Adaptive Multimedia Indexing Using Naïve Bayes Classification

Clement H. C. Leung
School of Science and Engineering
Chinese University of Hong Kong
Shenzhen, China
clementleung@cuhk.edu.cn

R. F. Ma & Jiayan Zhang
School of Data Science
Chinese University of Hong Kong
Shenzhen, China

***Abstract*— Current organizations increasingly depend on multimedia document repositories for their effective operation. However, unlike text-oriented objects, the retrieval of multimedia objects is often inhibited by limitations in their search and discovery mechanisms, since they do not readily lend themselves to automatic processing or indexing. Here, we describe the structure of an adaptive search mechanism which is able to overcome such limitations. The basic framework of the adaptive search mechanism is to capture human judgment in the course of normal usage from user queries in order to develop semantic indexes which link search terms to media objects semantics. This approach is particularly effective for the retrieval of such multimedia objects as images, sounds, and videos, where a direct analysis of the object features does not allow them to be linked to search terms, such as non-textual/icon-based search, deep semantic search, or when search terms are unknown at the time the media repository is built. An adaptive indexing mechanism is described which makes use of naïve Bayes classification approach. This approach allows for the efficient organizational creation and updating of media indexes, which is able to instill and propagate deep knowledge relating to the organizational functions into the media management system concerning the advanced search and usage of multimedia resources. The present learning approach will enable intelligent search of multimedia resources that are otherwise hard to be located and retrieved.**

Keywords – multimedia information indexing, reinforcement learning; multi-agent; naïve Bayes classifiers; stochastic game; probability generating function.

I. INTRODUCTION

Information search and retrieval has extended from textual based to multimedia content, with the characteristic of information search and retrieval shifting from pull to push applications. Instead of searching an accurate piece of information in a database, users are given selected choices of information [18]. In addition, affective indexing of multimedia content combines emotional responses generated by the users is sometimes employed, e.g. the psycho-physiological signals, galvanic skin response, face tracking, etc, [19].

There is now general consensus that involving users in the information search and retrieval process is able to improve the overall return results [22]. In [23], it is shown that using Markov decision process improves the efficiency of locating video frames in a video, and in [24], the distribution of visual words of multimedia data is found to be probabilistic in relation to the concept relationship formed [24]. Users often allocate the results based on some form of scoring metrics; for example, a linear combination of posterior probability is employed to refine the search results [25]. In [20], it is proposed that Reinforcement Learning (RL) approach is suitable for users exposing to raw and high-dimensional information [20], while instant rewards of the agents is generally able to impart significant improvements in the searching process [21]. In Reinforcement Learning (RL), an agent learns through the interaction with the dynamic environment to maximize its long-term rewards, in order to act optimally. Most of the time, when modeling real-world problems, the environment involved is non-stationary and noisy [1][4][6]. More precisely, the next state results from taking the same action in a specific state may not necessarily be the same but appears to be stochastic [2][7]. And the exploration strategies adopted in different categories of RL algorithms provide different levels of control to the exploration of unknown factors, which in turn give various possibilities to the learning results.

As a result, the observed rewards and punishments are often non-deterministic. For example, when one is trying to find a video for performing a particular task, a shortening of the searching time with respect to some anticipated norm may be regarded as a reward, while a lengthening of the same may be viewed as punishment. Likewise, when one is exploring a new advertising channel, a resultant significant increase in sales may be viewed as a reward, while failure to do so may be regarded as punishment. In situations like these, there are stochastic elements governing the underlying environment. In the new route to work example, whether one receives rewards or punishments depends on a variety of chance factors, such as weather condition, day of the week, and whether there happens to be road works or traffic accidents which may or may not be representative.

