A GENERAL NAIVE BAYES STYLE FUZZY PROBABILISTIC CLASSIFIER

Ronei Marcos de Moraes[1]

Abstract *— Some kinds of naive Bayes style networks have been proposed, such as the multinomial naive Bayes, possibilistic naive Bayes and fuzzy Gaussian naive Bayes. However, a general formulation for a naive Bayes style fuzzy probabilistic network was not proposed yet. In this paper, we proposed a formulation for this kind of supervised classifier, using random variables without specifying any statistical distributions. This approach can be useful for classification purposes, when random variables can have different statistical distributions. A brief discussion about applications for data classification from health sciences is provided too.*

Index Terms — Classification, Fuzzy Probability, Fuzzy Sets, Naive Bayes

INTRODUCTION

There are several kinds of Naive Bayes Networks for classification purposes, as: multinomial naive Bayes, possibilistic naive Bayes and fuzzy Gaussian naive Bayes. They are applied in the solution of a large number of problems in various areas of human knowledge, such as medical diagnostic decision support [27][28], classification of injury narratives [29], training assessment [12][13][14][15][16][17][18], decision making [8], medical diagnosis [7], machine learning [10], speech recognition [20] and others [1][5][24][25][26].

A general formulation which can unify the probabilistic and fuzzy version of Naive Bayes classifier using probability of fuzzy events was not found in the literature. This paper has as goal present present a possible formulation for that.

THEORECTICAL REVIEW

In this Section, theorectical aspects of some kinds of naive Bayes style networks are presented: classical naive Bayes and the versions for Gaussian distribution, Gaussian mixture models and modified naive Bayes. Fuzzy versions of naive Bayes are presented too. These formulations are presented in order to facilitate the understanding of the generalized version that will be presented in the next section.

Naive Bayes Method

Formally, let be the classes of performance in space of decision $\Omega=\{1,...,M\}$ where M is the total number of classes of performance. Let be w_i, $i \in \Omega$ the class of performance for an user. A Naive Bayes classifier [3] computes conditional class probabilities and then predict the most probable class of a vector of training data X, according to sample data D, where X is a vector with n features obtained when a training is performed, i.e. $X=\{X_1, X_2, ..., X_n\}$. Using the Bayes Theorem [21]:

$$\begin{aligned} &P(w_i \mid X) = [P(X \mid w_i)\, P(w_i)] / P(X) \Leftrightarrow \\ &\Leftrightarrow P(w_i \mid X_1, X_2, ..., X_n) = \\ &= [P(X_1, X_2, ..., X_n \setminus w_i)\, P(w_i)] / P(X) \end{aligned} \quad (1)$$

However, as $P(X)$ is the same for all classes w_i, then it is not relevant for data classification and can be rewritten as:

$$P(X \mid w_i)\, P(w_i) = P(X_1, X_2, ..., X_n \setminus w_i)\, P(w_i) \quad (2)$$

The equation (2) is equivalent to the joint probability model:

$$P(X_1, X_2, ..., X_n \setminus w_i)\, P(w_i) = P(X_1, X_2, ..., X_n , w_i) \quad (3)$$

Now, using successive applications of the conditional probability definition over equation (3), can be obtained:

$$\begin{aligned} &P(X_1, X_2, ..., X_n , w_i) = P(w_i)\, P(X_1, X_2, ..., X_n \setminus w_i) = \\ &= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2, ..., X_n \setminus w_i , X_1) \\ &= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i , X_1)\, P(X_3, ..., X_n \setminus w_i , X_1, X_2) \\ &... \\ &= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i , X_1) ... P(X_n \setminus w_i , X_1, X_2 , ...,X_{n-1}) \end{aligned}$$

