]. For incomplete decision tables, the

 ISBN: 978-1-61208-723-8

TABLE I. A DECISION TABLE

	Attributes			Decision
Case	Temperature	Headache	Cough	Flu
1	high	yes	*	yes
2	*	no	no	yes
3	very-high	*	yes	yes
4	high	no	*	yes
5	normal	yes	no	no
6	high	*	no	no
7	?	no	*	no
8	normal	yes	yes	no

definition of a block of an attribute-value pair is modified in the following way:

- if for an attribute a and a case x we have $a(x) = ?$, the case x should not be included in any blocks $[(a, v)]$ for all values v of attribute a;
- if for an attribute a and a case x we have $a(x) = *$, the case x should be included in blocks $[(a, v)]$ for all specified values v of attribute a.

For the data set from Table I, the blocks of attribute-value pairs are:

[(Temperature, normal)] = {2, 5, 8},
[(Temperature, high)] = {1, 2, 4, 6},
[(Temperature, very-high)] = {2, 3},
[(Headache, no)] = {2, 3, 4, 6, 7},
[(Headache, yes)] = {1, 3, 5, 6, 8},
[(Cough, no)] = {1, 2, 4, 5, 6, 7}, and
[(Cough, yes)] = {1, 3, 4, 7, 8}.

For a case $x \in U$ and $B \subseteq A$, the *characteristic set* $K_B(x)$ is defined as the intersection of the sets $K(x, a)$, for all $a \in B$, where the set $K(x, a)$ is defined in the following way:

- if $a(x)$ is specified, then $K(x, a)$ is the block $[(a, a(x))]$ of attribute a and its value $a(x)$;
- if $a(x) = ?$ or $a(x) = *$, then $K(x, a) = U$.

For example, for Table I and $B = A$,

$$K(1, Temperature) = [(Temperature, high)] = \{1, 2, 4, 6\},$$

$$K(1, Headache) = [(Headache, yes)] = \{1, 3, 5, 6, 8\}$$

and

$$K(1, Cough) = U$$

so $K_A(1) = \{1, 2, 4, 6\} \cap \{1, 3, 5, 6, 8\} \cap U = \{1, 6\}$.

Similarly,

$K_A(2)$ = {2, 4, 6, 7},
$K_A(3)$ = {3},
$K_A(4)$ = {2, 4, 6},
$K_A(5)$ = {5},
$K_A(6)$ = {1, 2, 4, 6},
$K_A(7)$ = {2, 3, 4, 6, 7}, and
$K_A(8)$ = {8}.

III. PROBABILISTIC APPROXIMATIONS

In this section, we will discuss two types of probabilistic approximations: global and saturated.

A. Global Probabilistic Approximations

An idea of the global probabilistic approximation, though restricted only to lower and upper approximations, was introduced in [15][16], and presented in a general form in [12]. Let X be a concept, $X \subseteq U$. A *B-global probabilistic approximation* of the concept X, based on characteristic sets, with the parameter α and denoted by $appr^{global}_{\alpha,B}(X)$ is defined as the following set

$$\bigcup \{K_B(x) \mid \exists\, Y \subseteq U\ \forall\, x \in Y,\ Pr(X|K_B(x))\ \geq \alpha\}. \quad (1)$$

In general, for given sets B and X and the parameter α, there exist many B-global probabilistic approximations of X. Additionally, the algorithm for computing B-global probabilistic approximations is of exponential computational complexity. Therefore, we decided to use a heuristic version of the definition of B-global probabilistic approximation, called a MLEM2 B-global probabilistic approximation of the concept X, associated with a parameter α and denoted by $appr^{mlem2}_{\alpha,B}(X)$, [12]. This definition is based on the rule induction algorithm MLEM2 [17]. The MLEM2 algorithm is used in the Learning from Examples using Rough Sets (LERS) data mining system [17]–[19]. The approximation $appr^{mlem2}_{\alpha,B}(X)$ is constructed from characteristic sets $K_B(y)$, the most relevant to the concept X, i.e., with $|X \cap K_B(y)|$ as large as possible and $Pr(X|K_B(y)) \geq \alpha$, where $y \in U$. If more than one characteristic set $K_B(y)$ satisfies both conditions, we pick the characteristic set $K_B(y)$ with the largest $Pr(X|K_B(y))$. If this criterion ends up with a tie, a characteristic set is picked up heuristically, as the first on the list [12].

In this paper, we study MLEM2 B-global probabilistic approximations based on characteristic sets, with $B = A$, and calling them, for simplicity, *global probabilistic approximations* associated with the parameter α, denoted by $appr^{mlem2}_{\alpha}(X)$. Similarly, for $B = A$, the characteristic set $K_B(X)$ is denoted by $K(x)$.

Let $E_\alpha(X)$ be the set of all eligible characteristic sets defined as follows

$$\{K(x) \mid x \in U, Pr(X|K(x)) \geq \alpha\}. \quad (2)$$

A heuristic version of the MLEM2 global probabilistic approximation is computed using the algorithm specified in Figure 1.

For Table I, all distinct MLEM2 global probabilistic approximations are

$$appr^{mlem2}_{1}(\{1, 2, 3, 4\}) = \{3\},$$

$$appr^{mlem2}_{0.75}(\{1, 2, 3, 4\}) = \{1, 2, 3, 4, 6\},$$

$$appr^{mlem2}_{0.6}(\{1, 2, 3, 4\}) = \{1, 2, 3, 4, 6, 7\},$$

$$appr^{mlem2}_{1}(\{5, 6, 7, 8\}) = \{5, 8\},$$

$$appr^{mlem2}_{0.5}(\{5, 6, 7, 8\}) = \{2, 4, 5, 6, 7, 8\} \text{ and}$$

$$appr^{mlem2}_{0.4}(\{5, 6, 7, 8\}) = \{2, 3, 4, 5, 6, 7, 8\}.$$

MLEM2 global probabilistic approximation algorithm
input: a set X (a concept), a set $E_\alpha(X)$,
output: a set T ($appr_\alpha^{mlem2}(X)$)
begin
  $G := X$;
  $T := \emptyset$;
  $Y := E_\alpha(X)$;
  while $G \neq \emptyset$ **and** $Y \neq \emptyset$
    begin
      select a characteristic set $K(x) \in Y$
      such that $|K(x) \cap X|$ is maximum;
      if a tie occurs, select $K(x) \in Y$
      with the smallest cardinality;
      if another tie occurs, select the first $K(x)$;
      $T := T \cup K(x)$;
      $G := G - T$;
      $Y := Y - K(x)$
    end
end

Figure 1. MLEM2 Global Approximation Algorithm

Saturated probabilistic approximation algorithm
input: a set X (a concept), index m,
a set $E_i(x)$ for i = 1, 2,..., n and $x \in U$,
output: a set T ($appr_{\alpha_m}^{saturated}(X)$)
begin
  $T := \emptyset$;
  $Y_i(x)$:= $E_i(x)$ for all i = 1, 2,..., m and $x \in U$;
  for j = 1, 2,..., m **do**
    while $Y_j(x) \neq \emptyset$
      begin
        select a characteristic set $K(x) \in Y_j(x)$
        such that $|K(x) \cap X|$ is maximum;
        if a tie occurs, select the first $K(x)$;
        $Y_j(x) := Y_j(x) - K(x)$;
        if $(K(x) - T) \cap X \neq \emptyset$
          then $T := T \cup K(x)$;
        if $X \subseteq T$ **then exit**
      end
end

Figure 2. Saturated Probabilistic Approximation Algorithm

B. Saturated Probabilistic Approximations

Another heuristic version of the probabilistic approximation is based on selection of characteristic sets while giving higher priority to characteristic sets with larger conditional probability $Pr(X|K(x))$. Additionally, if the approximation covers all cases from the concept X, we stop adding characteristic sets.

Let X be a concept and let $x \in U$. Let us compute all conditional probabilities $Pr(X|K(x))$. Then, we sort the set

$$\{Pr(X|K(x)) \mid x \in U\}. \quad (3)$$

Let us denote the sorted list of such conditional probabilities by α_1, α_2,..., α_n, where α_1 is the largest. For any i = 1, 2,..., n, the set $E_i(x)$ is defined as follows

$$\{K(x) \mid x \in U, Pr(X|K(x) = \alpha_i\}. \quad (4)$$

If we want to compute a saturated probabilistic approximation, denoted by $appr_\alpha^{saturated}(X)$, for some α, $0 < \alpha \leq 1$, we need to identify the index m such that

$$\alpha_m \geq \alpha > \alpha_{m+1}, \quad (5)$$

where $m \in \{1, 2, ..., n\}$ and α_{n+1} = 0. Then, the saturated probabilistic approximation $appr_{\alpha_m}^{saturated}(X)$ is computed using the algorithm specified in Figure 2.

For Table I, all distinct saturated probabilistic approximations are

$appr_1^{saturated}(\{1, 2, 3, 4\}) = \{3\}$,

$appr_{0.75}^{saturated}(\{1, 2, 3, 4\}) = \{1, 2, 3, 4, 6\}$,

$appr_1^{saturated}(\{5, 6, 7, 8\}) = \{5, 8\}$ and

$appr_{0.5}^{saturated}(\{5, 6, 7, 8\}) = \{2, 4, 5, 6, 7, 8\}$.

Note that $appr_{0.6}^{mlem2}(\{1, 2, 3, 4\})$ covers the case 7 in spite of the fact that the case 7 is not a member of the concept {1, 2, 3, 4}. The set {1, 2, 3, 4, 6, 7} is not listed among saturated probabilistic approximations of the concept {1, 2, 3, 4}.

C. Rule Induction

Once the global and saturated probabilistic approximations associated with a parameter α are constructed, rule sets are induced using the rule induction algorithm based on another parameter, also interpreted as a probability, and denoted by β. This algorithm also uses MLEM2 principles [20], with the algorithm details shown in Figure 3.

For example, for Table I and α = β = 0.5, using the saturated probabilistic approximations, the MLEM2 rule induction algorithm induces the following rules:

(Cough, no) & (Headache, no) $\rightarrow$ (Flu, no),
(Temperature, normal) $\rightarrow$ (Flu, no),
(Temperature, high) $\rightarrow$ (Flu, yes) and
(Temperature, very-high) $\rightarrow$ (Flu, yes).

IV. Experiments

The goal of this research is to select the best approach for rule set induction in data mining. The data sets were chosen to represent various types of data (symbolic or numeric) in an attempt to find the best method. For our experiments, we used eight data sets that are available in the University of California at Irvine *Machine Learning Repository*.

For every data set, a template was created. Such a template was formed by replacing randomly 35% of existing specified attribute values by *lost values*. The same templates were used for constructing data sets with "do not care" conditions, by replacing "?"s with "$*$"s. The reason that these data sets were selected is because they represent a reasonable distribution of types of data and data set sizes to measure the impacts of the experiments. Furthermore, 35% is selected as a missing attribute percentage is because it is the maximum percentage that is able to be replaced with all records having at least one value specified.

MLEM2 rule induction algorithm
input: a set Y (an approximation of X) and a parameter β,
output: a set $\mathcal{T}$ (a rule set),
begin
  $G := Y$;
  $D := Y$;
  $\mathcal{T} := \emptyset$;
  $\mathcal{J} := \emptyset$;
  while $G \neq \emptyset$
    begin
      $T := \emptyset$;
      $T_s := \emptyset$;
      $T_n := \emptyset$;
      $T(G) := \{t \mid [t] \cap G \neq \emptyset\}$;
      while ($T = \emptyset$ **or** $[T] \not\subseteq D$) **and** $T(G) \neq \emptyset$
        begin
          select a pair $t = (a_t, v_t) \in T(G)$ with maximum of $|[t] \cap G|$; if a tie occurs, select a pair $t \in T(G)$ with the smallest cardinality of $[t]$; if another tie occurs, select first pair;
          $T := T \cup \{t\}$;
          $G := [t] \cap G$;
          $T(G) := \{t \mid [t] \cap G \neq \emptyset\}$;
          if a_t is symbolic {let V_{a_t} be the domain of a_t}
            then
              $T_s := T_s \cup \{(a_t, v) | v \in V_{a_t}\}$
          else {a_t is numerical, let $t = (a_t, u..v)$} and $T_n := T_n \cup \{(a_t, x..y) \mid$ disjoint $x..y$ and $u..v\} \cup \{(a_t, x..y) \mid x..y \supseteq u..v\}$;
          $T(G) := T(G) - (T_s \cup T_n)$;
        end {while};
        if $Pr(X \mid [T]) \geq \beta$
          then
            begin
              $D := D \cup [T]$;
              $\mathcal{T} := \mathcal{T} \cup \{T\}$;
            end {then}
          else $\mathcal{J} := \mathcal{J} \cup \{T\}$;
        $G := D - \cup_{S \in \mathcal{T} \cup \mathcal{J}}[S]$;
    end {while};
  for each $T \in \mathcal{T}$ **do**
    for each numerical attribute a_t with $(a_t, u..v) \in T$ **do**
      while (T contains at least two different pairs $(a_t, u..v)$ and $(a_t, x..y)$ with the same numerical attribute a_t)
        replace these two pairs with a new pair $(a_t$, common part of $(u..v)$ and $(x..y))$;
    for each $t \in T$ **do**
      if $[T - \{t\}] \subseteq D$ **then** $T := T - \{t\}$;
  for each $T \in \mathcal{T}$ **do**
    if $\cup_{S \in (\mathcal{T} - \{T\})}[S] = \cup_{S \in \mathcal{T}}[S]$ **then** $\mathcal{T} := \mathcal{T} - \{T\}$;
end {procedure}.

Figure 3. MLEM2 Rule Induction Algorithm

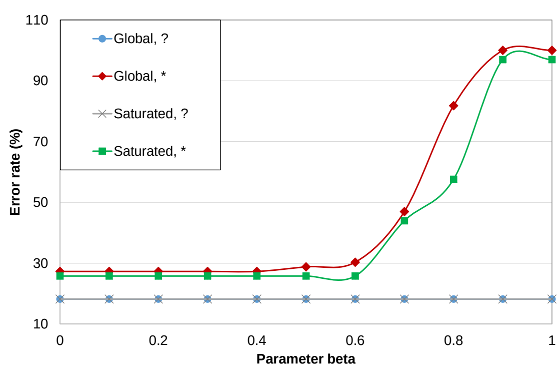

Figure 4. Error rate for the *Bankruptcy* data

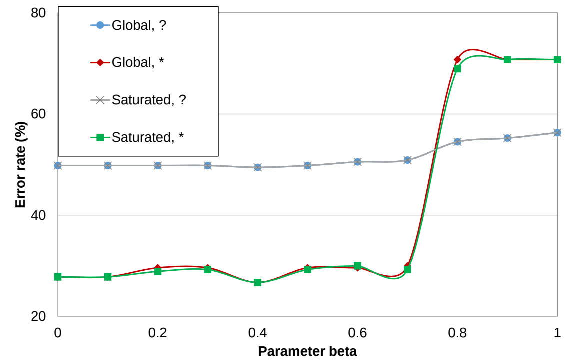

Figure 5. Error rate for the *Breast cancer* data set

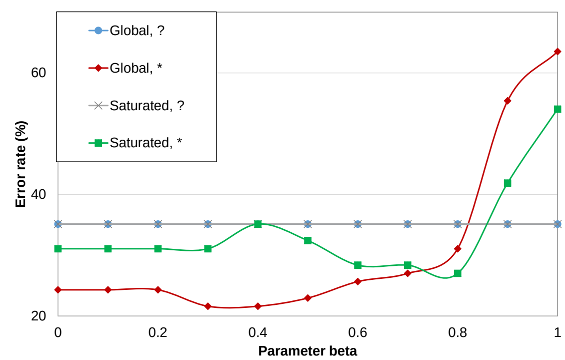

Figure 6. Error rate for the *Echocardiogram* data set

In our experiments, we used the MLEM2 rule induction algorithm. In all experiments, the parameter α was equal to 0.5. Results of our experiments are presented in Figures 4–11, where "Global" denotes a MLEM2 global probabilistic approximation, "Saturated" denotes a saturated probabilistic approximation, "?" denotes lost values and "$*$" denotes "do not care" conditions. In our experiments, four approaches for mining incomplete data sets were used, since we combined two options of probabilistic approximations: global and saturated with two interpretations of missing attribute values: lost and "do not care" conditions.