Noise in multimedia data is generally numerous and cannot be known or enumerated in a practical sense, and this tends to mask the underlying pattern. Indeed, if stochastic elements are absent, the learning problems involved could be greatly simplified and their presence has motivated early research in the area. As early as 1990s, mainstream research in RL, such as the influential survey assessing existing methods carried out by Kaelbling *et al.* [2], and the Explicit Explore or Exploit (E^3) Algorithm to solve Markov Decision Process (MDP) in polynomial time [3], adopts the common assumption of a stationary environment within a RL framework. Later on, with further advances in RL, theoretical analyses addressing the concern of non-stationary environment attracted great interests. One of the works by

 ISBN: 978-1-61208-812-9

Brafman and Tennenholtz introduces a model-based RL algorithm R-Max to deal with stochastic games [5]. Such stochastic elements can notably increase the complexity in multi-agent systems and multi-agent tasks, where agents learn to cooperate and compete simultaneously [6][10]. Autonomous agents are required to learn new behaviors online and predict the behaviors of other agents in multi-agent systems. As other agents adapt and actively adjust their policies, the best policy for each agent would evolve dynamically, giving rise to non-stationarity [8][9].

In most of the above situations, the cost of a trial or observation to receive either a reward or punishment can be significant, and preferably, one would like to arrive at the correct conclusion by incurring minimum cost. In the case of the advertising example, the cost of advertising can be considerable and one would therefore like to minimize it while acquiring the knowledge whether such advertising channel is effective. Similarly, in RL algorithms, we are always in the hope to rapidly converge to an optimal policy with least volumes of data, calculations, learning iterations, and minimal degree of complexity [11][12]. To do so, one should explicitly define the stopping rules for specifying the conditions under which learning should terminate and a conclusion drawn as to whether the learning has been successful or not based on the observations so far.

The problem of finding termination conditions, or stopping rules, is an intensive research topic in RL, which is closely linked to the problems of optimal policies and policy convergence [13]. Traditional RL algorithms mainly aim for relatively small-scale problems with finite states and actions. The stopping rules involved are well-defined for each category of algorithms, such as utilizing Bellman Equation in *Q*-learning [14]. To deal with continuous action spaces or state spaces, new algorithms, such as the Cacla algorithm [15] and CMA-ES algorithm [16], are developed with specific stopping criteria. Still, most studies on stopping criteria are algorithm-oriented and do not have a unified measurement for general comparison.

In this paper, we present an approach to RL by using a naïve Bayes classification framework, which explicitly incorporates the stochastic aspects of the environment in multimedia information search and retrieval. Applying naïve Bayes methods for classification problems are often employed in a variety of contexts [26][27], such as crowdsourcing and police surveillance. Here, we shall also learn and estimate the underlying stochastic structure of the environment by making use of the random classification labels gathered in the course of the learning process. Section II presents the fundamental model of a predefined general learning policy. The information search and retrieval success based on the rewards ratio is then studied in Section III. Based on the stochastic model, Section IV analyzes the probability of exceeding cost bounds. Section V views the relative occurrences of the binary classifications from the perspective of competing multi-agents, and the final conclusions are drawn in Section VI.

II. A PROBABILISTIC LEARNING FRAMEWORK WITH A FIXED NUMBER OF LABELS

We are concerned with a learning sequence of multimedia search and retrieval observations, each of which either results in a positive classification or negative classification. That is, we are dealing with a binary classification problem with two class labels, +1 or −1, where for convenience the former is referred to as success, and the latter, failure. Such a learning sequence in the present context corresponds to the proper association of given search terms to particular multimedia objects. We are interested in determining whether the sequential classifications indicate overall success or failure in the classification process. Evidently, if the number of +1 labels gathered is much greater than the number of −1 labels, then the conclusion drawn from the learning episode should be success, while if the opposite is true, then the corresponding conclusion should be failure. In the case of search terms to multimedia objects association, learning success would mean that the association in question is sound and should be incorporated as proper index, while failure would mean that the search term-object association cannot be established. In order to proceed with the analysis, we first let p and q (with $p + q = 1$) denote the probabilities of receiving a +1 or −1 label respectively for a given classification. Furthermore, we shall make use of the naïve Bayes property that different classifications are independent of each other. Later on, we shall derive estimates for p and q, which capture the stochastic structure of the learning environment. For example, if $p > q$, then clearly the final conclusion should be learning success. An error often committed is that when the first few observations are all −1, one would terminate prematurely and return a verdict of failure for the learning episode. Let us consider the following learning policy; such a policy is also studied in [26, 27] and is called majority voting.