The Naive Bayes classifier receives this name because its naive assumption of each feature X_k is conditionally independent of every other feature X_l , for all $k \neq l \leq n$. It means that knowing the class is enough to determine the probability of a value X_k. This assumption simplifies the equation above, due to:

$$P(X_k \setminus w_i , X_l) = P(X_k \setminus w_i) \quad (4)$$

for each X_k and the equation (3) can be rewritten as:

$$\begin{aligned} &P(X_1, X_2, ..., X_n , w_i) = \\ &= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i) ... P(X_n \setminus w_i) \end{aligned} \quad (5)$$

unless a scale factor S, which depends on $X_1, X_2, ..., X_n$. Finally, equation (1) can be expressed by:

$$P(w_i \mid X_1, X_2, ..., X_n) = (1/S)\, P(w_i)\, \Pi^n{}_{k=1}\; P(X_k \setminus w_i) \quad (6)$$

Then, the classification rule for Naive Bayes is done by:

$$\begin{aligned} &X \in w_i \;\; if \;\; P(w_i \mid X_1, X_2, ..., X_n) > P(w_j \mid X_1, X_2, ..., X_n) \\ &\text{for all } i \neq j \text{ and } i, j \in \Omega \end{aligned} \quad (7)$$

[1] Ronei Marcos de Moraes, Department of Statistics, Federal University of Paraíba. Cidade Universitária s/n, Castelo Branco, João Pessoa-PB. 58.051-900. Brazil. *ronei@de.ufpb.br*

and $P(w_* \mid X_1, X_2, \ldots, X_n)$ with $* = \{i, j \mid i, j \in \Omega\}$, is done by (6).

It is important to notice this formulation is general. The equation (6) can be used with several kinds of random variable X. in particular, X can assume multinomial distribution or it can be a discrete random variable.

To estimate parameters for $P(X_k \setminus w_i)$ for each class i, it was used a maximum likelihood estimator, named P_e:

$$P_e(X_k \setminus w_i) = \#(X_k, w_i) / \#(w_i) \qquad (8)$$

where $\#(X_k, w_i)$ is the number of sample cases belonging to class w_i in all sample data D and having the value X_k, $\#(w_i)$ is the number of sample cases that belong to the class w_i in all sample data D.

Gaussian Naive Bayes Method

As mentioned above, the NB Method must be applied over discrete or multinomial variables. However, its independence assumptions can be applied on variables, which can assume other statistical distributions. Some approaches were developed to use NB Method with continuous variables, as several discretization methods [7][30] were used in the first stage to allow the use of the Naive Bayes method after.

However, this approach can affect classification bias and variance of the NB method [31]. Other approach is use Gaussian distribution for X and to compute its parameters from D, i.e., mean vector and covariance matrix [10]. From equation (6) and using some mathematical simplification, it is possible to reduce computational complexity of that equation [15]:

$$\begin{aligned} &\log [P(w_i \mid X_1, X_2, \ldots, X_n)] \\ &= \log [(1/S)\, P(w_i)\, \Pi^n{}_{k=1}\ P(X_k \setminus w_i)] \\ &= \log (1/S) + \log P(w_i) + \Sigma^n{}_{k=1}\ \log[P(X_k \setminus w_i)] \qquad (9) \end{aligned}$$

As S is a scale factor, it is not necessary be computed in classification rule for GNB. Then:

$$\begin{aligned} &X \in w_i \quad if\ \{\log P(w_i) + \Sigma^n{}_{k=1}\ \log[P(X_k \setminus w_i)]\} > \\ &> \{\log P(w_j) + \Sigma^n{}_{k=1}\ \log[P(X_k \setminus w_j)]\} \\ &\text{for all } i \neq j \text{ and } i, j \in \Omega \qquad (10) \end{aligned}$$

Based on the same space of decision with M classes, all conditional class probabilities are computed and then predicts the most probable class of a vector of training data X, according to sample data D. The parameters are estimated from data and the conditional probabilities are estimated using equation (9) and the final decision about vector of training data X is done by equation (10) [14].

Naive Bayes Network Modeled by Gaussian Mixture Models

In several cases, statistical distributions of data cannot be assumed as Gaussian distributions (univariate or multivariate). Among others, a possible solution is modelling them using mixture models [9]. This section presents the GMM method, which is described as a maximum likelihood classifier. We follow [8][16][26] and Machado and Moraes explanations about GMM algorithm.