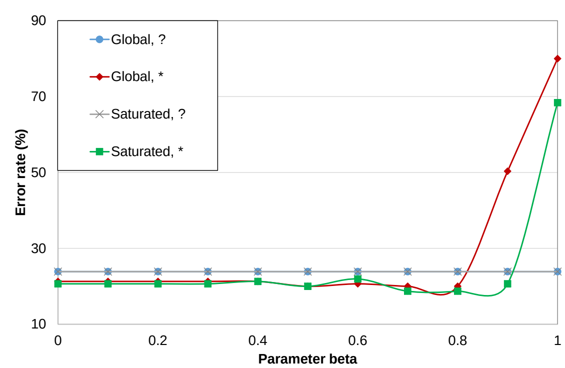

Figure 7. Error rate for the *Hepatitis* data set

 ISBN: 978-1-61208-723-8

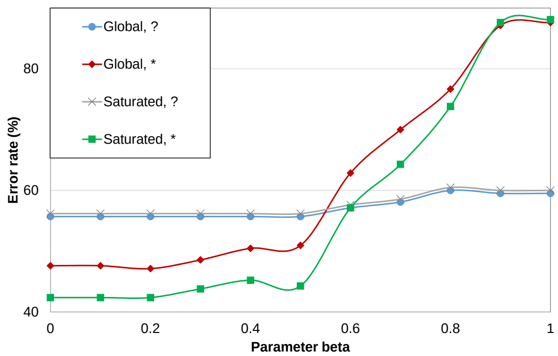

Figure 8. Error rate for the *Image segmentation* data set

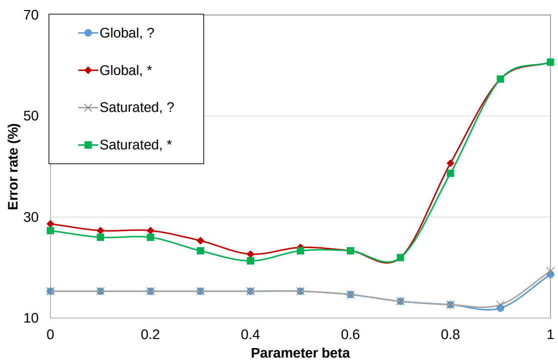

Figure 9. Error rate for the *Iris* data set

These four approaches were compared by applying the distribution free Friedman rank sum test and then by the post-hoc test (distribution-free multiple comparisons based on the Friedman rank sums), with a 5% level of significance. The null hypothesis H_0 of the Friedman test, claiming that differences between these approaches are insignificant, was rejected for *Breast cancer* and *Image recognition* as the only data sets. Results of the post-hoc distribution free all-treatments multiple comparisons Wilcoxon-Nemenyi-McDonald-Thompson test for the remaining six data sets are

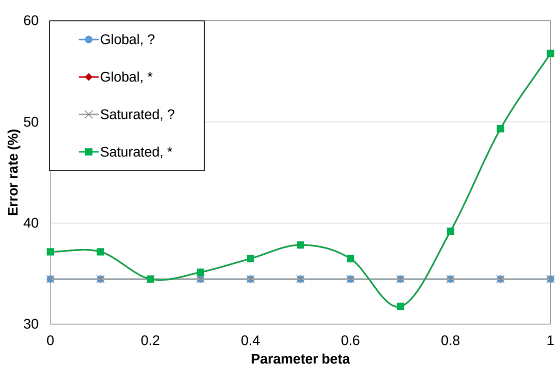

Figure 10. Error rate for the *Lymphography* data set

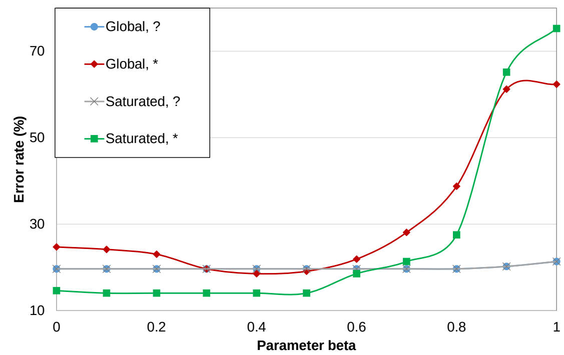

Figure 11. Error rate for the *Wine recognition* data set

TABLE II. RESULTS OF STATISTICAL ANALYSIS

Data set	Friedman test results (5% significance level)
Bankruptcy	(Global, *) is better than (Global, ?) (Saturated, *) is better than (Global, *) (Global, *) is better than (Saturated, ?) (Saturated, *) is better than (Saturated, ?)
Iris	(Global, *) is better than (Global, ?) (Saturated, *) is better than (Global, ?) (Global, *) is better than (Saturated, ?) (Saturated, *) is better than (Saturated, ?)
Lymphography	(Global, *) is better than (Global, ?) (Saturated, *) is better than (Global, ?) (Global, *) is better than (Saturated, ?) (Saturated, *) is better than (Saturated, ?)
Echocardiogram	(Global, ?) is better than (Global, *) (Saturated, ?) is better than (Global, *)
Hepatitis	(Global, ?) is better than(Saturated, *) (Saturated, ?) is better than (Saturated, *)
Wine recognition	(Global, *) is better than (Saturated, *)

presented in Table II. This table is divided into three parts. In the first part, for *Bankruptcy*, *Iris* and *Lymphography* data sets, data mining approaches based on "do not care" conditions are always better than approaches based on lost values. In the second part, for *Echocardiogram* and *Hepatitis*, it is the other way around. In the third part, for *Wine recognition*, the only conclusion is that for "do not care" conditions global probabilistic approximations are better than saturated ones. Due to the varying characteristics of the data, as follows from the experimental results, no particular combination was identified as the best for all situations.

V. Conclusions and Future Work

We compared four approaches for mining incomplete data sets, combining two interpretations of missing attribute values with two types of probabilistic approximations. Our criterion of quality was an error rate computed as a result of ten-fold cross-validation. As follows from our experiments, there were significant differences between the four approaches. However, the best approach, associated with the smallest error rate, depends on a specific data set. Thus, for a given data set, the base approach to mining must be selected by running experiments involved with all four approaches.

In future work, we will continue to study the experimental effects of saturation in an effort to form a theoretical basis for our ideas. Furthermore, these ideas will be extended to Maximal Consistent Blocks [21][22] to include the concepts of saturation with a thorough measurement of the experimental performance.

REFERENCES

[1] J. W. Grzymala-Busse and W. Ziarko, “Data mining based on rough sets,” in Data Mining: Opportunities and Challenges, J. Wang, Ed. Hershey, PA: Idea Group Publ., 2003, pp. 142–173.

[2] Z. Pawlak and A. Skowron, “Rough sets: Some extensions,” Information Sciences, vol. 177, 2007, pp. 28–40.

[3] Z. Pawlak, S. K. M. Wong, and W. Ziarko, “Rough sets: probabilistic versus deterministic approach,” International Journal of Man-Machine Studies, vol. 29, 1988, pp. 81–95.

[4] D. Ślęzak and W. Ziarko, “The investigation of the bayesian rough set model,” International Journal of Approximate Reasoning, vol. 40, 2005, pp. 81–91.

[5] S. K. M. Wong and W. Ziarko, “INFER—an adaptive decision support system based on the probabilistic approximate classification,” in Proceedings of the 6-th International Workshop on Expert Systems and their Applications, 1986, pp. 713–726.

[6] Y. Y. Yao, “Probabilistic rough set approximations,” International Journal of Approximate Reasoning, vol. 49, 2008, pp. 255–271.

[7] Y. Y. Yao and S. K. M. Wong, “A decision theoretic framework for approximate concepts,” International Journal of Man-Machine Studies, vol. 37, 1992, pp. 793–809.

[8] W. Ziarko, “Variable precision rough set model,” Journal of Computer and System Sciences, vol. 46, no. 1, 1993, pp. 39–59.

[9] ——, “Probabilistic approach to rough sets,” International Journal of Approximate Reasoning, vol. 49, 2008, pp. 272–284.

[10] J. W. Grzymala-Busse, “Generalized parameterized approximations,” in Proceedings of the 6-th International Conference on Rough Sets and Knowledge Technology, 2011, pp. 136–145.

[11] ——, “Rough set strategies to data with missing attribute values,” in Notes of the Workshop on Foundations and New Directions of Data Mining, in conjunction with the Third International Conference on Data Mining, 2003, pp. 56–63.

[12] P. G. Clark, C. Gao, J. W. Grzymala-Busse, T. Mroczek, and R. Niemiec, “A comparison of concept and global probabilistic approximations based on mining incomplete data,” in Proceedings of ICIST 2018, the International Conference on Information and Software Technologies, 2018, pp. 324–335.

[13] P. G. Clark and J. W. Grzymala-Busse, “Experiments on rule induction from incomplete data using three probabilistic approximations,” in Proceedings of the 2012 IEEE International Conference on Granular Computing, 2012, pp. 90–95.

[14] J. W. Grzymala-Busse, “LERS—a system for learning from examples based on rough sets,” in Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, R. Slowinski, Ed. Dordrecht, Boston, London: Kluwer Academic Publishers, 1992, pp. 3–18.

[15] J. W. Grzymala-Busse and W. Rzasa, “Local and global approximations for incomplete data,” in Proceedings of the Fifth International Conference on Rough Sets and Current Trends in Computing, 2006, pp. 244–253.

[16] ——, “Local and global approximations for incomplete data,” Transactions on Rough Sets, vol. 8, 2008, pp. 21–34.

[17] J. W. Grzymala-Busse, “MLEM2: A new algorithm for rule induction from imperfect data,” in Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2002, pp. 243–250.

[18] P. G. Clark and J. W. Grzymala-Busse, “Experiments on probabilistic approximations,” in Proceedings of the 2011 IEEE International Conference on Granular Computing, 2011, pp. 144–149.

[19] J. W. Grzymala-Busse, “A new version of the rule induction system LERS,” Fundamenta Informaticae, vol. 31, 1997, pp. 27–39.

[20] J. W. Grzymala-Busse, P. G. Clark, and M. Kuehnhausen, “Generalized probabilistic approximations of incomplete data,” International Journal of Approximate Reasoning, vol. 132, 2014, pp. 180–196.

[21] Y. Leung and D. Li, “Maximal consistent block technique for rule acquisition in incomplete information systems,” Information Sciences, vol. 153, 2003, pp. 85–106.

[22] P. G. Clark, C. Gao, J. W. Grzymala-Busse, and T. Mroczek, “Characteristic sets and generalized maximal consistent blocks in mining incomplete data,” in Proceedings of the International Joint Conference on Rough Sets, part 1, 2017, pp. 477–486.

Eager to Take Part in the Global AI Race? Better, Look for Traps Waiting There for You

Emanuel Diamant
VIDIA-mant
Kiriat Ono, Israel 5510801
e-mail: emanl.245@gmail.com

***Abstract*—The concept of Artificial Intelligence (AI) was introduced about 60 years ago. At the same time, Artificial Neural Networks (ANNs) were devised as a means for AI implementation. They were conceived as a collection of small interconnected computational units (called artificial neurons), which are supposed to imitate the biological neurons of the human brain. Essentially, ANNs were devised as data processing units – 60 years ago, all things in the world were considered computational. But today, the brain is thought as an information processing system. Therefore, biological neurons (and their artificial analogs) should be considered as information-processing units. This shift in the underlying assumptions is overlooked by almost all AI designers. I hope that a clarification of this issue will help AI designers to avoid dead ended trails and harmful pitfalls on their way to a successful and trustworthy AI realization.**

Keywords- Artificial Intelligence; Artificial Neural Network; information; information processing

I. INTRODUCTION

It is generally agreed and accepted that the human race is on the eve of a great and unknown time – the time of the AI era advent and victory. What once was considered a futuristic technology today begins to penetrate almost every aspect of our life.

The unprecedently rapid pace of AI infiltration is usually attributed to the latest Deep Learning Neural Nets (DLNN) explosion. Although Neural Nets have their roots a lot of years ago, DLNNs have rapidly became the best known technique in AI, yielding numerous state of the art results in a wide variety of domains such as speech recognition, image processing , language translation and as such tough and difficult tasks.

Wikipedia defines Deep learning as a set of machine learning algorithms that attempt to model high-level abstractions in data by using model architectures composed of multiple non-linear transformations. Deep learning software works by filtering data through a hierarchical, multilayered network of simulated neurons that are individually simple but can exhibit complex behavior when linked together, [1]. The might of Deep learning is now servicing AI research and development (R&D) supremacy.

Although in its long history AI's R&D has seen several ups and downs, the current surge of interest in AI is unmatched in its extent and enormity. The reason for this is the advent of the Big data era: Advances in sensor technologies, explosive growth of computing power, proliferation of broadband internet equipment, have led to an unprecedented flood of data inundating our surrounding. In such circumstances traditional practice of human-centered management of data volumes does not hold anymore and has to be delegated to a machine (a computer as we usually call it today). It is self-evident that such a computer has to possess many human-like cognitive abilities, which underpin understanding, analysis, and interpretation of the incoming data streams. In short, has to possess AI abilities.

The urgent need for AI solutions apt to meet the growing flood of Big data has led to an unprecedented rise in AI R&D efforts undertaken today around the world.

II. THE AI RACE

As the opportunities of AI technology become more recognized and appreciated, more and more nation states begin to consider their own AI strategies. Tim Dutton, in [2], has summarized a package of such national programs announced in the past two years.

In 2018, 25 European countries have joined up to make sure that the AI revolution doesn't leave them aside. The EU Commission has adopted a document that includes a commitment to increase the EU's investment in AI from €500 million in 2017 to €1.5 billion by the end of 2020, and the creation of the European AI Alliance, [2]. President Emmanuel Macron unveiled France's €1.5 billion plan to turn his country into a world leader for AI research and innovation. The U.K. Parliament has urged its government to draw up a policy to help the country become one of the world's AI leaders. Germany, too, has its grand ambitions in AI, [3], [4].

All these announcements are inspired by the idea that Europe must catch up with the United States and China in an AI arms race. Although the leadership of the United States and China is worldwide acknowledged and appreciated, the real state of affairs in AI R&D in US and China remains blurred and ambiguous.

The accepted (in the USA) free-market-oriented approach results in an entangled combination of classified research, public contracts from development organizations (like DARPA), and partnerships with the private sector, [4]. The extent of government funding in each of the sectors is blurred and unclear: In its unclassified 2017 budget, Pentagon reported on spending approximately $7.4 billion on

 ISBN: 978-1-61208-723-8

AI and the supporting fields, [3]. Billions more are invested in classified R&D, but how large are the figures is unknown. DARPA alone pledged to invest $2 billion in AI over the next five years through its AI Next initiative [5].

Similarly blurred and fuzzy are China's intentions in the AI. Only the budget figures are much more higher: The Chinese government pledges to build a US$150 billion AI industry by 2030, [4].

It is worth to be mentioned that human brain research, which is closely related and associated with AI's R&D, exhibits remarkably lower levels of government support and funding: The USA BRAIN Initiative is provided with $1 billion budget for ten years of project duration. European Human Brain Project is funded with 1.3 billion euro for the same 10 years long research span.

III. UNEXPECTED HURDLES

Despite the hype and the fascination that naturally surrounds AI undertakings, the real state of affairs is far from being satisfactory – despite of a long history of use and exploration, the operational principles of the DLNNs remain unclear and ambiguous. There is still little insight into their internal operations, and a lack of knowledge on why and how they achieve their impressive performances. From a scientific point of view, this is entirely unsatisfactory. Without a clear understanding of how and why the NNs work, the development of better models inevitably reduces to trial-and-error experimentation [6].

For that reason, the issue of explainability, interpretability, and transparency of ANNs applications has become a subject of hot public discourse and is repeatedly raised at many AI design related conferences, forums and gatherings, [7], [8], [9], [10].

An interesting turn in this discussion has been inspired by the recent announcement of Ali Rahimi (and his friends) that Machine Learning (ML) and AI "have slipped out of the bounds of science and engineering into alchemy", [11]. At the 2017 NIPS conference, Ali Rahimi declared that AI researchers do not know why some algorithms work and others don't. Relying on a trial and error strategy, AI algorithms have become a form of "alchemy." [12].

IV. OVERLOOKED REASONS

There is a common agreement among the participants of the public discussions about the measures that should to be taken in order to reach more "user friendly" and explainable ML/ANN constructions [13]. However, what is surprising in these recommendations is that, (agreeing with the statement that better theoretical understanding of the ANNs functional principles is compulsory for further successful development of ML/AI applications), no single attempt is undertaken to elaborate such theoretical principles. Therefore, I consider as my duty to try and to develop these overlooked theoretical grounds.

As to me, contemporary ANN designers are ignorant and unaware about the difference and the inconsistency between artificial and natural biologic neurons. While artificial neurons are intentionally designed to be **data processing** computational units, the natural biologic neurons are evolutionary developed to deal with and carry out **information processing**.

60 years ago, the world was at the dawn of a computing era, and the human brain was regarded as a computing device. Intelligence was considered as a human trait and as a product of human brain activity. Therefore, Intelligence was considered to be a computable function, and Computational Intelligence has evolved as a respected field of human-related computer science.

To facilitate human brain functioning studies, Artificial Neural Networks (ANNs) were conceived as a set of tightly interconnected simple computational units resembling the network of biological neurons in human brain. As it was just mentioned, ANNs units' functionality indisputably implies data processing ability. That is, **ANNs were designed and used as data processing (computational) devices**.

However, in recent decades, a significant paradigm shift in brain functionality understanding has occurred. Now, **human brain is considered as an information processing apparatus**. Intelligence, thus, should be seen as an **information processing outcome**. Subsequently, brain neurons should be seen as information processing units (not data processing instruments, as it was before). At the same time, Intelligence is no more an exceptionally human trait – it is a feature common to all living beings, with and without brains or nervous systems. (Another argument in favor of an information processing principle).

Contemporary AI designers are not aware about these revolutionary transforms. The difference between data and information is not seen and not understood properly by the contemporary research community and therefore the subject is overlooked and neglected. Often, the notions of data and information are used interchangeably, most often being mixed and blended. That is the legacy of Shannon's Information Theory or, let us be more accurate, the way of how people understand and use Shannon's Theory. I do not think it would be wise to deepen our understanding of these inconsistencies. Rather, I think, it will be more appropriate to understand what actually information is.

V. LET ME EXPLAIN

There is a widespread conviction that a consensus definition of Information does not exist. I do not agree with this. On several occasions, I have already published my opinion on the subject [14], [15], [16]. This time, with all fitting excuses, I would like to repeat some parts of these earlier publications in order to preserve consistency of this discussion.

Contrary to the widespread use of Shannon's Information Theory, my research relies on Kolmogorov's ideas on information, [17]. According to Kolmogorov, a not random binary string (called a separate finite object) can be represented by a compressed description of it (produced by a computer program in an algorithmic fashion) "in such a way that from the description, the original message can be completely reconstructed" [18]. The compressed description of a binary object has been dubbed as "algorithmic information" and its quantitative measure (the length of the descriptive program) has been dubbed as the description "Complexity".

Taking Kolmogorov's insights as a starting point, I have developed my own definition of information that can be articulated in the following way: "**Information is a linguistic description of structures observable in a given data set**".

To make the scrutiny into this definition more palpable I propose to consider a digital image as a given data set. A digital image is a two-dimensional set of data elements called picture elements or pixels. In an image, pixels are distributed not randomly, but, due to the similarity in their physical properties, they are naturally grouped into some clusters or clumps. I propose to call these clusters **primary or physical data structures**.

In the eyes of an external observer, the primary data structures are further arranged into more larger and complex agglomerations, which I propose to call **secondary data structures**. These secondary structures reflect human observer's view on the grouping of primary data structures, and therefore they could be called **meaningful or semantic data structures**. While formation of primary (physical) data structures is guided by objective (natural, physical) properties of the data, the subsequent formation of secondary (semantic) data structures is a subjective process guided by human conventions and habits.

As it was said, **Description of structures observable in a data set should be called "Information".** In this regard, two types of information must be distinguished – **Physical Information and Semantic Information**. They are both language-based descriptions; however, physical information can be described with a variety of languages (recall that mathematics is also a language), while semantic information can be described only by means of natural human language. (More details on the subject could be find in [19]).

Those, who will go and look in [19], would discover that every information description is a top-down-evolving coarse-to-fine hierarchy of descriptions representing various levels of description complexity (various levels of description details). Physical information hierarchy is located at the lowest level of the semantic hierarchy. The process of sensor data interpretation is reified as a process of physical information extraction from the input data, followed by an attempt to associate this physical information (about the input data) with physical information already retained at the lowest level of the semantic hierarchy.

If such an association is attained, the input physical information becomes related (via the physical information retained in the system) with a relevant linguistic term, with a word that places the physical information in the context of a phrase that provides the semantic interpretation of it (see also the block diagram in [14]). In such a way, the input physical data object becomes named with an appropriate linguistic label and framed into a suitable linguistic phrase (and further – in a story, a tale, a narrative), which provides the desired meaning for the input physical information. (Again, more details can be found on the website [19]).

VI. WHAT FOLLOWS FROM ALL THIS

Equipped with the "In this paper introduced" (ITPI) definition of Information we can now try to analyze what is going on in a typical ANN-based AI installation (it goes without further saying that DLNN, CNN, RNN, and all other NN versions are simply revisions of the same basic NN layout). Usually, as it is always proudly declared, after an act of training the NN transforms autonomously the data at its input into a semantic label or a semantic statement at its output. Terms "physical information" and "semantic information" are not known to the ANN design community. As a result, any information processing goal has not been foreseen and is not fulfilled in the course of ANN data processing activity.