Learning Policy I*: On gathering a total of r labels all belonging to either +1 or −1, the learning terminates and a decision is made in accordance with the accepted margin of the majority of voting of the classifiers.*

Here, we let the random variable $\boldsymbol{T}$ represent the number of classification labeling preceding the first positive classification; i.e. $\boldsymbol{T}$ may be viewed as the waiting time to the first positive classification,

$$\Pr[\boldsymbol{T} = k] = pq^k, \quad k = 0, 1, 2, 3, \ldots \tag{1}$$

The probability generating function $G(z)$ of $\boldsymbol{T}$ is given by

$$G(z) = \sum_{k=0}^{\infty} \Pr[\boldsymbol{T} = k]\, z^k = p \sum_{k=0}^{\infty} q^k z^k = \frac{p}{(1 - qz)}. \tag{2}$$

Note that after the occurrence of the first positive classification, the process probabilistically repeats itself again, so that we have for the waiting time $\boldsymbol{W_r}$ of the rth positive classification

$$\boldsymbol{W}_r = \sum_{k=1}^{r} \boldsymbol{T}_k, \quad (3)$$

where each $\boldsymbol{T_k}$ has the same distributional characteristics as $\boldsymbol{T}$. From [17], the probability generating function of $G_r(z)$ corresponding to $\boldsymbol{W_r}$ may be obtained

$$G_r(z) = G_1(z)^r = [\frac{p}{(1-qz)}]^r . \quad (4)$$

To gain a better understanding of behavior specified above, it is useful to obtain the average waiting time $\boldsymbol{W_r}$ and its variance when r positive labels are attained. From (4), the mean and variance of $\boldsymbol{W_r}$ can be derived

$$\mathrm{E}[\boldsymbol{W_r}] = G_r'(1) = \frac{rq}{p} , \quad (5)$$

$$\mathrm{Var}[\boldsymbol{W_r}] = G_r''(1) + G_r'(1) - G_r'(1)^2 = \frac{rq}{p^2} . \quad (6)$$

Furthermore, the probabilities $\Pr[\boldsymbol{W_r} = k]$ may be readily obtained from the expansion of (4) so as to study the probabilities for various waiting time,

$$\Pr[\boldsymbol{W_r} = k] = \binom{-r}{k} p^r (-q)^k, \quad k = 0, 1, 2, 3, \ldots \quad (7)$$

As $\boldsymbol{W_r}$ is the sum of independent identically distributed random variables, when r is appreciable, it may be approximated by the normal distribution [17]

$$\boldsymbol{W_r} \sim N\left(\frac{rq}{p}, \frac{rq}{p^2}\right), \quad (8)$$

whence we have, denoting by Φ the standard normal distribution,

$$\Pr[\boldsymbol{W_r} > b] = \int_{\frac{bp-rq}{\sqrt{rq}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = 1 - \Phi\left(\frac{bp-rq}{\sqrt{rq}}\right). \quad (9)$$

III. Learning Success Based on The Class Label Ratio

Let ρ be the ratio of the average number of negative labels to the number of positive labels, we have

$$\rho(p) = \frac{E[\boldsymbol{W_r}]}{r} = \frac{q}{p} . \quad (10)$$

From this, we determine the inherent stochastic structure of the environment by estimating p from actual observed labels ratio W/r, where W is the sample mean of $\boldsymbol{W_r}$. We can form our estimator from the above just by solving for p. We shall estimate the probability P_b that the learning cost for this component exceeding this bound. From (7) above, this is given by

$$P_b = 1 - \sum_{k=0}^{b} \Pr[\boldsymbol{W_r} = k] = 1 - \sum_{k=0}^{b} \binom{-r}{k} p^r (-q)^k . \quad (11)$$

Here, the normal approximation can be invoked. In many RL learning episodes, r tends to be under 100, as a lengthy iteration time is not feasible and most learning algorithms aim to converge in minimum time.