Let a feature X_k from previous section, where $k=\{1,\ldots,n\}$ for the class of performance w_i for an user. This feature can not assume a classical statistical distribution, but it can be modeled by a mixture of c Gaussian distributions. Thus, let $X_k = \{x_1, x_2, \ldots, x_T\}$ be a set of T vectors obtained from a feature k, which is measure during the training for that w_i. Since the distribution of these vectors is unknown, it is modeled as the weighted sum of c component densities, given by:

$$p(x_t \mid w_i, \lambda) = \Sigma^c{}_{j=1}\ \square_j\, N_{wi}(x_j, \mu_j, \Sigma_j), \qquad (11)$$

where t=1,...,T, λ denotes a prototype consisting of a set of model parameters $\lambda = \{\square_j, \mu_j, \Sigma_j\}$, $\square_j$, j=1,...,c are the mixture weights and $N_{wi}(x_j, \mu_j, \Sigma_j)$ are T-variate Gaussian component densities with mean vectors μ_j and covariance matrices Σ_j: for the class of performance w_i.

$$N_{wi}(x_t, \mu_j, \Sigma_j) = \exp\{(-1/2)(x_t - \mu_j)'\, \Sigma_j^{-1}\, (x_t - \mu_j)\} / (2\pi)^{d/2}\, |\Sigma_j|^{1/2} \qquad (12)$$

To train the GMM, those parameters are estimated such that they best match the distribution of X_k for the class of performance w_i. The maximum likelihood estimation is widely used as a training method. For a sequence of training vectors X_k for a λ, the likelihood of the GMM is done by:

$$p(X_k \mid w_i, \lambda) = \prod^T{}_{t=1}\ p(x_t \mid w_i, \lambda) \qquad (13)$$

The aim of maximum likelihood estimation is to find a new parameter model $\bar{\lambda}$ such that $p(X_k \mid \bar{\lambda}) \geq p(X_k \mid \lambda)$. Since the expression in (6) is a nonlinear function of parameters in λ, its direct maximization is not possible. However, these parameters can be obtained iteratively using the Expectation-Maximization algorithm [3]. In this algorithm, an auxiliary function Q is used as follows:

$$Q(\lambda, \bar{\lambda}) = \Sigma^T{}_{t=1}\ \Sigma^c{}_{i=1}\ p(w_i \mid x_t, \lambda) \log[\bar{\square}_j\, N_{wi}(x_t, \bar{\mu}_j, \bar{\Sigma}_j)] \qquad (14)$$

where $p(w_i \mid x_t, \lambda)$ is the *a posteriori* probability for performance class w_i, i=1,...,M and satisfies

$$p(w_i \mid x_t, \lambda) = [\square_j\, N_{wi}(x_t, \mu_j, \Sigma_j)] / \{\Sigma^c{}_{k=1}\ \square_k\, N_{wi}(x_t, \mu_k, \Sigma_k)\} \qquad (15)$$

The Expectation-Maximization algorithm is such that if $Q(\lambda, \bar{\lambda}) \geq Q(\lambda, \lambda)$ then $p(X_k \mid \bar{\lambda}) \geq p(X_k \mid \lambda)$ [20]. The setting of derivatives of Q function with respect to λ to zero, found the following reestimation equations:

$$\bar{\square}_j = 1/T \sum{}^T{}_{t=1}\ p(w_i \mid x_t, \lambda) \qquad (16)$$

$$\bar{\mu}_j = \sum^{T}_{t=1} [\, p(w_i \mid x_t, \lambda) x_t \,] \,/\, [\sum^{T}_{t=1} [\, p(w_i \mid x_t, \lambda)] \quad (17)$$

$$\bar{\Sigma}_j = \{\sum^{T}_{t=1} [\, p(w_i \mid x_t, \lambda)(x_t - \bar{\mu}_j)(x_t - \bar{\mu}_j)' \} \,/\, [\sum^{T}_{t=1} [\, p(w_i \mid x_t, \lambda)] \quad (18)$$

The algorithm for training a GMM is described as follows:

- Geenerate the *a posteriori* probability $p(w_i \mid x_t, \lambda)$ at random satisfying (15);
- Compute the mixture weight, the mean vector, and the covariance matrix following (16), (17) and (18);
- Update the *a posteriori* probability $p(w_i \mid x_t, \lambda)$ according to (8) and compute the Q function using (14);
- Stop if the increase in the value of the Q function at the current iteration, relative to the value of the Q function at the previous iteration, is below a chosen threshold; otherwise go to step 2.