It is unnecessary to remind the readers that according to ITPI theory, direct transition from primary (physical) information description to secondary (semantic) information description is not possible (does not exist). It is unreasonable, from a theoretical point of view. The grouping of primary data structures into secondary data structures is entirely an observer's privilege and prerogative. The rules of secondary structures organization are subjective, that is, they are solely observer's habits and concerns. Intelligence displaying systems (natural or artificial) acquire them as a grant, as a gift, a shared common agreement (a common knowledge base). Afterwards they all are preserved (conserved) in the system's memory. ANN training phase can be seen as a hint of this processing tread. But in ANN practice, which is devoid of any information processing intents, such things are even do not appear.

The term "information" is frequently seen in ANN R&D texts. But it is used in the sense of Shannon's Information Theory. That is, the term information is used as a substitution swap for data. Shannon's theory does not define what information is. It introduces and exploits a notion of "information measure". Which is equivalent to the "entropy" of a data set. Shannon himself has warned not to mix up the terms (information and information measure). But who cares?

An important outcome of the ITPI theory is the referential mode of information processing, which is unknown (to AI systems designers) and therefore is not addressed in any ANN-based AI designs. The long list of missed AI related properties that must be satisfied in an information processing AI system is not even mentioned here because of the limited article space.

VII. CONCLUSIONS

The purpose of this paper is not to reject or deny AI's virtues or to turn down its achievements. The purpose of this paper is to convince interested people that any further success in AI R&D requires immediate rejection of the ANN approach, which today is the main workhorse of AI modeling.

The notions of AI and ANN were introduced about 60 years ago. It was at the down of the computer era, when, according to the spirit of that time, every fact and every action were considered as a computational expression. All scientific fields were considered "computational". The brain and its functions were considered computational (recall "Computational Intelligence", which is alive and prosperous even in these days). Brain neurons, accordingly, were regarded as computational units. It is worth to mention that "computational" always implies "busy with data processing".

The beginning of this century was denoted by a rapid paradigm shift in scientific thinking – from data processing predisposition to information processing preference. Almost immediately, Computational Biology has become converted to Cognitive Biology, Computational Neuroscience to Cognitive Neuroscience, Computer Vision to Cognitive Vision. Almost every conventional Computational science was converted to an associated Cognitive science. (Here "Cognitive" implies "Information processing" aptness and ability). Unfortunately, ANN-based approach to AI modeling has not been affected by this general paradigm shifting.

At the same time, the information processing paradigm adopted by the whole spectrum of biological sciences has led to significant advancement in understanding the nature and special virtues of Intelligence. Intelligence is not anymore a uniquely human attribute. It is an evolutionary developed feature present in every living being, from bacteria to humans. Intelligence – as an ability to process information – is present now (evident, observable, discernable) at all levels of living beings spectrum. It does not require Neural Nets complexity, it is ruled by the same principles of information processing at all levels of living beings presence, [20]. Therefore, such things as Narrow Intelligence or General Intelligence for this constellation do not exist. You can guess that Artificial Intelligence and Natural Intelligence (in such a case and with all their diversities) differ only by the level of information complexity that is supposed to be processed or is actually processed in the system.

I hope that clarification of these issues will help AI designers to avoid dead ended trails and harmful pitfalls on their way to a successful and trustworthy AI realization.

References

[1] "Deep learning", From Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Deep_learning

[2] T. Dutton, "An Overview of National AI Strategies", https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

[3] T. Rabesandratana, "With €1.5 billion for artificial intelligence research, Europe pins hopes on ethics", Apr. 25, 2018, http://www.sciencemag.org/news/2018/04/15-billion-artificial-intelligence-research-europe-pins-hopes-ethics

[4] J. K. Delaney, "France, China, and the EU All Have an AI Strategy. Shouldn't the US?", May 20, 2018. https://www.wired.com/story/the-us-needs-an-ai-strategy/?mbid=social_twitter

[5] VentureBeat Events, "DARPA is betting $2 billion on your next AI innovation", October 8, 2018. https://venturebeat.com/2018/10/08/darpas-betting-2b-on-your-next-ai-innovation/

[6] M. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks", http://arxiv.org/abs/1311.2901

[7] R. Shwartz-Ziv and N. Tishby, "Opening the Black Box of Deep Neural Networks via Information", https://arxiv.org/abs/1703.00810

[8] B. Mittelstadt, *et al*, "Explaining Explanations in AI", https://arxiv.org/abs/1811.01439

[9] T. Xu, *et al*, "Deeper Interpretability of Deep Networks", https://arxiv.org/abs/1811.07807

[10] Z. C. Lipton, "The Mythos of Model Interpretability", https://arxiv.org/abs/1606.03490

[11] R. Letzter, "Google AI Expert: Machine Learning Is No Better Than Alchemy", May 7, 2018, https://www.livescience.com/62495-rahimi-machine-learning-ai-alchemy.html

[12] A. Rahimi and B. Recht, "An Addendum to Alchemy", Dec 11, 2017 http://www.argmin.net/2017/12/11/alchemy-addendum/

[13] Z. Lipton and J. Steinhardt, "Troubling Trends in Machine Learning Scholarship", https://arxiv.org/abs/1807.03341

[14] E. Diamant, "Unveiling the mystery of visual information processing in human brain", Brain Research, vol. 1225, pp. 171-178, 15 Aug. 2008, https://arxiv.org/abs/0807.0337

[15] E. Diamant, "Shannon's definition of information is obsolete and inadequate. It is time to embrace Kolmogorov's insights on the matter", Conference Paper, November 2016. https://www.researchgate.net/publication/311223095

[16] E. Diamant, "Advances in Artificial Intelligence: Are you sure, we are on the right track?" https://www.researchgate.net/publication/272478913

[17] A. N. Kolmogorov, "Three approaches to the quantitative definition of information", Problems of Information and Transmission, Vol. 1, No. 1, pp. 1-7, 1965. http://alexander.shen.free.fr/library/Kolmogorov65_Three-Approaches-to-Information.pdf

[18] P. Grunwald and P. Vitanyi, "Algorithmic Information Theory", 2008. http://arxiv.org/pdf/0809.2754.pdf

[19] E. Diamant, "Brain, Vision, Robotics and Artificial Intelligence". http://www.vidia-mant.info

[20] E. Diamant, "Designing Artificial Cognitive Architectures: Brain Inspired or Biologically Inspired?" https://www.researchgate.net/publication/329582475

Improving Measurement by Addressing the Sensor Fault Isolation Problem in Control System

Thanh Nga Thai
School of Electrical Engineering
Ha Noi University of Science and Technology
Ha Noi, Viet Nam
thaithanhnga@gmail.com

***Abstract*—In many processes, an intelligent system using a development simulation effective for the diagnosis system depends on the sensor fault isolation problem. This work presents a review of using advanced measurement technology relevant to the sensor fault detection field through the implementation of process instrumentations to improve safety and reliability in systems. The system's dynamics change in time as it is affected by multiple functions and nonlinear characteristic measurements. In such an environment, the controllers should adapt on the fly to ensure performance stability. An adaptable development method should be employed to improve measurement in control systems.**

Keywords-sensor fault types; nonlinear characteristic measurement; fault detection and isolation FDI; diagnosis; adaptable method.

I. INTRODUCTION

All real systems may encounter malfunctions or defects, which occur in sensors, actuators or the components of the system itself [1]. Abnormalities and component faults are increasingly studied as part of research meant to improve reliability and safety in industrial systems, especially in automatic control systems. Passive faults and active faults based on fault diagnosis can refer to the design and operation of measurement and control [2]. If a characteristic property or parameter of a system changes from the standard condition, the diagnosis system checks some predictable faults such as the blocking of an actuator or the loss of sensor or the disconnection of a system component. A diagnosis decision using previously acquired knowledge from a healthy system is usually made by using observation based methods or parameter estimation techniques. However, some nonlinear characteristics in the system may not be exact or observed in the estimating process. There is a need to require an effective adaptive development method for sensor fault diagnosis. Improving fault measurement via sensor isolation is the main way to achieve an intelligent environment for a diagnosis system as compared to other conventional fault diagnosis procedures that only work with fault tolerant system.

In a dynamic system, the intrinsic parameters of a system will modify the system's dynamic. For the measurement system, inaccurate values of physical variables in the sensor and the environment factors, such as poor or bad environmental conditions, affect to stable configuration structure or cause damage to the stability of the system. For example the bias, the drift or the freezing at low temperature of sensors can lead to seriously unstable configuration or can influence the system structure. These inner unstable characteristics are described by using the function of phenomenological analysis or general basis function, which are observed behaviors of the system. The change in the inputs affects the output and it can be the reason of some otherwise unobserved states in control system.

The step of detecting and measuring these nonlinear characteristics in a system, such as factors for structural damage, is critical towards measuring the quality of a system. Because the characterization of each component or each part in a system is completely different, such as for example random fault-variation of supply voltage, temperature changes or inaccurate instrument reading, the automatic recovery system needs some specific control techniques to improve safety and reliability. The methods for fault detection of these nonlinear characteristics have to operate under the constraint that there is potential inaccuracy in the model, or the model used may not precisely match the situation. The following optimization techniques in [4][7] present the identification of possible improvements of the onboard systems[6].

Improving measurements by sensor fault isolation in a control system to avoid accidents and optimization route is an efficient desired fault detection method to improve a diagnosis system. This work gives an overview of the various ways of measurements to better analyze and design for adaptable development methods. This also combines with modern control theory, provides a development simulation of nonlinear characteristic as a virtual tool for understanding the effects of input noise and time delay on the system operation and tolerance. The diagnosis system then identifies the relationship between fault detection and the allowable delay operations while ensuring a stable system performance. These analysis improvements based on the model free nonlinear control, a well defined controllability observer, provide more for the mathematical basis by multifunction faults in section II. Section III is the application adaptive method for some sensor fault groups. The conclusion will approach to the periodic diagnosis role in the complex system using this method.

 ISBN: 978-1-61208-723-8

II. Main problem

There are two main aspects to improving detection of when a fault happened in the control system. The first is the sensor fault type which is either open loop or close loop. Understanding the process control loop in the field of instrumentation by open loop- manual mode and close loop- automatic mode determines the solution and decision for the parameter faults involved in the process. The second is the method based on modelling to reflect the control system.

A. Maintaining the sensor fault types

There are different faults in an automatic control system such as sensor faults, actuator faults, component faults or control unit faults. At the same time, there are unknown inputs acting on the system causing disturbances generating system noise in a dynamic system. This paper gives some idea of sensor faults as occurring in the electronic part of an automatic control system working with mission critical aspects and operating in a well defined and protected environment.

Many measurement errors are caused by defects in sensors such as shortcut, offset, bias, power breakdown, sticking, scaling error, hysteresis. Approaching the fault free functioning in the electronic part for measured outputs is discussed as a problem to perfect performance. The open loop system is decomposed into three parts: sensors, actuator and plant dynamics. External system disturbances or process noise affect the dynamic of the plant and FDI needs robust to deal with these uncertainties and remain fault sensitive for diagnosis. Figure 1 presents some type of faults happening in an open loop.

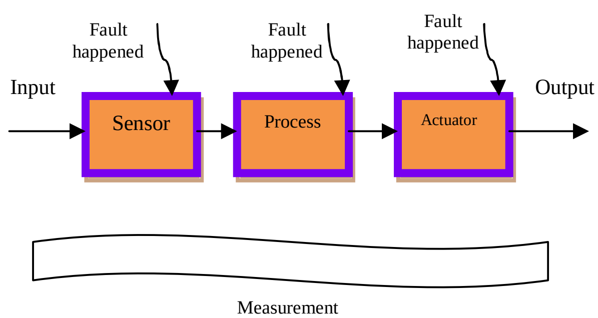

Figure 1. Fault happened in open loop system.

Starting at the input with the fault happening in sensor and then in each part of components in open loop, the diagnosis system uses dynamic switches to check the connection between them for the maintenance mode. Measurement of these faults as active faults or passive faults depends on using the system model. Based on the modeling approach as model based or model free, different FDI techniques are used for the system dynamic. Some technical details are needed in order to deal with the nonlinear characteristic while stabilizing a fault. The sensor fault types are the small and large bias, the slow and fast sensor drift, the freezing at the current sensor location or the oscillatory failure, increased sensor noise, non return to zero, the malfunction as short circuits, heating or erroneous signal conversion from analog to digital, aging, etc [8]. The basic faults detected in sensors are divided in passive faults and active faults. Passive faults are such as small and large bias, the slow and fast sensor drift, non return to zero. Active faults are two groups. The first group is based on the freezing at the current sensor location, oscillatory failure, increasing sensor noise. The second group considers the cause of sensor fault and external influences as the malfunction itself, the malfunction of the sensor heating or the aging for longtime, the erroneous signal conversion from the analog to digital or the influences as temperature or dust. Nonlinear characteristic sensor fault approach in this way is critical to the quality of measuring production processing capability. First, to determine the fault detection method in the operation system, the development adaptation measure is referred. Second, an intelligent control needs a development simulation to find complementary characteristic faults like type of faults, location fault or duration fault for diagnosis system. The reason is these characteristics in each part of the components are completely different from random faults to specific faults. A method implies nonlinear characterization fault detection for operating constraints in process system can be suggested. Based on the measurements obtained at the current and the past operating point, the position of the next operating points with respect to the previous is found by solving a nonlinear optimization problem in its application to the stabilization as a set of points. Regarding the measurement and control methods, this knowledge about the optimization solution is on the use of models for dynamic system. Extending in the system perspective, the plant model mismatch presented the uncertain or the unknown disturbances estimation becomes a solution model nonlinear compensation. This called characteristic structure is intrinsic to a corporation with the potential inaccuracy unobserved.

Another approach is an advance control method for nonlinear optimization based on the control law. It provides an overview of distinction between measure and control system as operating point and set point regulation. In this action, when the computer control is either reliable or prone to errors, the controller software handles using entirely different error and accommodation mechanisms. The stable controller, corresponding to each situation is designed off-line using model predictive control method with the fixed parameter sent to online controller redesign- design control system.

Three mains fault types in this approach are abnormal fault, the regular and the random fault or the control law as normal law, alternate law and direct law. These are called the characteristic structure control outside by law.

A structure of the definition control laws requires the entire process envelope. Since each automation operation of a dynamic system is characterized by the establishment of the operating points, the control system is interested in the dynamic behavior related to the observations. In figure 2, fault happened in a close loop related with the control law programming.

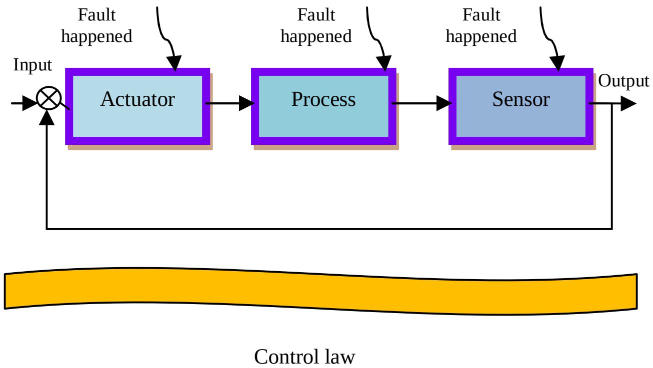

Figure 2. Fault happened in close loop system.

The requirement to reduce isolation faults based on the control method is still the subject of ongoing research. From the point of view of the control system, feedback is used to deal with active faults and adjustments are made to bring the system back in order. In fault detection, the first step is the stabilization assurance. The second step is the operating of threshold and false. The threshold includes the sensitivity to load, speed or noise, repeatability. Isolation fault consist of threshold, false, discrimination and severity. To achieve the best performance in a system, a predicted condition and remaining life need a suitable threshold. Modern control method approaches are to have the output which has characteristic input for detection as the prediction, and dynamic function as dysfunction. Control law provides set points, which are based on the best performance for the robust control system.

Depending on the fault characterization on sensors and control rules, the diagnosis system decides which is the suitable method based on the model based method and model free method. There are many research articles such as signal based fault detection and integration with the model based method fault detection to achieve full control system functionalities [1][4][8]. These can be considered as the combination between the method model based and model free. Each opinion makes some choices and generated somewhat different results. There are a few articles comparing the differences between them and also the application for each particular fault.

B. Model based and model free method

Considering the traditional model approaches to system engineering, we have traditionally relied on mathematic models. There are two main approaches: model based approaches and model free approaches. In the system dynamic, methods are model based which is done according to the structure of the model and the existing input-output relationships. Model based fault diagnosis is a fundamental role of fault detection for small faults with determining time, size and cause of fault in a dynamic system operation. Model free is used in real time measurements and process history when model based are unavailable or not useful [3]. The majority of FDI techniques have been designed assuming linearity while the majority of physical processes are strongly nonlinear, so the accurate models of dynamic system are nonlinear. In the model based method, analytical redundancy means to exploit mathematical relations between measured or estimated variables to detect possible dys-functions and should be understood as knowledge based dynamic model [3]. Residual generation uses the model of the system, in which the control inputs sent to the actuators and the system outputs, as measured by the injected sensors predicting the behavior of the system or part of it, comparing between the prediction actions and the actual action. Fault detection computes quantitative indexes of fault presence and the residual. The associated algorithms associate functional and non-functional properties of models to performance benchmarking and optimization. Modern control theory approaches are not only model based but also model free in which the output has a characteristic prediction of the input for fault detection. Model free is used when there is no explicit dynamical model of the control system, and it actually uses the redundancy and correlations of the data in a hidden manner. It exploits measurements acquired in real time, or available in a previously constructed database, such as using the behavioral signal model. The idea of this work is the residual generation which indices the fault presences, fault isolation problem changes into an appearance sensor fault. It integrates a characteristic prediction function in the passive fault situation via hidden data related with time delay and noise operation in the system. These sensors work as a real time measurement and process history data independent of the control system. The meaning is that a control based mechanism replaces measurements from a sensor in passive fault to ensure stable performance as per an adaptive model.

III. APPLICATION ADAPTABLE METHOD

Some methods approach to diagnosing fault system are divided into four groups. Group 1 and group 2 are following an obtained model method. Group 3 is the connection from the choice of group 1 and 2 to achieve suitable characterization threshold for structure state. Group 4 shows the combination for adaptable method as automatic tool for the supervision.

A. Application model based for sensor in group 1

Fault detection relies on the analytical redundancy principle using an accurate model of the system to mimic the real process behavior. If a fault occurs, the residual signal can be used to diagnose and isolate the malfunction such as a set point. In the case of noisy measurement, the identification technique gives the variances of the input-output noise signals. Fault detection by multi set points is based on the history data for the ability of time delay and noise operation as full control functionalities system. To analyze the

 ISBN: 978-1-61208-723-8

diagnosis effectiveness of the FDI system in the presence of abrupt changes or drift in measurements, fault modeled by step or ramp functions have been generated, indicate that the minimal detectable faults on the various process is a parameter of interest for industrial diagnosis applications. Comparisons between values of working parameters obtained in the simulation and measurement can make it possible to predict and design. Small and large bias, the slow and fast sensor drift, non return to zero are caused by the freezing at the current sensor location, and is considered as the malfunction of the sensor heating or the aging for longtime. Multi set points as multi malfunction sensor propose the technical layers are applied model based method [2][3][5].