Clearly, the selection of the maximum cost weight b will have a significant impact on P_b. Very often, it is more meaningful to relate b to $\mathrm{E}[\boldsymbol{W_r}]$ either additively or multiplicatively. Table I tabulates the values of P_b for different values of b. The first part of Table I considers b by adding a fixed value d, with $d = 5$ and $d = 10$, while the second part considers b by multiplying by a fixed multiple α, with $\alpha = 1.2$ and $\alpha = 1.5$; here, b is rounded to the nearest integer. In the first part of Table I, we see that for either value of r, when p is appreciably greater than q, the probability of exceeding cost bounds tends to be acceptably small, and this is especially so for $r = 20$. The reason is that, since d is a fixed value, its relative contribution to b increases as p increases, produces a relatively large cost bound weight compared to the average one, and accordingly lowers the probability of exceeding the bound. However, in the second part of Table I, the difference between $\mathrm{E}[\boldsymbol{W_r}]$ and b decreases as $\mathrm{E}[\boldsymbol{W_r}]$ decreases, so that P_b

TABLE I. ANALYSIS OF PROBABILITIES OF EXCEEDING COST BOUNDS

***b* Formula**	***r***	***p***	***q***	**E[W_r]**	***b***	***P_b***	***P_b'***	***Err***
***b* = *E[W_r] + d* (*d* = 5)**	20	0.5	0.5	20.00	25	0.215	0.186	0.029
		0.8	0.2	5.00	10	0.023	0.026	0.003
		0.9	0.1	2.22	7	0.001	0.004	0.003
	50	0.5	0.5	50.00	55	0.309	0.279	0.030
		0.8	0.2	12.50	17	0.127	0.108	0.019
		0.9	0.1	05.56	11	0.014	0.017	0.003
***b* = *E[W_r] + d* (*d* = 10)**	20	0.5	0.5	20.00	30	0.057	0.059	0.002
		0.8	0.2	5.00	15	0.000	0.001	0.001
		0.9	0.1	2.22	12	0.000	0.000	0.000

 ISBN: 978-1-61208-812-9

b Formula	r	p	q	E[W_r]	b	P_b	P_b'	Err
	50	0.5	0.5	50.00	60	0.159	0.147	0.012
		0.8	0.2	12.50	22	0.008	0.011	0.003
		0.9	0.1	05.56	16	0.000	0.000	0.000
$b = \alpha E[W_r]$ ($\alpha = 1.2$)	20	0.5	0.5	20.00	24	0.264	0.226	0.038
		0.8	0.2	5.00	6	0.345	0.253	0.092
		0.9	0.1	2.22	2	0.556	0.380	0.176
	50	0.5	0.5	50.00	50	0.159	0.147	0.012
		0.8	0.2	12.50	15	0.264	0.215	0.049
		0.9	0.1	05.56	7	0.280	0.207	0.073
$b = \alpha E[W_r]$ ($\alpha = 1.5$)	20	0.5	0.5	20.00	30	0.057	0.059	0.002
		0.8	0.2	5.00	7	0.212	0.156	0.056
		0.9	0.1	2.22	3	0.310	0.193	0.117
	50	0.5	0.5	50.00	75	0.006	0.010	0.004
		0.8	0.2	12.50	19	0.050	0.048	0.002
		0.9	0.1	05.56	8	0.163	0.121	0.042

tends to be large for higher values of p.