After parameters have been estimated, individual probabilities $p(X_k \mid \lambda)$ can be obtained from (13) using model parameters λ and for each feature k. Thus, equation (6) can be rewriten as:

$$P(w_i \mid X_1, X_2, \ldots, X_n) = (1/S)\, P(w_i)\, \Pi^{n}_{k=1}\; P(X_k \mid w_i, \lambda_k)$$
$$= (1/S)\, P(w_i)\, \Pi^{n}_{k=1}\; \Pi^{T}_{t=1}\; p(x_{tk} \mid w_i, \lambda_k)$$
$$= (1/S)\, P(w_i)\, \Pi^{n}_{k=1}\; \Pi^{T}_{t=1}\; [\, \sum^{c}_{j=1}\; ?_j\, N_{wi}\, (x_{jtk}, \mu_{jk}, \Sigma_{jk})]. \quad (19)$$

Finally, the assessment rule for Naive Bayesian Network, where each variable is modeled by a Gaussian Mixture Model is done by (6), where $P(w_i \mid X_1, X_2, \ldots, X_n)$ is given by (19).

Modified Naive Bayes

In some cases, the problem present quantitative and qualitative variables simultaneously [5] and to compute its parameters from D. Formally, let be $X_{cat}=\{X_1, X_2, \ldots, X_c\}$, c ($0 \leq c \leq n$) categorical or discrete variables obtained from training data, as in classical Naive Bayes and $X_{cont}=\{X_{c+1}, X_{c+2}, \ldots, X_n\}$, $n\text{-}c$ continuous variables, obtained from training data too. Thus, X is a vector with n features, with $X=X_{cat} \cup X_{cont}$. Now, the equation (6) can be rewritten as [3]:

$$P(w_i \mid X_1, X_2, \ldots, X_n) =$$
$$=(1/S)\, P(w_i)\, \Pi^{c}_{k=1}\; P(X_k \setminus w_i)\, \Pi^{n}_{k=c+1}\; P(X_k \setminus w_i) \quad (20)$$

The equation (20) defines the Modified Naive Bayes (MNB) method. Categorical variables can be modeled by multinomial distributions. Discrete variables can be modeled by count of events in sample data D or by a discrete statistical probability distributions. The continuous variables can be modeled by probability density functions. All these distributions models can be adjusted and verified by statistical tests over the data [11].

Based on the same space of decision with M classes, a MNB method computes conditional class probabilities and then predicts the most probable class of a vector of training data X, according to sample data D. The parameters of MNB method are learning from data. The final decision about vector of training data X is done by equation (7), where $P(w_* \mid X_1, X_2, \ldots, X_n)$ *with* $* = \{i, j \mid i, j \in \Omega\}$, is done by (20) [13].

Fuzzy Naive Bayes

Several authors proposed versions of Fuzzy Naive Bayes method. Tang et al. [23] proposed a Fuzzy Naive Bayes method with two stages: in the first one, a fuzzy clustering algorithm determines partitions in space of decision. In the second stage, the partitions obtained in the first stage are used to estimate the parameters for linguistic variables. With that methodology, it is possible to use continuous variables and to decrease the learning complexity of Naive Bayes method. In another paper, Tang et al. [24] analyses the model identification using weighted fuzzy production rules. After that, the accuracy of fuzzy production rules is investigated using genetic algorithms [25].

Other approaches were used for fuzzy models of Naive Bayes method. Borgelt [1] extended the Naive Bayes method to manipulate some kinds of fuzzy information. Nurnberger [19] made mappings from Naive Bayes method to a NEFCLASS modified algorithm to improve the first one. Borgelt [2] use Fuzzy Maximum Likelihood Estimation [6] to determine fuzzy partitions.

There is a third approach, in which fuzzy discretization methods [30] were used in the first stage to allow use Naive Bayes method after. However, this approach can affect classification bias and variance of Naive Bayes method.