B. Application model free for sensor in group 2

The dynamic nonlinear model has been developed by dividing the dynamic operation of the machine into elementary model by means thermodynamic and mechanic links. Reconfigure system from the different methods by the predicted suitable method. When uncertainties are caused by modeling errors, linearization errors, parameter variations, etc, such a disturbance decoupling approach cannot be directly applied, an approach suggests an exploiting estimated distribution matrices. From the group 1, group 2 continues with the adaptive method to observe the behavior in the system as model free in the operating system [1][2][5].

C. Equation relationship between them in structure state equation

Characterization applies to suitable threshold prediction as nonlinear characteristic to improve the safety and reliability in system from the structure state equations. The full nonlinear control applied to the positioning and orientation problem is assessed and adapted to the expected performances. Following the free model based, structure state equation change to the positioning and orientation relationship equations [1][2][8]. This promising method to solve the FDI problem consists of using an accurate model of the system to mimic the real process behavior by covering the difference between real system and model behaviors rather than the corporation with the potential inaccuracy of the model or model mismatch in parameter identifications.

D. Adaptive method for improving measurement

The way to ensure stability performance in the system becomes a step in the control based mechanism [5][9], related to the time delay and noise operations, the dysfunction for the stabilization and observation is applied to sensor fault isolation problem. Time delay and noise operation become the condition environment as a standard to adaptive method for improving measurement. The application includes automatic supervision of close loop operation as early as possible.

IV. CONCLUSION

An adaptable development method following this way can improve the measurement in control system. It shows that sensor fault isolation in control systems achieves the robust system by controlling passive faults and active faults. The material in sensor problem relates to the design of control structure allowing established autonomous position and orientation characterization. Coating material combines specific requirements towards the best available solution. Selecting the variety and technical guidance applied to the sensor fault isolation problem achieves the required standard structure. Future works will study the detail of the relationship between switch dynamics with the role sensor in the tolerance fault for the actuators. Finally, this justifies the periodic diagnosis point in complex dynamic system by securing the self-control problem.

REFERENCES

[1] S.Simani, C. Fantuzzi and R. J. Patton ''Model based fault diagnosis in dynamic systems using identification techniques'', Springer, 2003, pp. 2-28.

[2] R. Isermann, ''Fault diagnosis systems- An introduction from fault detection to fault tolerance'', Springer, 2006, pp. 15-25.

[3] L. Fernandez and J. Luis, ''Fault detection for nonlinear systems: a new approche'', thesis unpublished, pp. 8-20.

[4] P. Goupil, ''Fault diagnosis and fault tolerant control and guidance for aerospace vehicles'', Springer, 2014, pp. 2-18.

[5] T.N. Thai, '' Characteristic matrix faults for flight control system'', ICMMS, 2014, pp. 1-4.

[6] S. Chatty, S. Convesy, H. Gaspard and J.L. Vinot, ''The accident of flight 447 Rio-Paris: a case study for HCI research'', Hal, 2009, pp1-4.

[7] O. T. Pleter, C.E. Constantinescu and B.I.Jakab, ''Reconstructing the Malaysian 370 Flight trajectory by optimal search'', Journal of navigation, 2016, pp. 4-8.

[8] D. Ossmann, H.D. Joos and P.Goupil ''Enhanced sensor monitoring to maintain optimal aircraft handling in case of faults'', Journal of guidance control and dynamics, 2017, pp. 1-5.

[9] T.D. Nguyen, '' A conceptual approach to the problem of action angle variables'', ARMA, August 2018, Volume 229, Issue 2, pp. 789-794.

An Industry 4.0 Self Description Information Model for Software Components contained in the Administration Shell

Luís Neto*†§, Gil Gonçalves*†§, Pedro Torres‡¶, Rogério Dionísio‡¶, Sérgio Malhão‡‖
*SYSTEC (Research Center for Systems & Technologies)
†FEUP, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
‡Escola Superior de Tecnologia, Instituto Politécnico de Castelo Branco
§Email: {lcneto,gil}@fe.up.pt
¶Email: {pedrotorres,rdionisio}@ipcb.pt
‖Email: smalhao@ipcbcampus.pt

***Abstract*—Industry 4.0 is the movement towards a fourth industrial revolution that will consist in the digitization and integration of all value chain. In Europe, this movement is led by the German *RAMI 4.0* (Reference Architecture for Industry 4.0) proposal, which is attracting a lot of attention from industry, academia and other practitioners. Under the RAMI 4.0 scope there is an Administration Shell proposal to abstract physical and logical assets in a standardized way. Once abstracted, assets become Industry 4.0 Components and can be fully integrated in the Cyber Physical Production System or value chain. This work focuses on the utilization of software components within the Administration Shell. There is a necessity to represent software components and their relation to industrial asset. Therefore, control and monitoring applications involving software components and other assets can be represented in compliance with the I4.0 Component Model. To address this necessity the Smart Object Self Description information model is proposed and applied to a real case study scenario.**

***Keywords*–Information Model; Component Based Software Engineering; Smart Component; Administration Shell; Industry 4.0.**

I. Introduction

Industry 4.0 (I4.0) is the movement that aims to transform the traditional factory into a smart factory. There is a hype around this movement fueled by great expectations in the way industry will transform the value chain, business models and economy [1].

In terms of path to effectively create the smart factory, there are several of models with different specificities proposed by different countries [2]. In Europe, RAMI 4.0 [3], proposed by several German organizations gathered under Platform Industrie 4.0, seems to join the bigger consensus [4]. RAMI 4.0 most widespread concept is a tree dimensional map that combines: 1) the hierarchy of Industry 4.0 components, according to ISA 95 (International Society of Automation 95); 2) the product life cycle and it value chain, from conception to disposal; 3) the factory architecture perspectives, from assets to the whole organization and business processes. The main objective of this model is to create a clear understatement of all participants within the Industry 4.0 and across the value chain [1].

One of the core technological aspects for Industry 4.0 is *The Industrie 4.0 Component model* [5]. The components model for Industry 4.0 was developed by the participants of Platform Industrie 4.0 to help equipment producers and system integrators to create standardized I4.0 compatible hardware and software components [1]. This component model specifies that each asset (logical or physical) must be encapsulated by a standardized digital container – the *Adminsitration Shell* – that will enable description, collaboration and communication among all I4.0 Components.

A. Administration Shell

The *Administration Shell* (AS), acts as an interface connecting all physical and logical assets to the I4.0 compliant network, therefore creating an I4.0 compatible Cyber Physical Production System (CPPS). [6] is a proposal for the general structure of the AS, as proposed by Plattform Industrie 4.0, and therefore it was a reference to this work. The things abstracted by an AS are diverse and some are manufacturer dependent, so the AS maintains an internal interface specific for each asset, as in Figure 1. The AS has also an external interface, which is responsible for communication with the I4.0 network. Another peculiarity of the AS is that it can represent passive and active assets. One example of a passive asset is a purely analogical machine or tool which might be important to represent digitally. On the other hand, an example of an active asset is a complex machine incorporating digital control units capable of processing and communication. The asset itself can be composed of other assets as is the case of a machine whose sub-systems can be represented individually, it is also the case of a production line or even the entire factory. The representation of an asset, once abstracted by an AS, is also commonly called the *digital twin*.

Once an asset is encapsulated by an AS it becomes an I4.0 Component (Figure 1) and can participate in the I4.0 compatible CPPS, which is formed by other I4.0 Components. A descriptive diagram explaining the process of encapsulation and the main advantages of the I4.0 Component Model are described in detail in [1] [5].

B. PRODUTECH-SIF Project

The work described in this paper was performed in the scope of PRODUTECH-SIF (Solutions for the Industry of the Future) project [7]. This is a Portuguese National initiative with a research agenda that comprises a set of R&D activities in key domains with the objective of enabling the digital transformation of the Portuguese industry. One of the base activities is the study and implementation of base technologies to create CPPS, whose first results are described in this paper. As already discussed, the AS is a base technology to

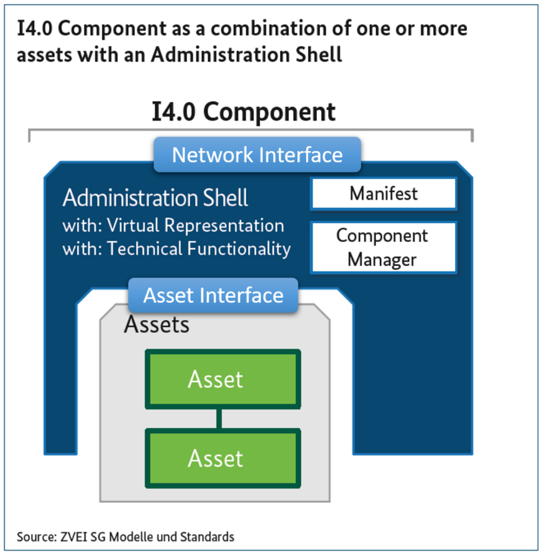

Figure 1. Representation of the Administration Shell application to an Asset to form an I4.0 Component [6].

create CPPS. As this concept is proposed under the scope of RAMI 4.0, the AS was chosen to convert the participating companies assets in I4.0 Components. Therefore, preparing the Portuguese industry for the upcoming establishment of I4.0 based technologies.

The methodology followed started with requirements assessed by means of inquiries to all companies participating in the project. Thereafter, a company responsible to host the pilot demonstration of the project was visited to prepare a complete case study, upon which all developments presented in this paper were based. All hardware requirements are tackled by a SmartBox described in Section II. The AS of the project is realized by the SmartObject concept, which is described in Section III. The case study, based on the pilot demonstration of the project, is presented in Section IV. A series of problems and the motivation for this work are presented in Section V. The main contribution of this work is a proposal for an information model used for self-description and (re)configuration of SmartObject's, presented in Section VI. The paper finishes with some conclusions and future work in Section VII.

II. SmartBox

The *SmartBox* is a smart hardware developed over the National Instruments (NI) cRIO-9040 platform that runs the *SmartObject* and enables the remote connectivity with machines in the shop floor. NI cRIO is a Programmable Automation Controller (PAC) that allows extremely high speed measurements and also allows to perform software-defined hardware through an internal FPGA. One important component is the modular Real-Time PAC platform that enables flexible data acquisition and actuation based on hardware modules that can be incorporated, including machine vision systems. In particular, the cRIO-9040 has a dual-core 1.3 GHz processor, 2 GB DRAM and 4 GB Storage with 4 slots for different I/O modules.

In the scope of the PRODUTECH-SIF project, the *SmartBox* was programmed for Device-to-Device (D2D) connections supported by the OPC-UA protocol and Device-to-Server (D2S) connections with the MQTT protocol. A *SmartBox* can be installed to manage one machine or a groups of machines. In both cases, the *SmartBox* acts as an OPC-UA server or a MQTT publisher of the IIoT system architecture, as an Edge-Node between sensors and actuators and the information systems. Figure 2 illustrates the *SmartBox* that communicates over EtherCAT with Remote I/Os for shop floor scenarios where data or command information flows between machine physically distant from each other. Currently installed on the *SmartBox*, there is multifunctional module (NI-9381) with digital and analogue I/O to receive data from digital sensors or analogue voltage values between 0V - 5V with medium to low resolutions needs. The same module is used for actuation. The *SmartBox* has also an AC differential input module (NI-9215) installed, with 4 channel of ± 10 V, 16 bits for acquisition of analogue signals such as acoustic or vibrations. Another available module (NI-9239) has a 24 bits ADC, and 4 channels. It is used for vibrations and small signals measurements produced by magneto-resistive sensors.

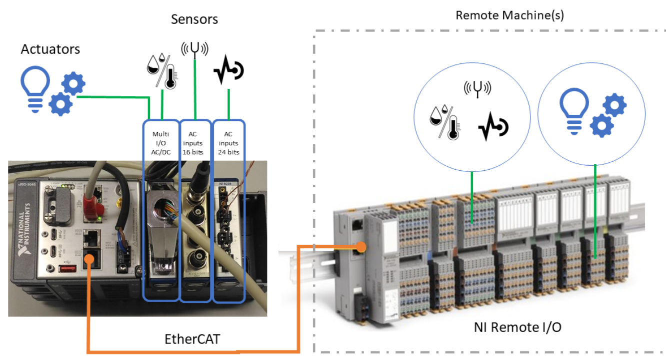

Figure 2. *SmartBox* with Remote I/Os for scenarios with multiple machines installation.

What distinguishes the *SmartBox* from other conventional controllers is the capacity to implement machine learning techniques. For example, it can be used to measure vibrations in induction motors and run classification methods that automatically detects and alerts for anomalies like possible misalignments of the rotors and worn out bearings and gears. Another possibility is the forecasting of temperature in specific machine areas based on the history and based on this, act in the machine or simply monitor the current state. These are only 2 examples already tested with the *SmartBox*, but another supervised, unsupervised or reinforced machine learning techniques can be implemented. The *SmartBox* can be reconfigured by adding or removing specific hardware modules, and each can be also reconfigured by software. This versatility is an asset for the implementation in different scenarios presented by the industrial partners of PRODUTECH-SIF project.

III. SmartObject

The SmartObject (SO) is an implementation of the *SmartComponent* concept [8], defined by Smart Component's community [9] [10]. In the scope of the PRODUTECH-SIF project, the SO implementation is being extended to consider the AS requirements defined in [6]. Therefore, the SO will act as the AS for encapsulating and converting the case study assets in *I4.0 Component's*. To make assets transparent to each other, i.e., capable of mutual cooperation and understatement, each SO will maintain and make available to the CPPS a *manifest* describing each asset. This paper focuses on the common data model created for that purpose, the *Smart Object Self Description (SOSD)*.

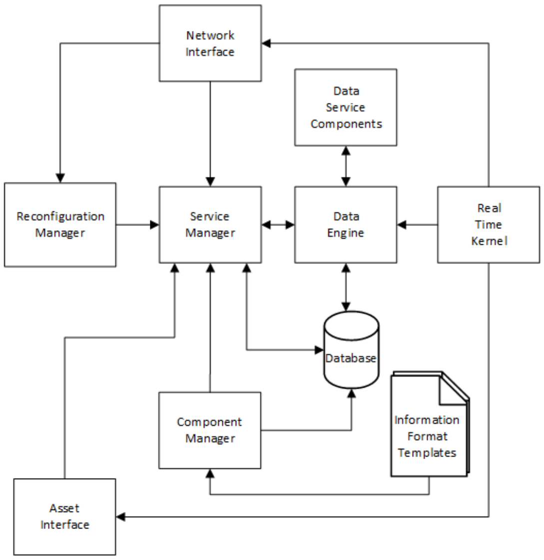

Figure 3. SmartObject Block Diagram.

In software terms, the SO is a component framework [11] constituted by a composition tool and a runtime environment. The diagram of Figure 3 shows the main components of the runtime environment, which will be deployed in each SmartBox. The composition tool allows to develop, maintain and deploy software components compatible with the runtime environment. During deployment an user can create a new composition or modify any one running in a certain SO. The deployment environment embeds a canvas that displays the composition and allows new components to be dragged in and interconnected by means of it's interfaces. This way, a control engineer doesn't need to know any specific details of software. All he sees is a set of black boxes whose functionality and interfaces are well documented and that can be used to build control or monitoring applications. Neto and Gonçalves [12] explain and survey component frameworks applied to industrial environments. A comprehensive state of the art of software engineering in industrial automation is presented by Vyatkin [13].

IV. CASE STUDY

The pilot demonstration of the PRODUTECH-SIF project defines the case study for this work. A Portuguese company produces labels and technical narrow fabrics for clothing and other applications is the application target. A part of the production line consists in a variety of looms that produce the labels. Due to its business nature, the company has been acquiring new looms across the years, having now a diversified set of machines, from older to new ones. A set of intermediate aged looms were chosen to constitute the pilot and therefore the application requirements. To support the pilot description one can refer to Figure 4, which illustrates the physical architecture of the case study. These machines have some control electronics and are capable of signaling some errors related to severe failures and others related to small production issues, like rupture of the threads and fabrics. Despite that, these machines are not capable of communication and the only way to acknowledge production problems is through a warning light tower connected to the loom. An operator regularly checks the light towers for faults. A small screen used to upload the label design can also be used to check error codes. The company identified a set of problems that should be tackled by the proposed combination of SmartBox and SmartObject:

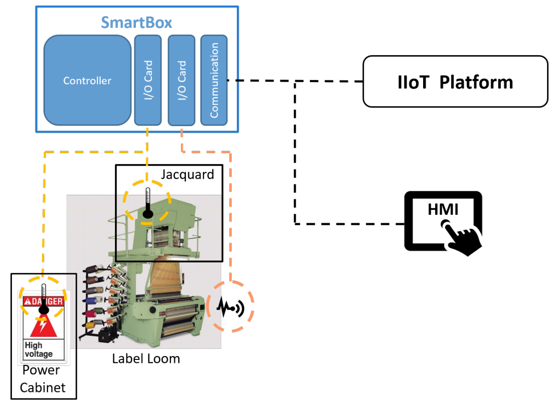

Figure 4. Case Study Physical Architecture.

- **P1**: The power cabinet of the machines is cooled by a fan. Due to mechanical wear, or due to the amount of dust in the air, generated by the threads being processed, the fan fails compromising the cooling. The machine is not prepared to react to this problem and sometimes happens that the power components heat up too much leading to severe failure.
- **P2**: The loom has a cutting system based on electrical resistance elements which can be positioned to cut the fabric according with the required label dimensions. These elements sometimes fail and the machine has no sensing to detect the issue.
- **P3**: The part of the loom which controls the needles is a complex electric-mechanical system called *Jacquard*. This part can sometimes overheat disrupting the mechanical elements of the system and leading to unexpected failures that the machine cannot predict or warn due to the lack on sensing.
- **P4**: There are several issues related with the threads and fabric processed as these break very often. The warning is given by the light tower and there is no automatic way to warn the loom is stopped. Also, related to this, there is no way to know for how long the machine has stopped and what is the rate of production.

V. PROBLEM, REQUIREMENTS AND MOTIVATION

Table I shows the solution proposed for each problem of the case study. A set of requirements result from these proposals. All the hardware aspects are solved by the SmartBox modular capacity. All the software aspects are dealt by the SO as follows. The SO will act as a runtime environment and AS for the assets outlined in the case study. People from the company in charge of the production line will configure the SO through the composition tool discussed in Section III. The tool will have a list off all assets in the shop floor, so that the operator can drag and drop the assets in the canvas to create the desired application. New associations can be made between machines, components, sensors, actuators and software components to create control or monitoring applications defined by the composition workflow. Figure 7 illustrates the vision that the control engineer will have of the monitoring application for the described case study.