In Table I, column P_b' gives the exact calculation using (11), while column P_b employs the normal approximation using (9). The absolute error between the exact calculation and the normal approximation is given by column *Err*. We see that the normal approximation is quite acceptable in most cases with absolute error less than 0.1. Note that no matter whether having b additively or multiplicatively related to E[W_r], a higher value of d or α always gives smaller absolute error. We therefore suggest that the approximation should only be used when r, d and α are sufficiently large.

IV. MULTI-AGENT LEARNING

In *Learning Policy I* above, the termination of a learning episode is triggered whenever a fixed number of positive labels r is obtained, irrespective of the number of negative labels accumulated in the process of doing so. Sometimes, however, this may not be desirable, especially when an inordinate number of negative labels have been accumulated, in which case, termination should take place earlier along with the conclusion of learning failure. Therefore, one is comparing the number of positive labels gathered against the number of negative labels, and the learning is concluded as success or failure according to which of these achieve the majority.

More precisely, this may be viewed as a multi-agent tournament with two competing agents A and B, in which A is responsible for giving out the positive labels, while B, the negative labels. This framework is not unlike the game theoretic approach in statistical decision theory, where both the statistician and nature are regarded as players in the game of estimation, and also this may be regarded as a kind of stochastic game [5]. While we shall focus on the agents A and B, we note that there is a further agent, the learner, so that three agents exist in this situation. Here, when a classification results in a positive labels, then A would gain a score of one, while when an observation results in a negative labels, then B would gain a score of one. When either ±1 label first reaches a given threshold h, then this will trigger a termination and the learning episode is concluded as success or failure according to which agent attains the threshold score first. Therefore, we have the following stopping rule:

Learning Policy II*: The learning process terminates when either agent, A or B, first reach the threshold of accumulating h +1 or −1 classifications, which can be concluded as a success or a failure according to which agent attains the threshold first.*

Here, without loss of generality, we shall let $h = 2m+1$ be odd, where m is an integer, and similar to Section II, we let p and q, with $p + q = 1$, signify the probabilities of receiving a positive labels, and negative labels, respectively for a particular classification. In other words, for a given classification, agent A wins with probability p, while agent B wins with probability q. In order to attain h for either agent, the number of classifications Ω will fall within the range

$$2m + 1 \le \Omega \le 4m + 1 .$$

If f_k represents the probability that A wins at classifications number $4m+1-k$, which occurs if and only if A scored $2m$ successes in the first $4m-k$ observations, and subsequently score a final success, then f_k is given by

$$f_k = \binom{4m-k}{2m} p^{2m+1} q^{2m-k} .$$

The probability that A reaches the threshold first, irrespective of the classification number, is therefore given by

$$P_m = \sum_{k=0}^{2m} f_k = \sum_{k=0}^{2m} \binom{4m-k}{2m} p^{2m+1} q^{2m-k} .$$

That is, P_m gives the probability that the learning is successful (i.e. agent A wins) according to *Rule B*.

Table II computes P_m for different values of p, q, and m. We see that, as expected, when $p = q = 1/2$, $P_m = 1/2$, since neither A nor B has any advantage over its opponent. As p increases, however, P_m will increase, reaching almost certainty as p increases beyond 0.8. If we regard p as a

TABLE II. PROBABILITIES OF LEARNING SUCCESS

m	p	q	P_m	m	p	q	P_m
1	0.5	0.5	0.5000	5	0.5	0.5	0.5000
	0.6	0.4	0.6826		0.6	0.4	0.8256
	0.7	0.3	0.8369		0.7	0.3	0.9736
	0.8	0.2	0.9421		0.8	0.2	0.9990
	0.9	0.1	0.9914		0.9	0.1	1.0000
2	0.5	0.5	0.5000	10	0.5	0.5	0.5000
	0.6	0.4	0.7334		0.6	0.4	0.9035