Störr [22] proposed another approach which has some advantages: is fast; is able to work with few training examples; is able to deal with missing attributes; can be used for incremental learning and if all fuzzy membership functions assume values in {0;1}, that approach has the same behavior than Naive Bayesian classification algorithm.

Based on the same space of decision with M classes, the Fuzzy Naive Bayes method computes conditional class probabilities and then predict the most probable class of a vector of training data $X=\{X_1, X_2, \ldots, X_n\}$, according to sample data D. In this case, it is assumed each X_k, $k=1,\ldots,n$, is a linguistic variable and it is expressed by linguistic values, with normalized membership functions $\mu_i(X_k)$, where $i=1,\ldots,M$. Those functions in Fuzzy Naive Bayes method [22] are interpreted as conditional information of class w_i done by a variable X_k. Then, let $X = \{ X_1 = A_{1i}, \ldots, X_n = A_{ni} \}$ and using Bayes:

$$P(w_i \setminus X) = [P(X \setminus w_i)\, P(w_i)\, \mu_i(X)]/P(X) \quad (21)$$

As before, the Naive Bayes method assumes conditionally independent among the events in X. It modifies the equation (21) to:

$$P(w_i \setminus X) = (1/S)\ P(w_i)\ \Pi^n{}_{k=1}\ \ [P(X_k \setminus w_i)\ \mu_i(X)] \quad (22)$$

Then, the classification rule for Fuzzy Naive Bayes is done by:

$$X \in w_i \text{ if: } P(w_i \setminus X) > P(w_j \setminus X),$$
$$\text{for all } i \neq j \text{ and } i, j \in \Omega \quad (23)$$

where $P(w_i \setminus X) = P(w_i \setminus X_1 = A_{1i}, \ldots, X_n = A_{ni})$, is done by (22) [14].

GENERAL FUZZY PROBABILISTIC NAIVE BAYES

As mentioned before, a general formulation for a naive Bayes style fuzzy probabilistic network was not proposed yet. In this paper, we proposed a formulation for this kind of supervised classifier, using random variables without specifying any statistical distributions. This approach can be useful for classification purposes, when random variables can have different statistical distributions.

A New Formulation

Zadeh proposed the probability measure for fuzzy events [32]. The probability space is a triplet (R^n, **B**, **P**), where **B** is the σ-field of Borel sets in R^n and **P** is a probaility measure over R^n. Let A in **B** be a fuzzy event, with a membership function μ_A: $R^n \rightarrow [0,1]$, then the probability of a fuzzy event A is defined by the Lebesgue-Sieltjes integral:

$$\mathbf{P}(A) = \int_A\ dP = \int_{R^n}\ \mu_A(x)\ \ dP = E(\mu_A) \quad (24)$$

Then, the probability of a fuzzy event is the expectation of its membership function [32]. The equation (24) can be written as:

$$\mathbf{P}(A) = \int_{R^n}\ \mu_A(x)\ \ P(x)\ \ dx \quad (25)$$

As the membership function for a crisp event is equal to the characteristic function, then the equation (25) can be changed for:

$$\mathbf{P}(A) = \int_{R^n}\ P(x)\ \ dx \quad (26)$$

Therefore, the equation (6) for Generalized Naive Bayes method can be written as:

$$\boldsymbol{P}(w_i \mid X_1, X_2, \ldots, X_n) = (1/S)\ \boldsymbol{P}(w_i)\ \Pi^n{}_{k=1}\ \ \boldsymbol{P}(X_k \setminus w_i) \quad (27)$$

This expression is a generalization of naive Bayes method, where P(•) can be done by equation (25) or (26). Then, using (27), the new classification rule from equation (7) for Generalized Naive Bayes is given by:

$$X \in w_i \ \ \text{if} \ \ \boldsymbol{P}(w_i \mid X_1, X_2, \ldots, X_n) > \boldsymbol{P}(w_j \mid X_1, X_2, \ldots, X_n)$$
$$\text{for all } i \neq j \text{ and } i, j \in \Omega \quad (28)$$

According to the statistical distribution which X assumes, all equations above for Gaussian Naive Bayes Method (eq. 9 and 10), Naive Bayes Network Modeled by Gaussian Mixture Models (eq. 19), Modified Naive Bayes (eq. 20), Fuzzy Naive Bayes (eq. 22 and 23) can be easily rewritten. Other versions of Naive Bayes, which can be found in the literature, can be also generalized using the same idea.