TABLE I. CASE STUDY PROPOSED SOLUTIONS.

Problem	Solution
P1	Install a temperature sensor in the power cabinet. Use a software component programmable alarm to alert for excessive temperature.
P2	Install a temperature sensor in the cutting element. Use a software component programmable alarm to alert for excessive temperature.
P3	Install a temperature sensor in the Jacquard. Use a software component programmable alarm to alert for excessive temperature.
P4	Install an inductive/encoder sensor in the loom. Use a software component to convert the sensor impulses into meters.

For all the functionalities discussed so far to be possible, there was a necessity to use an information model that would be capable of tackling the following requirements:

- The SO must abstract all machines and its components (loom and respective power cabinet and Jacquard), sensors and actuators.
- The SO must abstract all software components used and the respective instances.
- The SO must allow for any external actor to subscribe data produced by any device or software component instance.
- The SO must allow to define hierarchy and subscription relations among all devices, software component instances and external actors.

Although the RAMI 4.0 defines ontologies and information models to support communication and representation of several industrial assets throughout its respective life cycles [14] [15], it seems to lack in specification for the composition of logic assets – such as software components [16] – in workflows. A set of applications that are of great importance for a smart factory, such as: condition monitoring, predictive maintenance, self-reconfiguration, quality control and fault detection; depends of software components. Therefore, there should be an information model which could represent all software component peculiarities and its relation with industrial assets to form applications. In this paper, we propose such an information model.

VI. SMART OBJECT SELF DESCRIPTION

The SOSD defines classes and properties for all assets specified in the use case, taking in consideration the following requirements:

- **Physical Assets:** The model is capable of representing: machines, machine components, sensors and actuators. It also must be capable of representing dependencies and connections between these. This is of major importance if we want to contextualize information of a given sensor, or if components need to be represented as parts of some machine. As an example, in Figure 7, it can be seen that there are: 1) relations between a machine and it's constituent parts, the hierarchical references between the *Loom* and it's *Power Cabinet* and *Jacquard* components; 2) relations of contextualization between machines or machine parts and sensors, as between the *Loom* and *Inductive Sensor*, or between the *Jacquard/Power Cabinet* and the *Temperature Sensors*. **For representing physical assets the class *SOSD:DeviceType* (left in Figure 7) is used.** This class has sub-types: 1) *Device*, to represent machines; 2) *Component*, to represent machine components; 3) *Sensor*, to represent sensors; 4) *Actuator*, to represent actuators.
- **Services and Service Instances:** A service, in the SOSD context, represent some algorithm or computational process that is available in the CPPS network. A SO or any other node in the CPPS can announce its services and respective capabilities, e.g. data processing services like a Fast Fourier Transform (FFT) or a simple alarm. A service corresponds to a software component that can be instantiated by including it in the composition design. Each new instance created can be interfaced with providers and subscribers, e.g. the *Alarm* and *Impulses to Meters* instances in the view of Figure 7 are fed by sensors and feed external nodes in the CPPS. **Class *SOSD:ServiceDescriptionType*, represented in Figure 8 under the folder *ServiceDescriptionSet*, is used to represent software components maintained by the SO composition tool. Class *SOSD:ServiceInstanceType*, represented on the left side of Figure 7, is used to represent instances of software components running in a SO.**
- **Points:** A Point represent a node in the CPPS network, e.g. an Human Machine Interface (HMI) device used by the operators to check production variables or an Industrial Internet of Thing's (IIoT) Platform to maintain production telemetry data. In the composition of Figure 7, the *HMI* and *IIoT Platform* are notified each time the *Alarm* or *Impulses to Meters* service instances produce a new output value. **The class *SOSD:PointDescriptionType*, on the left side of Figure 7, is used to represent other nodes in the CPPS with which the SO can communicate.**
- **Variables, Parameters and Methods:** All static and dynamic variables and parameters associated to assets, services or endpoints must be represented. Static variables represent information about some entity, e.g. in case of machine, the manufacturer, model and serial number. Dynamic variables represent information generated during production, e.g. a sensor value or a service output. Parameters represent values that can be changed to modify or tune some process, e.g. a welding machine laser power or an alarm minimum and maximum thresholds. A Method represents an simple routine that can be invoked, e.g. a calibration method for a sensor or a stop routine for a machine. For simplicity and compatibility reasons, each Variable and Parameter defined under the classes proposed

 ISBN: 978-1-61208-723-8

by SOSD, are represented using *BaseDataTypes*, *VariableTypes* and *Method semantic* from the OPC UA Information Model (OPC UA Specification: Part 5) [17]. In Figure 8, under the main classes representation, examples of properties and variables are illustrated.

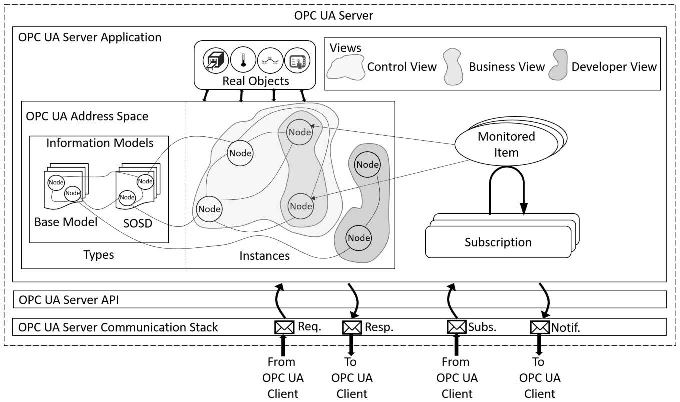

Figure 5. OPC UA Server with SOSD (Adapted from [18]).

OPC UA is the *de facto* standard communication protocol under the RAMI 4.0 proposal [19]. For that reason, the SO network interface embeds an OPC UA Server. The SOSD model was entirely mapped in the OPC-UA native and Data Access types and information model. Once an asset is physically or wireless connected to the SmartBox, its information will be mapped in the SOSD model (Figure 5), and make available to the CPPS by a local OPC-UA server instance.

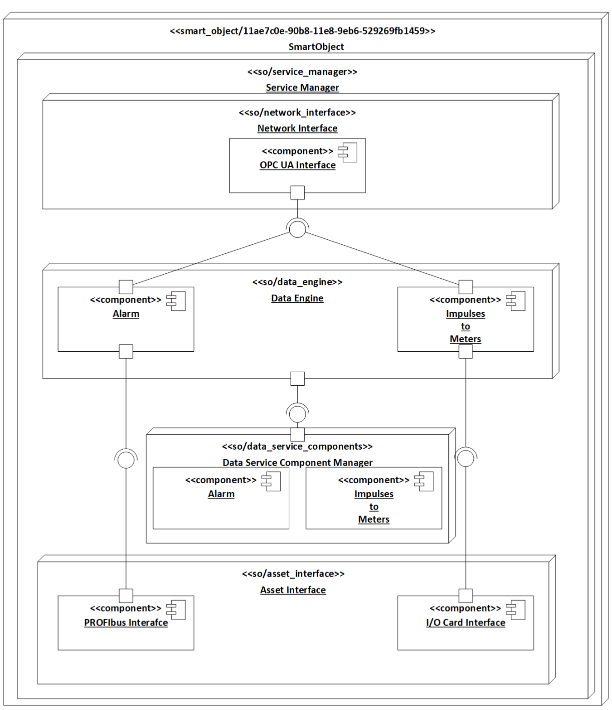

Figure 6. SmartObject Component Developer View.

Figure 5 illustrates the functionality details of an OPC UA server embedded in the SO. Each node represented by a SOSD class will have a direct dependency to the SOSD model. Each property, variable or method of an SOSD class will reference the OPC UA Base Model.

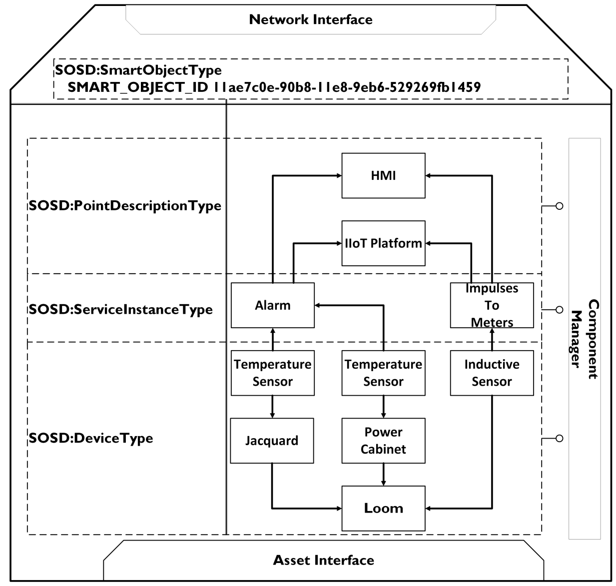

Figure 7. SmartObject Control Engineer View.

The combination of the SOSD with the OPC UA views allows to separate the visions of the software developer and the industrial domain expert. Figure 6 relates to how the software components developer sees the composition running in the SO. In the bottom there are the components that implement the drivers for the temperature and impulse sensors. In the middle there are the components that implement the alarms and impulse to meters algorithms. In the top there is a component with an embedded OPC UA driver. The software architect can grasp the SO composition and develop, modify and reuse components as needed by the domain experts. Figure 7 relates to how the industrial domain expert sees the composition running in the SO. Each block in the figure corresponds either: 1) to a physical asset as discussed in the case study (Section IV); 2) to a software component as discussed previously. The figure shows a monitoring application that the control engineer could assemble by dragging and dropping the blocks using the SO composition tool.

Another important concept provided by OPC UA is the views. Figure 5 gives a general idea of views: *Control View*, *Business View* and *Development View*. These views define which nodes are presented to different users groups. A control engineer, who has the expertise to build control and monitoring applications, only cares about nodes relevant to build or watch technical compositions, as in Figure 7. A software component developer only cares about nodes relevant to the SO architecture, as in Figure 6. The business view, although not represented, can be used to specify users only interested in see components related to production performance and other Key Performance Indicators. By creating only the essential classes and structures, and combining these with the OPC UA features, the SOSD was demonstrated to tackle all requirements of the Case Study, constituting a solid start point for software component models assembled under the scope of AS and SO. Figure 8 shows a tree view of the SOSD model embedded in a working SO OPC UA server. Due to constraints in size and

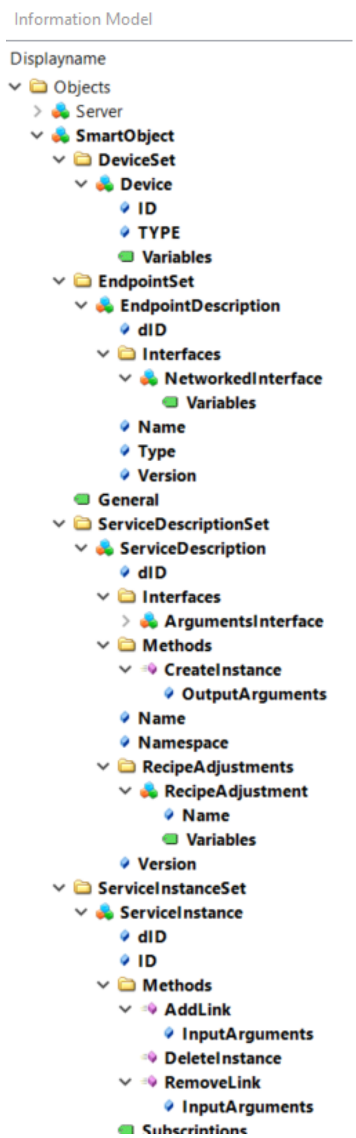

Figure 8. SmartObject OPC UA Server SOSD Model View.

simplification this representation cannot include all details, but it gives a general idea of how the model works in practice.

VII. Conclusions and Future Work

This work proposes the SOSD information model, whose objective is to establish a standard for the representation of software components, assets and compositions between them. This model was successfully embedded in a OPC UA server and used to model a real case study.

Future work will involve the modeling of more complex scenarios and the extension to support real-time quality of service requirements in the composition. The capability of defining specific views for different components also needs to be explored to support a business perspective.

Acknowledgment

This research was supported by the project *PRODUTECH-SIF - Solutions for the Industry of the Future*, financed by the Portuguese National program COMPETE 2020 and Portugal 2020.

References

[1] F. Zezulka, P. Marcon, I. Vesely, and O. Sajdl, "Industry 4.0–an introduction in the phenomenon," IFAC-PapersOnLine, vol. 49, no. 25, 2016, pp. 8–12.

[2] G.-G. et al., "Towards a semantic administrative shell for industry 4.0 components," in 2016 IEEE Tenth International Conference on Semantic Computing (ICSC). IEEE, 2016, pp. 230–237.

[3] M. Hankel and B. Rexroth, "The reference architectural model industrie 4.0 (rami 4.0)," ZVEI, vol. 2, 2015, p. 2.

[4] M. Hermann, T. Pentek, and B. Otto, "Design principles for industrie 4.0 scenarios," in 2016 49th Hawaii international conference on system sciences (HICSS). IEEE, 2016, pp. 3928–3937.

[5] M. A. . C. K. Hoffmeister Festo. Industrie 4.0: The industrie 4.0 component. [retrieved: 6, 2019]. [Online]. Available: \url{https://www.zvei.org/fileadmin/user_upload/Themen/Industrie_4.0/Das_Referenzarchitekturmodell_RAMI_4.0_und_die_Industrie_4.0-Komponente/pdf/ZVEI-Industrie-40-Component-English.pdf}

[6] P. e. a. Adolphs, "Structure of the administration shell. continuation of the development of the reference model for the industrie 4.0 component," ZVEI and VDI, Status Report, 2016.

[7] "PRODUTECH-SIF - solutions for the industry of the future," http://mobilizadores.produtech.org/en/produtech-sif, [retrieved: 6, 2019].

[8] L. Neto, J. Reis, R. Silva, and G. Gonçalves, "Sensor selcomp, a smart component for the industrial sensor cloud of the future," in 2017 IEEE International Conference on Industrial Technology (ICIT). IEEE, 2017, pp. 1256–1261.

[9] Pinto, Rui; Reis, João; Silva, Ricardo; Pesch, Michael, and Gonçalves, Gil, "Smart sensing components in advanced manufacturing systems," International Journal On Advances in Intelligent Systems, vol. 9, no. 1 & 2, [retrieved: 6, 2019]. [Online]. Available: {https://repositorio-aberto.up.pt/bitstream/10216/109876/2/239571.pdf}

[10] S. Micheler, Y. M. Goh, and N. Lohse, "Green paper: R&d priorities for smart manufacturing components and systems," 05 2017.

[11] Neto, Luis; Madsen, Anders; Jeppesen, Nicolaj Søndberg; Silva, Ricardo; Reis, João; McIntyre, Peter, and Gonçalves, Gil, "A Component Framework as an Enabler for Industrial Cyber Physical Systems," in IEEE International Conference on Industrial Cyber-Physical Systems (ICPS 2018). [Online]. Available: {https://ieeexplore.ieee.org/abstract/document/8387682/}

[12] Neto, Luis and Gonçalves, Gil, "Component Models for Embedded Systems in Industrial Cyber-Physical Systems," in The International Symposium on Intelligent Manufacturing Environments InManEnv 2018, [retrieved: 6, 2019]. [Online]. Available: {https://www.researchgate.net/profile/Luis_Neto8/publication/326693373_Component_Models_for_Embedded_Systems_in_Industrial_Cyber-Physical_Systems.pdf}

[13] V. Vyatkin, "Software engineering in industrial automation: State-of-the-art review," IEEE Transactions on Industrial Informatics, vol. 9, no. 3, 2013, pp. 1234–1249.

[14] F. e. a. Zezulka, "Communication systems for industry 4.0 and the iiot," IFAC-PapersOnLine, vol. 51, no. 6, 2018, pp. 150–155.

[15] X. Ye, T. Y. Park, S. H. Hong, Y. Ding, and A. Xu, "Implementation of a production-control system using integrated automation ml and opc ua," in 2018 Workshop on Metrology for Industry 4.0 and IoT. IEEE, 2018, pp. 1–6.

[16] H. Bedenbender, A. Bentkus, U. Epple, and T. Hadlich, "Industrie 4.0 plug-and-produce for adaptable factories: Example use case definition," Models, and Implementation, 2017, pp. 56–62.

[17] S.-H. Leitner and W. Mahnke, "Opc ua–service-oriented architecture for industrial applications," ABB Corporate Research Center, vol. 48, 2006, pp. 61–66.

[18] J. Imtiaz and J. Jasperneite, "Common automation protocol architecture and real-time interface (capri)," in Kommunikation unter Echtzeitbedingungen. Springer, 2013, pp. 79–88.

[19] L. Rauchhaupt, G. Kadel et al., "Network-based communication for industrie 4.0-proposal for an administration shell," Federal Ministry for Economic Affairs and Energy (BMWi), 2013.

Path Planning for an Industrial Robotic Arm

Zahid Iqbal, João Reis, Gil Gonçalves

SYSTEC, Research Center for Systems and Technologies
Faculty of Engineering, University of Porto
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
Email: {zahid,jpcreis,gil} @fe.up.pt

***Abstract*—Enabling humans and robots to work together in modern industrial environments can increase production volumes and reduce costs. However, a robot must be equipped to perceive humans and redirect its actions under hazardous events or for cooperative tasks. Thus, dealing with dynamic obstacles appears as an essential exercise. This work presents a motion planning algorithm for robots based on Probabilistic Road Maps (PRM). For efficient nearest neighbour search, we use kd-trees in learning and query phases of the algorithm. We construct the roadmap as an undirected graph in the free space. We implement the method for a simple configuration space in $\mathbb{R}^2$ and a point robot is considered to navigate between given initial and goal configurations added to the roadmap, all specified in two dimensions. We use Euclidean distance when finding the closest neighbours. The shortest path between the start and goal configurations is found using Dijkstra's algorithm. The method is easy to implement. After the learning phase, the method can answer multiple queries. We propose to use this method in combination with a labelled voxel-based grid for solving multiple path planning queries efficiently.**

Keywords–sampling-based planning, configuration space, grid-based search, kd trees, PRMs

I. Introduction

Today, standard industrial practices use robots for improving production volumes, bringing down production costs, and for better precision and accuracy of the production process. Additionally, as robots capabilities increase, they can take on jobs that might be impossible or dangerous for humans [2]. To ensure safety, typical assembly environments within industries confine robots to separate operation spaces isolated from human workers. Such a setting, on the one hand, incurs significant space and installation costs, and on the other, loses cooperation opportunity with humans. Such systems frequently bring robot operation to a halt if a human entered its operational area [3].