	0.7	0.3	0.9012		0.7	0.3	0.9964
	0.8	0.2	0.9804		0.8	0.2	1.0000
	0.9	0.1	0.9991		0.9	0.1	1.0000

measure of *A*'s winning ability per trial, then when $p >> q$, most trials will be scored by *A*, so that winning the entire game (i.e. reaching *h* first) is almost a certainty, and this is especially so for higher values of *h*. It is interesting to see that when *h* or *m* is sufficiently high (e.g. *m*=10), a moderate advantage for *A* (e.g. $p = 0.6$) is enough to almost guarantee success. On the other hand, $1-P_m$ gives the probability that agent *B* wins, where the measure of *B*'s winning probability per trial is given by *q*. For instance, when *q*=0.4, then *B* stands a chance of around 27% of winning the game when *m*=2, and a chance of winning of around 10% when *m*=10.

Returning to the estimation problem, by observing P_m, i.e. by computing the observed proportion of time that agent *A* wins, it is possible to infer the underlying probability *p*. While unlike in Section II, where an explicit formula exists linking directly the observations to the estimate, such explicit relationship is not available here. Nevertheless, as can be observed from Table II, useful estimation bounds can be drawn to determine whether $p > ½$ or $p < ½$. We see that it is quite reasonable to estimate $\hat{p} > ½$ whenever $P_m > ½$, and this would seem sufficient for most practical purposes.

V. CONCLUSION

Since multimedia information search environments are often noisy and seldom static nor deterministic, the use of stochastic methods is therefore an unavoidable necessity. Indeed, if stochastic elements are absent, the same outcome will always occur, obviating the need for repeated observations.

In this paper, we first consider a situation where the cumulative number of classifications is pre-specified and fixed, which constitute the criterion for stopping the learning process. By observing the random positive to negative labels ratio, a meaningful estimation of either learning success or failure may be arrived at. In most practical situations, the cost of securing a classification can be significant, and this has been incorporated into our model, with the probabilities of exceeding the classifications cost bounds also derived.

We also consider a multi-agent framework where the handing out of positive and negative labels are viewed as being performed by agents. Thus, the final learning outcome is determined by a kind of stochastic game with the agents competing against each other. The termination criterion here is determined by when and how the game is won. The respective probabilities of learning success and failure are also explicitly derived. Closed-form expressions of other relevant measures of interest are obtained. A procedure for estimating the underlying stochastic structure from the observed random agent winning frequencies is also employed.

In this study, we have adopted the naïve Bayes assumption and assumed that positive labels and negative labels occur independently. In future, it may be useful to relax this assumption and incorporate single-step or multi-step Markov dependency into the analysis. It is likely, however, that the corresponding estimation procedures will be considerably more involved.

REFERENCES

[1] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum Entropy Inverse Reinforcement Learning," *Proc. Twenty-Third AAAI Conference on Artificial Intelligence (AAAI 08)*, vol. 8, pp. 1433-1438, 2008.

[2] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey" *Journal of artificial intelligence research*, vol. 4, pp. 237-285, 1996.

[3] M. Kearns and S. Singh, "Near -optimal reinforcement learning in polynomial time," In *Int. Conf. on Machine Learning*, 1998.

[4] H. Santana, G. Ramalho, V. Corruble, and B. Ratitch, "Multi-agent patrolling with reinforcement learning," *Proc. Third International Joint Conference on Autonomous Agents and Multiagent Systems*, vol. 3, pp. 1122-1129, IEEE Computer Society, 2004.

[5] R. I. Brafman and M. Tennenholtz, "R-max-a general polynomial time algorithm for near-optimal reinforcement learning," *Journal of Machine Learning Research*, vol. 3, pp.213-231, 2002.

[6] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous agents and multi-agent systems*, vol. 11, no. 3, pp. 387-434, 2005.

[7] E. Ipek, O. Mutlu, J. F. Martínez, and R. Caruana, "Self-optimizing memory controllers: A reinforcement learning approach," *ACM SIGARCH Computer Architecture News*, vol. 36, no. 3, IEEE Computer Society, 2008.