APPLICATIONS

Data classification is a technique which can be used in several areas of human knowledge. In particular, in health sciences use classification for many purposes. medical diagnostic decision support, classification of injury narratives, skills assessment in medical training, medical diagnosis and others.

In Epidemiology, results obtained from Naive Bayes classifiers or its derived forms can help decision makers to redefine or create public policies, reassign people to reinforce the staff working in priority areas and create specific profiles for diseases and/or patients.

Acknowledgment

The authors would like to thank to Brazilian Council for Scientific and Technological Development (CNPq), grant 310470/2012-9, which partially supported this work. This work is related to the National Institute of Science and Technology for Medicine Assisted by Scientific Computing (process 181813/2010-6) also supported by CNPq.

REFERENCES

[1] Borgelt, C.; Gebhardt, J.; A Naive Bayes Style Possibilistic Classifier. Proceedings of *7th European Congress on Intelligent Techniques and Soft Computing* (EUFIT'99), 1999. [cdrom].

[2] Borgelt, C.; Timm, H.; Kruse, R.; Using fuzzy clustering to improve naive Bayes classifiers and probabilistic networks. Proceedings of *Ninth IEEE International Conference on Fuzzy Systems* (FUZZ IEEE 2000), p. 53-58, 2000.

[3] Borgelt, C., Kruse, R., Graphical Models: Methods for Data Analysis and Mining. Wiley, 2002.

[4] Dempster, A.P., Laird, N.M. and Rubin, D.B. Maximum likelihood from incomplete data via EM algorithm. Journal of Royal Statistical Society, Ser. B. 39, pp.1-38, 1977.

[5] Doring, C. Borgelt, C. and Kruse, R., Fuzzy clustering of quantitative and qualitative data, Proc. of the 2004 NAFIPS, pp. 84-89, 2004.

[6] Gath, I.; Geva, A.B.; Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.11, n. 7, p. 773-780, 1989.

[7] Kononenko, I., Inductive and Bayesian learning in medical diagnosis. Applied Artificial Intelligence 7(4), 317–337, 1993.

[8] Machado, L.S., Moraes, R.M. Intelligent Decision Making in Training Based on VR. In:Da Ruan(Org.) Computational Intelligence in Complex Decision Systems. Cap 4: p.85-123. Atlantis Press, 2010.

[9] McLachlan, G. and Peel, D. Finite Mixture Models. New York: Wiley-Interscience, 2000.

[10] Mitchell, T. Machine Learning, McGraw-Hill, 1997.

[11] Mood, A. M., Graybill, F., and Boes, D. C.; Introduction to the Theory of Statistics. McGraw-Hill, 3rd ed., 1974.

[12] Moraes, R. M.; Machado, L. S., Assessment Based on Naive Bayes for Training Based on Virtual Reality. Proceedings of International Conference on Engineering and Computer Education (ICECE'2007). Março, Santos, Brasil, pp 269-273, 2007.

[13] Moraes, R. M.; Machado, L. S., A Modified Naive Bayes to Online Training Assessment in Virtual Reality Simulators. 3th International Conference on Intelligent System and Knowledge Engineering (ISKE'2008). Xiamen, China, 2008.

[14] Moraes, R. M.; Machado, L. S., Another Approach for Fuzzy Naive Bayes Applied on Online Training Assessment in Virtual Reality Simulators. Proceedings of Safety Health and Environmental World Congress 2009 (SHEWC'2009). 26-29 Julho, Mongaguá, Brasil, p. 62-66, 2009.

[15] Moraes, R. M.; Machado, L. S., Gaussian Naive Bayes for Online Training Assessment in Virtual Reality-Based Simulators. Mathware & Soft Computing, v. 16, n.2, p. 123-132, 2009a.