However, with technical progress and to realize their full potential, it became common to put robots in more open unstructured environments. Such environments can leverage the benefits of cooperation by assigning specific production tasks to robot and humans. Moving from a highly structured to an unstructured environment poses several challenges for motion planning, an important one being that robot can only possess a partial knowledge of its surroundings [4]. Most planners assume that manipulator operates in a static environment. However, for many situations, such as the collaborative environment where humans work in proximity to the manipulator, the robot must account for dynamic environments. For such environments, we need methods with two goals; firstly, to avoid collision of robots with foreign objects (environment obstacles and humans) and secondly, with the ability to detect potential collisions in advance.

The work presented in this paper concerns the first goal, i.e., obstacle avoidance for an industrial robotic arm or so-called manipulator. Mounted on a stationary platform, its links with revolute joints and end-effector can move with a certain degree of freedom. The solution deals with forward and inverse kinematics of the robotic arm. With forward kinematics, we aim to calculate the end-effector pose from the position of joints. Inverse kinematics deals with the opposite problem; given the position of the end-effector in the configuration space, we must work out the angles that each joint should have to reach that configuration. Given a robot with a description of its kinematics, a description of the environment (representation of free and occupied spaces), an initial state, and a goal state, the motion planning problem is to find a sequence of inputs that drive the robot from its initial state to the goal state while obeying the rules of the environment, e.g., not colliding with the surrounding obstacles (Figure 1). The manipulator will follow this path to reach the goal position. This work, however, explores the idea for static scenario and later we develop it for the dynamic case.

We follow a sampling-based path planning approach called Probabilistic Roadmap Planner (PRM) [5]. The method randomly samples configurations from the configuration space of the manipulator. Following this, it builds a roadmap graph of free space. Finally, it connects the initial and goal configurations to this roadmap and finds the shortest path from the start to the target using some standard algorithms such as A* [6] or Dijkstra's algorithm [7].

Our work is in the context of the project INDTECH 4.0 [8], which aims at developing new technologies for intelligent manufacturing such as Collaborative Robotics, Autonomous Drive Systems, Decision Support Systems (DSS). In the lines of INDTECH 4.0, different works have focused on these areas [9][10][11][12]. Reusable software components are major building blocks for modern Cyber-Physical Systems (CPS), efficiently managing the complexity by supporting modular development and composability. Neto and Gonçalves [9] studies component models for many industrial CPS for understanding design choices and application targets, and Neto et al. [10] designs a component framework following component-based software engineering principles. Reis et al. [11] demonstrates a collaboration scenario between human and a simple robotic manipulator in the context of Cyber-Physical Production Systems (CPPS).

The rest of the paper is organized as follows. In Sec-

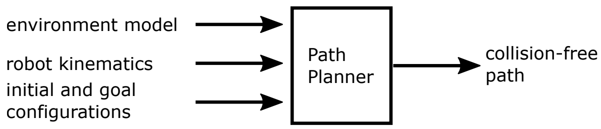

Figure 1: Robot motion planner, a top-level view.

tion II, we present some concepts and background important to the work. Section III lists some works regarding robotics path planning. Section IV presents the developed approach as well as some preliminary evaluation. In Section V, we propose a path planning architecture based on a voxelized grid. Section VI discusses merits and potential of the proposed approach. Finally in Section VII, we conclude the paper and present some future work directions.

II. Background and Concepts

Fundamental to the path planning problem is the concept of *configuration* and *configuration space*. The configuration of a robot is a set of independent parameters that characterize the position of every point in the object whereas *configuration space* (denoted $\mathcal{C}$) is the space of all possible placements of the moving object, i.e., space of all configurations [13]. It is important to distinguish between *work space* and the *configuration space*. *work space* represents the actual environment or the world where the manipulator and the obstacles exist; *configuration space* is the representation of the *work space* that we use in our implementations to aid the path planning solutions. In the following, we review two common approaches for path planning, namely, configuration space-based and sampling-based approach.

A. Path planning - configuration space approach

In order to find a trajectory in the configuration space $\mathcal{C}$, the main difficulty lies in the high computation complexity in constructing $\mathcal{C}$. The dimension of the configuration space is determined by the degrees of freedom of the robot. A configuration q is given as a vector of robot joint values. For instance, a robotic arm with six revolute joints has six degrees of freedom and its configuration is denoted $q = \langle q_1, q_2, q_3, q_4, q_5, q_6 \rangle$ where q_i denotes i^{th} joint angle. The configuration space $\mathcal{C}$ is a space of these configurations, i.e., for 6 dimensions, we have $\mathcal{C} = R^6$. Any path finding strategy places configurations in $\mathcal{C}$ into two categories, those that are free $\mathcal{C}_{free}$, and others that are in collision $\mathcal{C}_{forb}$, i.e., occupied by obstacles. A configuration $q \in \mathcal{C}_{free}$ if the robot placed at q does not intersect with work space obstacles. A path is a continuous sequence of configurations in $\mathcal{C}_{free}$ connecting initial and goal configurations. Constructing $\mathcal{C}$ entails creating a map of $\mathcal{C}$ such that collision regions can be identified. One way is to discretize $\mathcal{C}$ and test each discretized configuration if it is collision-free. The resolution or granularity of the discretized $\mathcal{C}$ has a significant impact on the performance of path finding algorithm. A fine resolution will increase the search space as well as computation time whereas a coarse resolution might miss a valid path when there exists one. We illustrate this with an example of a robot manipulator with six revolute joints and each joint can rotate between $-180°$ to $180°$. With a discretization resolution of $1°$, size of $\mathcal{C}$ is $360^6 \approx 2.18 \times 10^{15}$ points. Thus, it leads to a huge search space when explicitly constructing collision regions in $\mathcal{C}$.

B. Sample-based planning

To avoid the great computational complexity and the state explosion problem, motion planning algorithms often employ sampling-based planning together with a collision detection module. Such methods first generate a roadmap representing connectivity of $\mathcal{C}$ and later path search requests are processed on this roadmap. Sampling based planners can often create plans in high-dimensional spaces efficiently.

Approaches to the path planning problem can be divided into two classes *single-query* and *multiple-query* planning. For single-query planning, a one-time solution to a unique problem is defined, without preprocessing. Multiple-query planning applies to cases where we need to solve different problems for the same environment. Thus, such approaches form a map of the workspace, and then, multiple queries are issued at runtime to find trajectories in that map. The problems relating to dynamic environments fit into the class of single-query planners.

C. kd-trees

Initially examined in [14], a kd-tree is a data structure for storing and searching finite points in a k-dimensional space. Each node represents a subset of the dataset as well as partitioning of the subset. Each leaf node is a k-dimensional point. Each non-leaf node generates an implicit hyperplane splitting the space into two halves. Alike a binary tree, values to the left of the hyperplane are less than or equal to the value at the parent and constitute the left sub-tree, while values to the right of the hyperplane are larger and form the right sub-tree. Further, any node is assigned to one of the k dimensions with hyperplane perpendicular to that dimension's axis. The splitting dimension for each node is selected based on its level in the tree; we obtain the splitting axis for each level by cycling through the k dimensions in order, given by the following rule, $D = L \ mod \ k + 1$ where D is the splitting axis for level L and the root is defined to be at level zero [14].

Nearest neighbour search is an essential component of many path planning strategies. Some approaches generate thousand of vertices; finding closest vertices to connect to in huge search space is a challenging task. kd-trees can efficiently aid the nearest neighbour search by quickly eliminating large portions of the search space. Partitioning strategy in k-d trees allows keeping the average number of examinations small when searching for the best matches to a query item. For the experimental exercise reported in this paper, we naively use kd-trees, i.e. based on Euclidean distance. However, Yershova and LaValle [15] shows that the distance metric can be tailored to account for complex spaces encountered in motion planning scenarios.

III. Related Work

In the past years, much research has focused on path planning for industrial robots. A fundamental robotic task is to navigate from an initial position to a goal position while avoiding the set of obstacles. The developed approaches encompass static scenarios with fixed operational space of robot or dynamic situations where humans or obstacles can

 ISBN: 978-1-61208-723-8

move in the functional area of the robot. Other classifications include configuration space approaches vs the random sampling approaches, or single query based planners vs multiple-query based planners. For a detailed review of different path planning approaches, refer [16]. Below, we describe a few basic methods and some works that employ these approaches to solve motion planning problems.

Rapidly-exploring random trees (RRT) [1] and its several improvements, such as RRT-connect [17], RRT* [18] are sample-based, single-query path planning algorithms. The basic idea of randomly sampling free space is similar to PRMs [5]. The tree is rooted at the start configuration. A random free sample x_{rand} is generated, and if it is within a distance ϵ to the closest tree node $x_{nearest}$, a direct link is created between the two. Otherwise, a new node x_{new} is generated, which is at most within ϵ from $x_{nearest}$ and in the direction of x_{rand}. A connection is formed between $x_{nearest}$ and x_{new}, and x_{rand} is discarded. Random samples can be seen as controlling the expansion direction of an RRT. Local planner checks for collision avoidance when making connections. RRT effectively biases the search towards unexplored regions of the search space. As opposed to PRM, RRT remains connected even with a small number of edges. Instead of defining a goal state, a goal region can be defined. When the tree expansion falls in this region, we have a possible path connecting start to the goal. The RRT-connect approach differs from the basic RRT in two aspects. Firstly, instead of the incremental growth by ϵ, it grows the tree to the new random node x_{rand} or until an obstacle is encountered. Secondly, it uses two RRTs, one rooted at q_{init} and the second at q_{goal}. The trees are maintained until they get connected and hence a path is found. Within RRT*, the cost of reaching each vertex from q_{init} is recorded. Alike RRT, initially, RRT* extends the nearest tree node $x_{nearest}$ towards the random sample x_{rand} thus giving the new node x_{new}. However, it then examines a neighbourhood of vertices in a fixed radius around x_{new}. If any such vertex x_{near} incurs the minimum accumulated cost to reach x_{new}, then, it is made the parent of x_{new}. Another difference is that RRT* rewires the tree. It tests all x_{near} and if any such vertex can be accessed through x_{new} with a smaller cost, then x_{near} is rewired to x_{new} making the path more smooth.

PRMs (see Section IV) belong in the category of multiple-query sample-based planners. Having constructed the roadmap of $\mathcal{C}$, we can put multiple queries to the roadmap, each time specifying different end configurations.

Bertram et al. [19] presents a strategy for finding a solution when goal configuration is unreachable as it lies in a disconnected component of the configuration space with respect to the initial configuration. The main idea is to integrate the IK solution directly into the path planning process. Instead of specifying an explicit goal configuration, the planner evaluates configurations that might belong to a valid goal region. A heuristic workspace goal function calculates proximity to the target given the end-effector pose. It uses forward kinematics to find the end-effector pose from current configuration q. The function uses a weighted sum of different factors to characterize the goal region. The foremost factor is the Euclidean distance between the origin of the coordinate frame attached to the end-effector and the centre of the target object; penalty terms are added to constrain the orientation of the end-effector. It uses a modification of the RRT algorithm to guide the search. It ranks configurations based on their distance to an obstacle and their goal distance. New node goal distance must be smaller than its parent. A node which is in a collision or its goal distance is not lower than its parent is removed from the search after some failure count. The tree grows in configurations that reduce the goal distance.

Qin and Henrich [20] presents a parallel randomized approach for the path-finding problem. This work uses a discretized version of the C-space already discounting C-space regions where the arm has mechanical limitations to reach and where obstacles lie. The idea is to generate many sub-goals in free C-space. Then, it uses several parallel processes where each process will find a path connecting initial configuration with the goal configuration through a sub-goal. All processes terminate whenever a process returns a valid path. The transition from the current state to the next state is based on a cost function which selects the candidate with the minimum cost, also considering that it is within the workspace and collision-free.

Henrich et al. [21] proposes a strategy to reduce C-space size using different discretizations along each coordinate of the configuration vector, with joints closer to the base having a finer discretization resolution. The so-called optimized discretization is such that each joint results in the same movement in the Cartesian space when rotated by the chosen angle. The authors provide a formula for computing the desired angle that takes into account the distance from the joint centre to the farthest point on the end effector.

In this work, we implement a PRM approach for a simple space in $\mathbb{R}^2$, and we consider a point robot that must navigate the area. However, as the method developes, we integrate a collision detection module, that makes use of a voxel-based grid.

IV. Approach

This work is a preliminary exercise towards finding a solution to the robotic arm path planning problem. Here, we consider a static environment, and we follow a sampling-based approach called probabilistic roadmap (PRM) planner [5]. The method consists of two phases, a *learning phase* and a *query phase*. Rather than computing the configuration space explicitly, we sample it during the learning phase. This phase constructs the roadmap as a graph where sampled configurations are vertices and connections between configurations are the edges. In the query phase, we ask for a path between two free configurations. Collision detection can be an independent module and can serve in different phases of the construction of trajectory. To allow adequate connectivity of $\mathcal{C}$, we sample many configurations widely distributed over the free space.

A. Learning phase

The algorithm works as follows. We begin with an empty graph G. In the *learning phase*, we repeatedly sample a random configuration q from $\mathcal{C}$. There can be a collision detection test that checks whether the selected configuration is free i.e., $q \in \mathcal{C}_{free}$. If this check succeeds, q is added to the set of vertices V. This process continues until a given number of nodes n have been added to V. n is a parameter of the algorithm and is the desired number of nodes in the graph. In the second step, for each vertex of G, we find its k nearest neighbours denoted $P = \{p_1, p_2, \cdots, p_k\}$ according to some $dist$ metric.

 ISBN: 978-1-61208-723-8

Then, an edge is created from q to its nearest neighbours, i.e., $\forall p_i \in P, E = E \cup (q, p_i)$ where E is the list of edges in G. At the end of the *learning phase*, we have an undirected graph $G = (V, E)$. Due to the random nature of sampling, it is possible that roadmap G has disconnected components.

B. Query phase

The next step is the *query phase*. We are given the initial and goal configuration of the robot denoted respectively q_{init} and q_{goal}. The algorithm must find a feasible path connecting q_{init} and q_{goal} or return failure. We find k nearest neighbours of q_{init} and q_{goal} from V using some distance metric $dist$. These sets are denoted $P_{q_{init}}$ and $P_{q_{goal}}$ respectively for q_{init} and q_{goal}. Then, we check each element e of $P_{q_{init}}$ in order and add an edge $E = E \cup (q_{init}, e)$. Similarly, we add an edge connecting $P_{q_{goal}}$ in the roadmap. Finally, we can use any shortest path algorithm such as Dijkstra's algorithm [7] to find the path between initial and goal configurations. Algorithm might fail to return a path when roadmap has disconnected components. The working of the PRM algorithm can be explained as shown in Figure 2. We assume an Euclidean space $\mathbb{R}^2$. Robot configuration is given by a point in $\mathbb{R}^2$, depicted with an empty circle whereas shaded regions represent the obstacles. Figure 2(a) shows many random samples of the free space which are nodes of the roadmap. In Figure 2(b), we form edges between nodes using k closest neighbours. k is three, but the degree of some nodes can be greater when it is included in the closest neighbours of different nodes, or it can be smaller if it cannot find k closest neighbours due to an obstacle. In Figure 2(c), we solve a query in the roadmap. The configurations q_{init} and q_{goal} are connected to the roadmap. Then, a graph-based search algorithm returns the shortest path denoted by thick lines.

C. Remarks

Several essential aspects affect the performance of PRM; for instance, how can we choose the random configurations such that sampling of the configuration space is uniform. The original PRM uses a local planner. This local planner is instrumental for creating connections between any two sampled configurations hence creating a feasible local path. The task of the local planner is to interpolate the motion of the robot between the two samples, checking for collisions at a given resolution. If no configuration of the robot between the samples collides with an obstacle, then an edge is inserted to the roadmap. Both phases of the PRM algorithm employ the local planner. In the presented work, however, we have not used a local planner. Instead, having found the nearest neighbours using kd-tree, we make connections creating edges in the graph. We use kd-tree in both phases of PRM. In the reported implementation, kd-tree use Euclidean distance when finding nearest neighbours. However, due to the complex topology of the configuration space, such distance metric may not be directly applicable to the path planning algorithms.

D. Preliminary evaluation

We have implemented the approach in a simple 2d test scenario. We create a two-dimensional point grid that represents configuration space $\mathcal{C}$. We also specify the obstacle points which is a subset of $\mathcal{C}$ and represents the occupied or forbidden part of the space, i.e., $\mathcal{C}_{forb}$. The number of points in $\mathcal{C}$ and $\mathcal{C}_{forb}$ is a parameter of the algorithm. When creating the roadmap, we also specify the desired number of points to sample and also the number of closest neighbours to connect to. We use kd-tree and we find the nearest neighbours based on Euclidean distance, i.e., for points $p = (x_1, y_1)$ and $p_2 = (x_2, y_2)$, we calculate distance $d(p, q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. When forming edges, d serves as the weight on the respective edge.

Figure 3 shows an example configuration space with 40 free points (blue) and 10 collision points (red). In the next step, we sample $\mathcal{C}$ and construct a roadmap in the free region. The following 15 points were randomly selected as vertices of the roadmap.

$V = \{(11, 4), (12, 8), (14, 6), (39, 12), (12, 7), (24, 27), (46, 3), (22, 33), (15, 0), (23, 25), (11, 9), (8, 32), (6, 16), (4, 37), (29, 46)\}$

We use 2 nearest neighbours while forming the edges between nodes. Figure 4 shows the roadmap. Finally, we query the roadmap to find a path between an initial configuration $(22, 33)$ and a goal configuration $(11, 9)$. For this exercise, these configurations have been selected from the constructed roadmap. Generally, q_{init} and q_goal must be added to the roadmap at this stage. Using Dijkistra's algorithm and employing $d(p, q)$ as the cost of going from p to q, the following shortest path is returned (Figure 5). $\{(22, 33), (23, 25), (39, 12), (46, 3), (15, 0), (14, 6), (12, 8), (11, 9)\}$

In this way, we can solve a query as long as the roadmap is connected. The connectivity problem gets more complicated as the size of the configuration space increases. To this end, different measures can help, such as increasing the number of nearest neighbours to look at or increasing the number of sampled points.