[8] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems, Man, And Cybernetics-Part C: Applications and Reviews*, vol. 38, no. 2 2, 2008.

[9] S. V. Albrecht, and P. Stone, "Autonomous agents modelling other agents: A comprehensive survey and open problems," *Artificial Intelligence* 258, pp. 66-95, 2018.

[10] A. Tampuu, T. Matiisen, D. Kodelja, I. Kuzovkin, K. Korjus, J. Aru, and R. Vicente, "Multiagent cooperation and competition with deep reinforcement learning," *PloS one*, vol. 12, no. 4: e0172395, 2017.

[11] A.W. Moore and C.G. Atkeson, "Prioritized sweeping: Reinforcement learning with less data and less time," *Machine learning*, vol. 13, no.1, pp. 103-130, 1993.

[12] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning,", unpublished.

[13] Q. Wei, F. L. Lewis, Q. Sun, P. Yan, and R. Song, "Discrete-time deterministic Q-learning: A novel convergence analysis," *IEEE transactions on cybernetics*, vol. 47, no. 5, pp. 1224-1237, 2017.

[14] C. J. Watkins and P. Dayan, "Q-learning," Machine learning 8.3-4 pp. 279-292, 1992.

[15] H. Van Hasselt and M.A. Wiering, "Using continuous action spaces to solve discrete problems," *Proc. International Joint Conference on Neural Networks (IJCNN 09)*, pp. 1149-1156. IEEE, 2009.

[16] N. Hansen, S. D. Müller, and P. Koumoutsakos, "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)," *Evolutionary computation*, vol. 11, no. 1 pp. 1-18, 2003.

[17] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 1, 3rd Edition, Wiley & Sons, 1968.

[18] Q. Huang, A. Puri, Z. Liu, "Multimedia search and retrieval: new concepts, system implementation, and application". *IEEE transactions on circuits and systems for video technology*, 2000, 10.5: 679-692.

[19] R.Gupta, M. Khomami Abadi, J. A.Cárdenes Cabré, F. Morreale,T. H. Falk, and N. Sebe, "A quality adaptive multimodal affect recognition system for user-centric multimedia indexing". In: Proceedings of the 2016 ACM on international conference on multimedia retrieval. ACM, p. 317-320, 2016.

[20] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A.,"Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, 34(6), 26-38, 2017.

[21] X. Yao, J. Du, N. Zhou, and C. Chen, "Microblog Search Based on Deep Reinforcement Learning," *In Proceedings of 2018 Chinese Intelligent Systems Conference* (pp. 23-32). Springer, Singapore, 2019.

[22] Y.C. Wu, T. H.Lin, Y. D. Chen, H. Y Lee, and L. S. Lee, "Interactive spoken content retrieval by deep reinforcement learning". arXiv preprint arXiv:1609.05234, 2016.

[23] S. Lan, R. Panda, Q. Zhu, and A. K. Roy-Chowdhury, "FFNet: Video fast-forwarding via reinforcement learning", *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6771-6780, 2018.

[24] R. Hong, Y. Yang, M. Wang, and X. S. Hua, "Learning visual semantic relationships for efficient visual retrieval", *IEEE Transactions on Big Data*, 1(4), pp.152-161, 2015.

[25] R. Yan, A. Hauptmann, and R. Jin,"Multimedia search with pseudo-relevance feedback. In International Conference on Image and Video Retrieval" (pp. 238-247). *Springer*, Berlin, Heidelberg, 2003.

[26] E. Manino, L. Tran-Thanh, and N. R. Jennings. On the Efficiency of Data Collection for Multiple Naïve Bayes Classifiers. Artificial Intelligence, 275: 356–378, 2019.

[27] E. Manino, L. Tran-Thanh, and N. R. Jennings. On the Efficiency of Data Collection for Crowdsourced Classification. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, pp. 1568-1575, 2018.

 ISBN: 978-1-61208-812-9