[16] Moraes, R.M.; Machado, L.S., An Online Training Assessment Based on Naive Bayes Network Modeled by Gaussian Mixture Models for Medical Simulators. VII International Conference on Engineering and Computer Education (ICECE'2011). Guimarães, Portugal. 5p., 2011.

[17] Moraes, R. M.; Machado, L. S., Online Assessment in Medical Simulators Based on Virtual Reality Using Fuzzy Gaussian Naive Bayes. Journal of Multiple-Valued Logic and Soft Computing, v.18, n.5-6, pp.479-492, 2012.

[18] Moraes, R. M.; Machado, L. S., Simultaneous Assessment of Teams in Collaborative Virtual Environments Using Fuzzy Naive Bayes. 2013 IFSA World Congress and NAFIPS Annual Meeting (IFSA'2013) . 24-28 Junho, Alberta, Canadá, p. 1343 – 1348, 2013.

[19] Nurnberger, A.; Borgelt, C.; Klose, A.; Improving naive Bayes classifiers using neuro-fuzzy learning. Proceedings of *6th International Conference on Neural Information Processing* (ICONIP '99), p. 154-159 , 1999.

[20] Rabiner, L.R and Juang, B-H. Fundamentals of Speech Recognition. Prentice Hall PTR, New Jersey, 1993.

[21] Ross, S., A First Course in Probability. Pearson, 8th ed., 2009.

[22] Störr, H.-P. A compact fuzzy extension of the naive Bayesian classification algorithm, Proceedings of the Third International Conference on Intelligent Technologies and Vietnam-Japan Symposium on Fuzzy Systems and Applications, Hanoi, Vietnam, 172-177, 2002.

[23] Tang, Y.; Pan, W.; Li, H.; Xu, Y.; Fuzzy Naive Bayes classifier based on fuzzy clustering. Proceedings of *2002 IEEE International Conference on System, Man and Cybernetics*. October, 2002.

[24] Tang, Y.; Pan, W.; Qiu, X.; Xu, Y.; The identification of fuzzy weighted classification system incorporated with Fuzzy Naive Bayes from data. Proceedings of *2002 IEEE International Conference on System, Man and Cybernetics*. October, 2002.

[25] Tang, Y.; Xu, Y.; Application of fuzzy Naive Bayes and a real-valued genetic algorithm in identification of fuzzy model. Information Sciences, v.169, p.205-225, 2005.

[26] Tran, D.; Pham, T. and Wagner, M. Speaker recognition using Gaussian mixture models and relaxation labeling. Proc. 3rd World Multiconf. on Systemetics, Cybernetics and Informatics/ 5th Int. Conf. Information Systems Analysis and Synthesis (SCI/ISAS99), 6, pp. 383-389, 1999.

[27] Wagholikar K. B., Sundararajan V., Deshpande A. W., Fuzzy Naive Bayesian model for Medical Diagnostic Decision Support. Proceedings of 31st Annual International Conference of the IEEE EMBS. September, Minnesota, USA, pp. 3409-3412, 2009.

[28] Wagholikar K. B., Sundararajan V., Deshpande A. W., Modeling Paradigms for Medical Diagnostic Decision Support: A Survey and Future Directions. Journal of Medical Systems, v.36, n.5, pp. 3029-3049, 2012.

[29] Wellman, H.M., Lehto,M., Corns, H., A combined Fuzzy and Naive Bayesian strategy can be used to assign event codes to injury narratives. Injury Prevention, v.17, pp. 407-414, 2011.

[30] Yang, Y.; Webb, G. I.; A Comparative Study of Discretization Methods for Naive-Bayes Classifiers. Proceedings of *2002 Pacific Rim Knowledge Acquisition Workshop* (PKAW'02), p. 159-173, 2002.

[31] Yang, Y. and Webb, G.I., On Why Discretization Works for Naive-Bayes Classifiers, Lecture Notes on Artificial Intelligence, v. 2903, pp. 440-452, 2003.

[32] Zadeh, L. A., Probability Measures of Fuzzy Events. Journal of Mathematical Analysis and Applications, v.23, n.2, 1968.