V. Proposed Architecture

In our proposed architecture, we integrate the probabilistic planner with a grid-based approach for efficiency. The work in [22] explores such a relation. Figure 6 shows the main components of the architecture. In particular, the framework receives the initial and goal configurations q_{init} and q_{goal} and returns a collision-free path. The main component of the architecture is the PRM which we explained in Section IV. Essential tasks of the planner are to build a roadmap and establish local paths. These tasks rely on efficient nearest neighbour search and collision detection, which we add as independent modules within the architecture. Nearest neighbour search can use kd-trees, but it also needs information for collision-free nodes as well as an appropriate distance metric. Many works have used slight modifications of Euclidean distance [5][19] whereas others specify different metrics for more complex spaces [15]. Local planner decides whether we can add an edge to the roadmap. It has the task to interpolate robot motion between two given configurations (nodes from the roadmap). For this purpose, it uses the kinematics model of the robot and checks in the voxel grid if a given configuration is collision-free. The voxel-based grid renders the entire scene of the collaborative environment as a grid composed of cubes with a given dimension, known as voxels. We can describe the granularity of the grid with the size of one side of the cube. The resolution can be set higher or lower by choosing the dimension of the unit cube

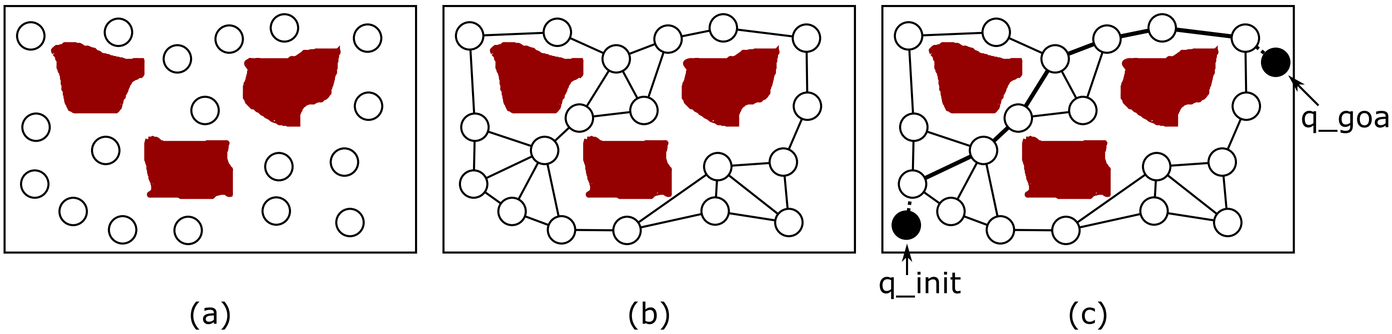

Figure 2: An example of a roadmap in a two-dimensional Euclidean space.

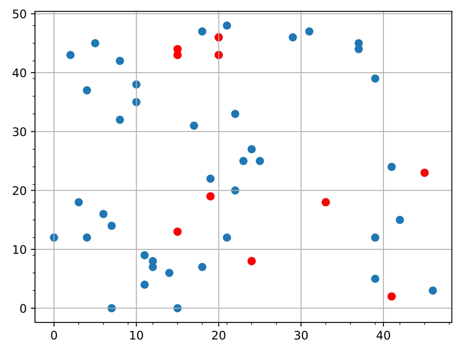

Figure 3: Two dimensional configuration space.

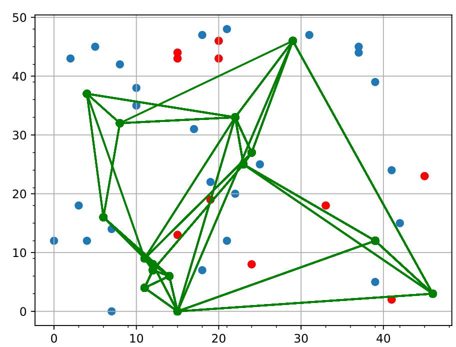

Figure 4: Randomly sampled points. Straight lines between points correspond to edges. The number of k closest neighbours for the construction of roadmap is 2.

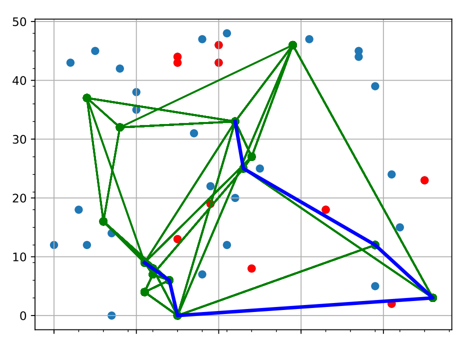

Figure 5: An example of solving a query with the roadmap. Shortest path between initial and goal configurations is depicted with thick lines.

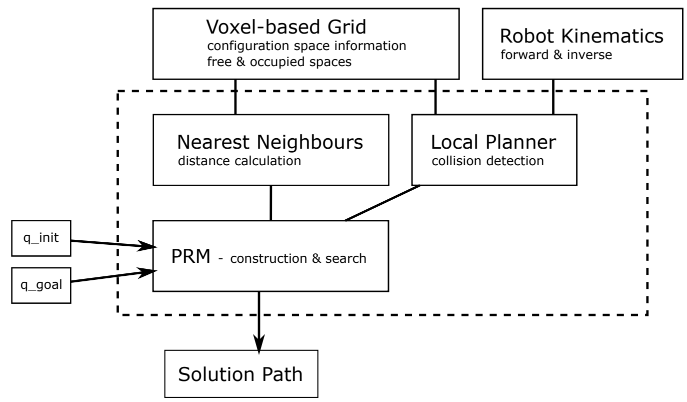

Figure 6: Robot motion planning approach - architecture, environment information is given with a labeled voxel-grid.

in the grid, i.e., for a higher resolution, we choose a smaller size of the voxel, thus corresponding to more voxels in the grid, and vice verse. When voxels in the grid are labelled, we can identify, firstly, if an object in the collaborative environment occupies the voxel, and secondly, which type of object, i.e., either a robotic part itself, human, or an obstacle object. The resolution has a direct impact regarding collision detection as the local planner interpolates motion at the granularity of grid resolution. When using PRM and targeting static scenario, construction of a labelled grid can be a *preprocessing* step without impacting on the processing of queries online. The knowledge of free and occupied space is a vital component of the collision detection module within the path planning algorithm.

The modular architecture allows easy experimentation; for instance, we can use different planners or strategies for nearest neighbour searching or local planning, in future.

VI. Discussion

A desirable property of the developed solution is its real-time performance. High computational complexity is the main limitation of many path planning algorithms, preventing online recalculation of trajectories within the response time of the manipulator [23]. Thus, dynamic environments, which inevitably occur during collaborative scenarios, may not be handled in real-time, and most planners remain applicable to static environments, only. We contend that using a sampling-based approach combined with a voxel-based grid search would support real-time performance, as well as enhance the accuracy of the plan. In the planning algorithm, we can attribute a significant computation time to the following key components, i.e.,

 ISBN: 978-1-61208-723-8

nearest neighbour searching, and collision checking. Firstly, we see how the voxel-based grid supports faster computations for the above two elements. Secondly, we see in which ways can we optimize the nearest neighbour searching and collision checking to enhance algorithm performance.

Nearest neighbour search implicitly checks for collision when it finds a new neighbour. We can efficiently find neighbours in the search space using data structure, such as kd-trees. Since we build the kd-tree from the roadmap, we must account for the time it takes to construct the tree. With an offline preprocessing phase, as in PRM, this computation effort is irrelevant. However, with online planning, we must find the tradeoff between the time taken to construct the tree, and time saved in neighbour searching.

We envision that voxel-based grid search would offer most savings in collision checking for both static and dynamic scenarios. The collision checking in PRM is managed with a so-called local planner. Local planner interpolates robot motion between configurations q_s and q_f. For this purpose, at a given discretization level δ of the configuration space, it advances coordinates of q by δ to reach an intermediate configuration q_i. It keeps advancing until it reaches q_s, testing each q_i if it is collision free. Checking for a collision at a given q_i is equivalent to getting the voxels corresponding to a q_i (robot joint values) in the grid, and their occupancy status. Hence, the local planner can easily find step path. Apart from testing for environment obstacles that lie outside the robot body, collision checking must also check for self-collisions which involve motions where different links might obstruct each other.

The proposed approach will be more flexible in terms of implementation. Since path planning and voxel grid are independent modules, thus, other planners could be combined with the voxel grid.

For a uniform random sampling of $\mathcal{C}$, when choosing a configuration q, we use a uniform probability distribution over an interval of values corresponding to dof of each coordinate in q.

In terms of practical impact, this approach will benefit the collaborative environments. Collaborative spaces reduce costs and improve production volumes by allowing humans and robots to work side-by-side. In this way, humans and robots can perform specific tasks they are best at, e.g., robots perform heavy tasks while humans can inspect for quality. The main concern here is the safety; robot must react in real-time in case of hazardous events, necessitating efficient motion planners. An example of manipulator deployment is within an industrial assembly line which typically distributes the workload among several robots. The main concern is to make the production process efficient such that the final product fulfils its functional requirements, as well as reduces the production time and time to market. A small variation in single parts may propagate such that the final product does not comply with specifications. In such a setting, deriving collision-free paths for each worker is of particular importance. To maximize the number of units assembled means reducing the time needed at each station (robot) to perform its specific task, referred to as the process cycle time [24]. Practical efficiency favours sample-based planners over the configuration space approach. Such planners can derive more robust motions with shorter cycle times [25].

VII. Conclusion and Future Work

The robotic path planning is a classic problem. In this paper, we presented a simple implementation of a robot motion planner based on a sampling approach. We used kd trees for efficient nearest neighbours search. Random sampling might result in roadmaps with disconnected components, and thus, it will fail to find a path when start and goal configurations lie in disconnected components. We would investigate methods to find adequate connectivity of $\mathcal{C}_{free}$. Therefore, to guide the sample selection, our future work would study techniques that reduce the number of samples as well as improve the final roadmap quality. Next step involves testing the presented approach for real scenarios. For such cases, we need to represent configurations in terms of robot joint values instead of the point robot, and to implement the kinematics equations as shown in Figure 1. However, we can find a way to combine the two planning approaches, namely the configuration space approach and the sample based approach for efficiency. A voxel-based representation of the configuration space with additional information on the environment (e.g., the *reachability grid* [26]) can simplify the local planner used to detect collisions in the sampling approach.

Acknowledgments

INDTECH 4.0 – New technologies for intelligent manufacturing. Support on behalf of IS for Technological Research and Development (SI à Investigação e Desenvolvimento Tecnológico). POCI-01-0247-FEDER-026653

References

[1] S. M. LaValle, “Rapidly-Exploring Random Trees: A New Tool for Path Planning,” Department of Computer Science, Iowa State University, Tech. Rep. TR 98-11, October 1998.

[2] “MHI - The Industry That Makes Supply Chains Work,” http://www.mhi.org/fundamentals/robots, accessed: 2019-05-19.

[3] P. Anderson-Sprecher, “Intelligent Monitoring of Assembly Operations,” Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-12-03, June 2011.

[4] D. Katz, J. Kenney, and O. Brock, “How Can Robots Succeed in Unstructured Environments,” in *Workshop on Robot Manipulation: Intelligence in Human Environments at 4th Robotics: Science and Systems Conference (RSS 2008)*. Citeseer, June 2008.

[5] L. E. Kavraki, P. Svestka, J.-C. Latombe, and M. H. Overmars, “Probabilistic Roadmaps for Path Planning in High-Dimensional Configuration Spaces,” *IEEE Transactions on Robotics and Automation*, vol. 12, no. 4, pp. 566–580, August 1996.

[6] P. E. Hart, N. J. Nilsson, and B. Raphael, “A Formal Basis for the Heuristic Determination of Minimum Cost Paths,” *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, July 1968.

[7] E. W. Dijkstra, “A Note on Two Problems in Connexion with Graphs,” *Numerische Mathematik*, vol. 1, pp. 269–271, December 1959. [Online]. Available: https://doi.org/10.1007/BF01386390

[8] “INDTECH 4.0 – Novas tecnologias para fabricação Inteligente, INDTECH 4.0 - New Technologies for Intelligent Manufacturing,” https://site.groupe-

psa.com/mangualde/pt-pt/atualidades/atividade/centro-de-mangualde-do-groupe-psa-na-vanguarda-4a-revolucao-industrial-projeto-indtech-4-0-em-execucao/, accessed: 2019-05-29.

[9] L. Neto and G. M. Gonçalves, "Component Models for Embedded Systems in Industrial Cyber-Physical Systems," in *Proceedings of the 7th International Conference on Intelligent Systems and Applications (INTELLI 2018)*. IARIA, June 2018, pp. 24–29.

[10] L. Neto *et al.*, "A component framework as an enabler for industrial cyber physical systems," in *Proceedings of the 1st IEEE International Conference on Industrial Cyber-Physical Systems (ICPS 2018)*. IEEE, May 2018, pp. 339–344.

[11] J. Reis, R. Pinto, and G. Gonçalves, "Human-Centered Application using Cyber-Physical Production System," in *Proceedings of the 43rd Annual Conference of the IEEE Industrial Electronics Society (IECON 2017)*. IEEE, November 2017, pp. 8634–8639.

[12] L. Antão, R. Pinto, J. Reis, G. Gonçalves, and F. L. Pereira, "Cooperative Human-Machine Interaction in Industrial Environments," in *Proceedings of the 13th APCA International Conference on Control and Soft Computing (CONTROLO 2018)*. IEEE, June 2018, pp. 430–435.

[13] T. Lozano-Pérez, "Spatial Planning: A Configuration Space Approach," in *Autonomous Robot Vehicles*. Springer-Verlag, 1990, pp. 259–271.

[14] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software*, vol. 3, no. SLAC-PUB-1549-REV. 2, pp. 209–226, 1976.

[15] A. Yershova and S. M. LaValle, "Improving Motion Planning Algorithms by Efficient Nearest-Neighbor Searching," *IEEE Transactions on Robotics*, vol. 23, no. 1, pp. 151–157, February 2007.

[16] S. M. LaValle, *Planning algorithms*. Cambridge university press, 2006.

[17] J. J. Kuffner Jr and S. M. LaValle, "RRT-Connect: An Efficient Approach to Single-Query Path Planning," in *Proceedings of the 17th IEEE International Conference on Robotics and Automation (ICRA 2000)*, vol. 2. IEEE, April 2000, pp. 995–1001.

[18] S. Karaman and E. Frazzoli, "Sampling-based Algorithms for Optimal Motion Planning," *International Journal of Robotics Research*, vol. 30, no. 7, pp. 846–894, 2011.

[19] D. Bertram, J. Kuffner, R. Dillmann, and T. Asfour, "An Integrated Approach to Inverse Kinematics and Path Planning for Redundant Manipulators," in *Proceedings of the 23rd IEEE International Conference on Robotics and Automation (ICRA 2006)*. IEEE, May 2006, pp. 1874–1879.

[20] C. Qin and D. Henrich, "Path Planning for Industrial Robot arms - A Parallel Randomized Approach," in *Proceedings of the 4th International Symposium on Intelligent Robotic Systems (SIRS 1996)*. Citeseer, July 1996, pp. 65–72.

[21] D. Henrich, C. Wurll, and H. Wörn, "Online path planning with optimal C-space discretization," in *Proceedings of the 11th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 1998). Innovations in Theory, Practice and Applications*. IEEE, October 1998, pp. 1479–1484.

[22] S. M. LaValle, M. S. Branicky, and S. R. Lindemann, "On the Relationship Between Classical Grid Search and Probabilistic Roadmaps," *The International Journal of Robotics Research*, vol. 23, no. 7-8, pp. 673–692, 2004.

[23] T. Petrič, A. Gams, N. Likar, and L. Žlajpah, "Obstacle Avoidance with Industrial Robots," in *Motion and Operation Planning of Robotic Systems*. Springer, March 2015, pp. 113–145.

[24] N. Papakostas, G. Michalos, S. Makris, D. Zouzias, and G. Chryssolouris, "Industrial Applications with Cooperating Robots for the Flexible Assembly," *International Journal of Computer Integrated Manufacturing*, vol. 24, no. 7, pp. 650–660, June 2011.

[25] D. Spensieri, "Planning Robotic Assembly Sequences," Ph.D. dissertation, Chalmers University of Technology, March 2017.

[26] P. Anderson-Sprecher and R. Simmons, "Voxel-Based Motion Bounding and Workspace Estimation for Robotic Manipulators," in *Proceedings of the 29th IEEE International Conference on Robotics and Automation (ICRA 2012)*. IEEE, May 2012, pp. 2141–2146.

 ISBN: 978-1-61208-723-8

Human-Machine Interaction in Cyber-Physical Production Systems

Henrique Lopes Cardoso*, João Reis*, and Gil Gonçalves*

*SYSTEC, Research Center for Systems and Technologies
*FEUP, Faculdade de Engenharia Universidade do Porto, Portugal
*Email: {up201700417, jpreis, gil}@fe.up.pt

***Abstract*—As factories thrive to produce better and more precise products, the cognitive load increases and the factory workers performance starts to affect the results. This paper presents a possible solution that aims to help the assembly line workers by reducing their cognitive load when looking for the tasks they need to execute for each product and consequently help the business. The proposed solution focuses on a support interface designed to grab the worker's attention to the most important information and uses it to integrate the human in the loop of a Cyber-Physical System (CPS). In this paper, we present the architecture of the developed and constructed CPS, which contains the interface, a robotic arm and an artificial vision system.**

Keywords–CPPS; Support Interface; Cognitive Load; Expertise Level

I. Introduction

As the manufacturing capacity of companies continues to grow, extending their variety spectrum, the workload felt by the human element of these production environment increases [1], and their situational awareness during the operation decreases. This solution has direct impact on the company's operators, since it is where the problem resettles and has an indirect change on the company's business. As for the assembly line operators, it is expected that, with the resolution of the presented problem, they are more attentive throughout the work shift leading to a more precise performance which will prompt a higher success rate and consequently boost their confidence and motivation. With a successful worker comes a successful business. Due to the reduced failure level from the operators, the company is able to produce more, faster and with a higher quality. This improvement will avoid human and material resources waste as the products will more likely get a positive grade when being evaluated by the quality assurance department.

There could be an enormous variety of combinations of tasks in each assembly station, which raises the probability of error if the operator is not certain of how and what set of tasks to complete to fulfill a specific product's requirements, and it can get worse if the factory is capable of producing a large number of different products. For each product that comes on the assembly line there is a list of tasks to be completed on each station on a short time lapse (around 3 minutes, for the PSA Group) and when there are multiple products in a row that require the same tasks the operator starts to ignore the screen because it is not clear that the tasks changed or the information has bad resolution [2] and simply performs the same tasks that completed on the previous product. Mistakes start to appear when suddenly a different product shows up on the assembly line and the operator was not able to identify the changes on the support interface. Most of this mistakes, come from the fact that the operator was not capable to detect the changes on the screen between each product that shows up because the interface was developed without taking into consideration the **cognitive load** of the employees, due to the long working shifts, neither their **expertise level**. In most production factories that contain an assembly line and require human intervention during the process, the working shifts can reach the 16 hours mark and a total of more than 60 hours a week, which is physically and mentally heavy [3], [4] so there is a need to reduce their mental effort when it comes to software that was created to assist the operator. Human operators should not be treated like machines nor expect that they preform like so as they will not perform proficiently as soon as they integrate the production line. There is a learning curve that must be taken into consideration and treated carefully. There is a need to redevelop the support interface system replacing the current one with another capable of adapting to the user's expertise level and also reduce the operator's mental effort when interacted with so that it really becomes a **Support Interface** instead of just a screen with information.

In Section 2 - Motivation and Related work, it is described the motivation to address the problem and some related work with this topic as well as the objective of the proposed. In Section 3 - Research question and proposed concept, the reader will be presented with the developed solution and all the aspects that where taken into account during its implementation related with the Interface. In Section 4 - Scenario, is where the CPS architecture, elements and sequence of activities are described and explained. During Section 5 Discussion, the reader is given a set of arguments that support the motivation for this problem's solution, some direct and indirect expected results with the developed approach and also a few restriction and frailties of the system. Finally on Section 6 - Conclusion, there is given a wrap up of the whole problem proposal and it's solution.

II. Motivation and related work

Modern production lines benefit from the Internet of Things (IoT) and Industrial Cyber Physical Systems (ICPS) technological advances to create environments where Smart Connected Products can influence its own production process and companies can benefit from Service Based business models [5].

With the Industry 4.0 appearance, the human integrates the CPS with its own Digital Twin representation, this way, not only the machines are aware of the existence of other machines but also of humans with which they can and must interact. This integration came from need to change the roles on the collaboration between both parts, from demanding an adaptation from the human operator to the machine [6], [7], to becoming the machine to adapt to the human. With the evolution of the manufacturing paradigm, balancing and changing the volume-variety relationship became a necessity. From Mass Production where factories produce large amount of similar products raising the volume and decreasing the variety, to the Mass Customization, increasing the product variety and reducing the volume manufactured of each model [8]. With this evolution, companies are more agile and capable of responding to the client's will which emphasizes the importance of the human intervention on the assembly line, bringing the human's flexibility to adjust to any expected and unexpected changes [9]. The interplay between humans and a CPS occurs either by direct manipulation, or with the help of a mediating user interface and it is best applied in a Cyber-Physical Production System (CPPS) which comes from the fusion of cyber, physical, and socio spaces through Industry 4.0 [10] and where the machines are responsible for performing the heavier and repetitive operations, while the human employees are responsible for handling shop-floor equipment and supervising processes for high-level decision making [6], aiming this way for the delivery of a high quality product, taking advantages of both human and machine's best attributes to complete a task.

Normally, workload and awareness issues are handled by changing the physical element's (robotic manipulator) autonomy levels, increasing them in order to ease the human operator's intervention. However we propose that changes are made to the support interface, and to the manner in which it is conveyed to the human operator, providing them with a tool that affects the operator workload and awareness in a positive fashion [11]. By now, we have identified the two unexplored variables to the development of a support interface for an assembly line worker. The main objective of this article is to contribute for an easier, more intuitive and adaptable interaction of the production line operator with the interface that provides all the tasks to be executed during the process as follows:

- **Providing the worker with an interface that is able to adapt itself to the the user's expertise and reduce cognitive load.**

This interface will be able to change based on the user's success rate over time which is gathered from an evaluation team that makes sure that every task was executed correctly. In order to have an adaptive interface and since this is inserted in a Cyber-Physical Production System, the system needs to know who is working at each shift which denotes to the need of having a virtual representation of each operator, creating a Digital Twin that will communicate with the remaining of the components as well as hold the information about the operator so that it can adapt the interface accordingly.

III. Research question and proposed concept

The interface is the main aspect of the presented article as it is the bridge between the Human and the remaining of the system, it is the machine with which the Human communicates and interacts.

As a way to identify the worker that is interacting with the interface, a simple login page was created where the worker has to insert a four digit login code that represents him in the system and from this point forward the system has all the required information to present the correct interface and the tasks that the worker is capable of executing.

The developed interface is mainly composed of ”cards” that contain the information regarding each task that the worker must perform. The interface contains a top bar containing complementary information such as the Station in which the worker is working on and the stopwatch indicating how much time the worker has to finish the tasks for that product. Right below this top bar there is a section with a more relevant information about the product in construction showing the product's name and extra features if the product demands so, also in this section is placed the STOP and RESUME button.

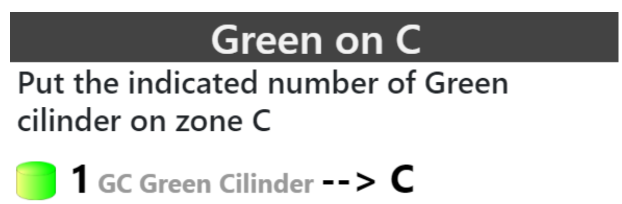

Figure 1: Card with information about the task

These ”cards” (Figure 1) are essentially a rectangle where all the necessary information about the task is presented, containing the task's title on the top part of the card, the task's description right bellow the title where it is explained in a more extensive version how the task must be performed and after this description, a list with the necessary Assets is presented, indicating the quantity, code, name, image and the Zone's letter in which the worker must place the Asset if applied.

A. Grab user's attention

By researching about the topic [12], it was found a way of how one can grab the user's attention through an interface and using the acquired knowledge, Figures 2 and 3 where designed and developed in order to grab user's attention making sure he is aware of what is happening on the system.

Figure 2: Card with information about a task that is only present when an extra package is requested

The figure above represents a task that has to be executed and that is why it has the same structure as regular task. The only difference between the Figure 2 and Figure 1 is how frequently they appear on the screen. For this project it was defined a Default set of tasks that must be performed for every product that may show up for assembly and those tasks are presented has Figure 1 demonstrates. Whenever a product has a specific set of tasks that only belong to that product they are presented as Figure 2 by changing the border color to red that represents attention and caution and the title's background by adding a colored patterned as a way to enhance their appearance and grab the worker's attention to that new and different task that is crucial for that products overall correct assembly.

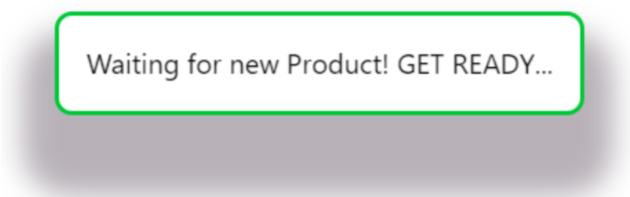

Figure 3: Message shown when waiting for the new product in line

While the Vision system is analysing the working area and the Robotic Arm is reaching out for the new box of assets, the worker is presented with the Message that Figure 3 is showing. This happens so that the user is aware that the product is complete and a new one is being prepared to get assembled and by clearing the interface, we are able to attract the user to the new information. This message is presented with a fat rounded green border passing out a feeling of success and achievement. Note that for this message it is used the Depth and Size illusions to get the user to absorb the most important information which in case would be the "GET READY" part of the message [13], [14], [15].

B. Cognitively Light Interface

In order to make all of the system's interfaces as light as possible for the worker, a list of suggestions was taken into consideration and used as a guide to design them. The interface contains a clear core area that contains the most relevant information and instructions which are simple and concise, as introduced before. Although there is a change on the interface regarding the level of expertise of a user, the format is consistent throughout the different interfaces either if the expertise level changes or if the tasks are unique for a product.

As to color schema, the taken approach was to keep the most relevant information with a darker color and the less important make it clear while having a white background.

C. Adaptive Interface

As it has been referred throughout the document, one of the goals of this proposal is to develop an interface that is able to adapt to the user's expertise level, and that, is accomplished by assigning an interface to each level of expertise available which for the tests environment it was defined as being three levels.

Although these three interfaces vary from one another, their changes are minimal keeping the fundamental structure intact avoiding a contrary effect from the one it is aimed for. This minimal changes rely on the format that the information is shown to the user regarding the assets as it is the only type of information that is flexible to changes.

TABLE I: FORMAT OF INFORMATION AVAILABLE FOR EACH LEVEL OF EXPERTISE

	Image	**Name**	**Code**
Novice	Yes	Yes	Yes
Intermediate	No	Yes	Yes
Expert	No	No	Yes

Table I gives us information of what is the format of the task's information that is presented for each expertise level. As it is visible, the **Novice** user is the one with less experience which means that it is the one who needs more support and so for each asset that is necessary for a task the card holds an image, the name and a code that represent the asset. For an **Intermediate** user, the asset is described with the name and the code, finally for an **Expert** user, the asset is only represented by it's code.

IV. SCENARIO

As a way to simulate an assembly line section as can be seen in an industrial environment, a small system was developed using the previously described interface, a robotic arm and a camera. This section describes how we were able to get as close as possible to a real world environment justifying the appearance of every intervener on the system.

A. Human-in-the-loop of CPPS

The architecture created to test the proposed solution, demonstrated on Figure 4, is divided into three main components each with a smart component (Digital Twin) [16] that represents it in the digital world and a forth smart component that is responsible for the connections between all the other three components and controls the action flow of the system and it is nominates Process Smart Component.

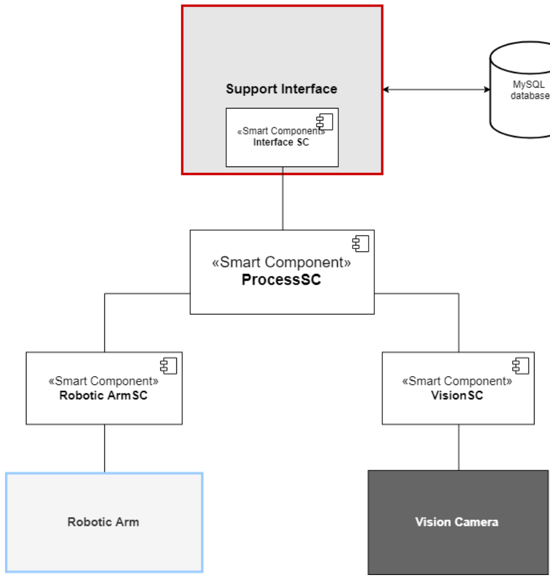

Figure 4: System's Architecture

The first component is the **Support interface** developed, with which the worker interacts where the tasks are displayed as the product progresses in the assembly line. The communication with the smart component happens every time a new product enters the station where the smart component sends an Id that represents the product, as well as the Time Period in milliseconds and the support interface communicates with a database where every task needed to be executed is provided and then displayed to the worker.

The second component is the **Robotic Arm** whose job is to deliver to the worker a box with every asset necessary to complete the tasks. For this component, the smart component is given the box position and the final position and the arm picks up the box and drops it at the final position where the worker starts to preform the tasks.

The final component present in the system is the **Vision Camera** that is responsible of verifying the tasks executed by the worker and rate them in a specific scale with which the algorithm will calculate the expertise level of the worker and consequently adjust the interface to the user.

B. Test the scenario

A solution found to test the System presented on Figure 4 has the following workflow:

1) Worker joins the Working Area
2) Worker enters user code
3) Robotic Arm moves assets box to near the worker
4) Interface shows tasks
5) Worker reads tasks
6) Worker executes tasks
7) Work time ends
8) Camera analyses the area
9) Repeats processes 3 to 8

The test case aims to simulate one Station of the production/assembly line from the moment a worker starts the working shift until the evaluation moment.

In this process, the worker must execute all the tasks presented on the **Interface** by placing the assets on the different areas defined, within the **Time Period** previously defined and shown on the interface in a countdown format. The process starts when the user enters the working station and enters his user code, after this, the **Interface** receives the Product being assembled at the time and gets all the tasks associated with it from the database, ate the same time, the **Robotic Arm** grabs the **Box** with the Assets and gives it to the user. At this point, the user has gathered all the conditions to start executing the tasks. When the time ends, the **Vision Camera** analyses the **Table Area** and verifies the final result. In case of emergency, i.e, the time is ending and the user feels that he wont be able to finis the tasks, the user has the ability to press the **STOP** button which will stop the whole system including the **Robotic Arm** and **Timer**.

C. How it works

Now that the actions flow for a proper test was presented we may go a step further on understanding what is actually happening "behind the scenes" concerning the communications and logic associated with the previously described actions flow.

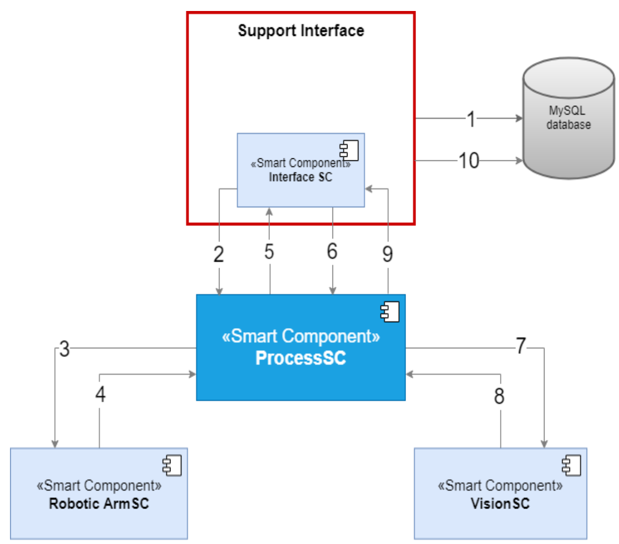

Figure 5: Flow connections between smart components

The above Figure 5 is composed of all the components of the system that are evolved with the communications that happen during the test execution, each arrow represents one of those communications and it's numerations indicate the data flow throughout the test case.

 ISBN: 978-1-61208-723-8

1) User Login;
2) **Interface** tells the **Process Component** that the user is ready and the system may start;
3) The **Process Component** tells the **Robotic Arm** to start the movement;
4) The **Robotic Arm** informs that the movement is complete;
5) The **Process Component** tells the interface which product is being assembled and how much time the worker has and so show the tasks to the worker;
6) The **Interface** indicates that the time has terminated;
7) The **Process Component** tells the **Vision system** to start analysing;
8) The **Vision system** answers with the objects position on the working area;
9) The **Process Component** passes it out to the **Interface**;
10) The **Interface** saves the evaluation rate of the tasks after comparing the objects real positions with the expected ones.

Besides these, there are two other connections that can happen at any time during the system's execution and they are activated by the worker when the Emergency button is pressed. This button is implemented on the Interface to simulate the Emergency Stop Handle that is available throughout the entire assembly line as a last minute resource if the worker feels that he will not be able to complete all the tasks in time and the whole system must freeze. For so there are two complementary connections between the **Interface** and the **Process Component**:

1) The **Interface** indicates that the emergency button was clicked to STOP;
2) The **Interface** indicates that the emergency button was clicked to RESUME

V. Discussion

With the developed support interface, the main goal is to affect in a positive manner the way the assembly line workers do their job and this way, as a result of the workers improvement, contribute for the success of the business in which it is inserted. This interface was developed having in mind the workers, their cognitive effort and their expertise level as these are two of many other factors that influence the worker's performance. With the Cognitively Light approach to the interface the main goal is to reduce the user's precious effort when reading from it and this way keep them more focus throughout the whole working shift which itself is mentally and physically demanding. The Expertise side of this solution brings a huge advantage to the workflow of the production line as a whole as it helps new employees to better integrate the production line and the same way bring the best of the expert worker's performance by adjusting the way the information is presented making it easy for novices and fast for experts.

With this solution implementation in a production line environment there are a group of aspects that are expected to improve either directly or indirectly. Are considered **Direct improvements** any positive aspect related with the production line workers and **Indirect improvements** any better aspect associated with the factory and company in which this solution is implemented. Here are the expected positive topics:

Direct improvements:

- Higher motivated workers
- More focus workers
- More energized workers

Indirect improvements:

- Faster production
- Less waste
- Happier clients
- More production variety

The presented interface was designed and developed having in mind that it could be used by any type of user with full mental and physical capabilities, regardless their gender, age our ability with technology and that is why the it was thought to have the least input from the user as possible, making it an almost read only interface.

A. Restrictions/Frailties of the CPS

As everything in the world, this CPS also has a few restrictions and frailties from each component on the system that limit the tests and consequently delay the evolution process demanding a higher number of tests and participants since the tests complexity is also limited.

Interface:

- Requires user input from external mouse controller;

Robotic Arm:

- Maximum payload weight of 5 kilograms;
- Maximum reach 0,85 meters;

Artificial Vision system:

- Only able to detect Yellow, Green and Blue objects;
- Not able to distinguish objects by their shape;

The described restriction are related to the current state of the Cyber Physical System (CPS) which could be improved with extra implementations.

VI. Conclusion

Expertise and cognitive levels are two of many aspects that influence a worker's performance and with the presented article we are introduced to a new approach for the support interface that guides the workers throughout the assembly process that helps to reduce the cognitive level and adapt to the expertise and this way help to grow the business. By bringing the worker to the CPPS and enabling a digital communication between him and the rest of the system it is possible to adapt the machines to the humans. Throughout this article, the reader is presented with the design of this support interface and a proposal of a CPS in which the human is part of.

References

[1] B. J. Pine, B. Victor, and A. C. Boynton, "Making mass customization work," Harvard business review, vol. 71, no. 5, 1993, pp. 108–11.

[2] M. Sondalini. Unearth the answers and solve the causes of human error in your company by understanding the hidden truths in human error rate tables. Available at https://www.lifetime-reliability.com/cms/tutorials/reliability-engineering/human_error_rate_table_insights/.

[3] F. L. Association. Examining the impact of long hours on factory workers. Available at http://www.fairlabor.org/blog/entry/examining-impact-long-hours-factory-workers. [retrieved: September, 2011]

[4] A. Burney. The life of a chinese factory worker. Available at https://woman.thenest.com/life-chinese-factory-worker-9239.html.

[5] L. Neto, A. L. Madsen, R. Silva, J. Reis, P. McIntyre, and G. Gonçalves, "A component framework as an enabler for industrial cyber physical systems," 2018.

[6] L. Antão, "Cooperative human-machine interaction in industrial environments," Ph.D. dissertation, Faculdade de Engenharia da Universidade do Porto, 2017.

[7] L. Antão, R. Pinto, J. Reis, G. Gonçalves, and F. L. Pereira, "Cooperative human-machine interaction in industrial environments," 2018.

[8] S. J. Hu, "Evolving paradigms of manufacturing: From mass production to mass customization and personalization," Proceedings of the Third Conference on Object-Oriented Technologies and Systems, 2013.

[9] I. Graessler and A. Poehler, "Integration of a digital twin as human representation in a scheduling procedure of a cyber-physical production system," 2017.

[10] F. Ansari, M. Khobreh, U. Seidenberg, and W. Sihn, "A problem-solving ontology for human-centered cyber physical production systems," August 2018.

[11] C. Fuchs, S. Ferreira, G. Gonçalves, and J. Sousa, "Adaptive consoles for supervisory control of multiple unmanned aerial vehicles," 2013.

[12] J. Johnson, "Designing with the mind in mind: Simple guide to understanding user interface design rules," 2010.

[13] S. Bradley. 11 ways to add depth to a design. Available at https://vanseodesign.com/web-design/pictorial-depth-cues/. [retrieved: January, 2012]

[14] W. Zieliński. How to use contrast in ui design. Available at https://blog.prototypr.io/how-contrast-works-in-ui-design-21bf75a5a2bf. [retrieved: December, 2016]

[15] J. Kliever. Designing with contrast: 20 tips from a designer. Available at https://www.canva.com/learn/contrasting-colors/.

[16] J. Reis, R. Pinto, and G. Gonçalves, "Human-centered application using cyber-physical production system," 2017.