Determination of Distant Learner's Sociological Profile Based on Fuzzy Logic and Naïve Bayes Techniques

Youness Chaabi(✉), Lekdioui Khadija
Ibn Tofaïl University, Kenitra, Morocco
`chaabiyouness@gmail.com`

Fatine Jebbor
Mohammed V University in Rabat, Morocco

Rochdi Messoussi
Ibn Tofaïl University, Kenitra, Morocco

Abstract—The present article is elaborated in the context of e-learning software systems that provide assistance and functionalities to learners engaged in distance learning. Our contribution consists of a system that estimates a behavioral (sociological) profile for each student. This estimation is based on automatic analysis of students' textual asynchronous conversations. In general, the automatic analysis of textual conversations is based on speech acts for classification and categorization of messages. This technique has several disadvantages like the absence of standardization of speech acts for determining social behaviors of learners. To overcome this, we propose a multi-agent system based on fuzzy logic reasoning and supervised learning technique for automatic classification and categorization of textual conversation. The determined profiles are proposed to teachers to assist them during tutoring tasks. The objective of this article is to share our reflections around these issues by presenting our experience in the analysis of asynchronous online discussion forums. In this paper, we specifically propose (i) definitions of the used sociological profiles that characterize the roles that learners play in the group and (ii) introduce the architecture of the Multi-Agent System (MAS) that determines the profiles. The system was experimented with the students of the Master Program "Software Quality" in the Ibn Tofail University. The results obtained from this experience, presented and discussed in this paper, show that the proposed approach can be of interest.

Keywords—Collaborative learning, Multi-Agent System, Fuzzy logic, Naïve Bayes, Speech act, Social Profiles.

1   Introduction

The work presented in this paper is part of the area of Human Learning technologies and more specifically: Computer-Supported Collaborative Learning (CSCL) software environments. These environments are designed depending on: models of domain knowledge, objects of learning and models intended to diagnose knowledge and the difficulties of learning and teaching styles strategies for adapting education to the profile of each learner. Even if distance learning is so interesting and technically feasible today, we would emphasize already identified problems as: the learner sociological insulation, the loss of motivation, the empowerment of the Learner, the assessment of the educational progress of the group [1].

Numerous researches in the field of environments of human learning are strongly interested in tracking user activities while interacting with communication tools provided by each learning platform. These tools of asynchronous communications such as e-mails and forums [2], allow learners to communicate between themselves and to organize their activities in order to achieve online learning. As part of the automatic analysis of collaborative learner's activities, we propose an approach for the analysis of textual asynchronous conversations of learners to determine social behavior [3, 4], the roles that are attributed and the organization they establish to carry out collaborative works. These observations provide the teacher with indicators that allow him to situate the work of learners compared to other learners, or groups to other groups. The aim of this approach is to support learners during the learning process. More specifically, the idea is to identify students who are isolated for a period of the projects in order to avoid abandons [1]. To reach this goal, we have chosen to analyze the contents of online discussion forums that could provide information on the learners' social behaviors.

In this paper, we propose an approach for the automatic analysis of asynchronous text conversations, based on a supervised learning algorithm inspired by the naïve Bayes method for the classification and categorization of messages after preprocessing. It is one of the most basic text classification techniques with various applications in document categorization, email spam detection, personal email sorting and language detection [5, 6, 7]. Naive Bayes classifiers perform well in many complex problems. Even though they are often outperformed by other techniques such as Decision Tree, Support Vector Machines, etc., naive Bayes classifiers are very efficient since they are less computationally intensive (in both CPU and memory) [8] and require a small amount of training data, which is particularly interesting for the problem addressed in this paper. Indeed, CSCL forums are frequently small in terms of quantity of messages. Moreover, the training time with Naive Bayes is significantly smaller as opposed to alternative methods. Our article presents a comparison of two classification methods that are Speech Acts and Naïve Bayes in order to determine the learner's social behavior. This paper is organized as follows: Section 2 is the background which presents the related works, the architecture of the proposed system, the analysis agents and their interactions. Section 3 describes the proposed robust classification approach. In section 4, we present an experiment to assess the relevance of the proposed system. Finally, Section 5 presents our conclusions.

2 State of the art

2.1 Related works

The analysis of conversations in discussion forums often relies on approaches which consist in applying category analysis defined for the study of face-to-face exchanges mediated by computer [9, 10, 11, 12, 13, 14]. This kind of approaches [9, 10, 11, 12, 13, 14] raises a number of methodological problems. For example, many existing systems assume that the messages are semi-structured which allow learners to follow a template and predefined syntax. Although, the employment of the systems can sometimes hamper the communication for the realization of collaborative work [15]. For example, "the message content is different of the discussions topics" [16] and this generates problems issues for the analysis of discussions.

Several research have focused on conversational analysis to determine the social behaviors of learners and, for this analysis, have used the theory of speech acts [17, 4]. A speech act is an utterance that serves a function in communication. It might contain just one word, as in "Propose!": "I propose an approach for text classification", as in "Sorry!" to perform an apology, or several words or sentences: "I'm sorry I forgot your request." Speech acts include real-life interactions and require not only knowledge of the language but also an appropriate use of that language within a given situation. The underlying idea is to identify speech acts in the analyzed conversations. There are two methods for identifying speech acts. The first is to analyze the content of interventions to try to determine these acts. This is the field of automatic natural language processing (NLP). In his book, Luzzati [18] notes the difficulties of the automatic analysis notably linked to the understanding of the underlying meaning by the system. Despite the progress made in this area, the results are not always reliable. The most efficient systems are those that analyze conversations made in a very specific and limited context. The second solution to identify speech acts in an intervention is to ask the user. It is a basic principle that is used particularly in CSCL: if the system can't recognize the necessary information for his analysis, then it must ask the user: "avoid guessing, get the student to tell you what you need to know" [19]. These techniques have some shortcomings:

— The results are not always reliable;
— It may hinder students when writing a message;
— It does not permit learners to focus more on the task.

In our case, there is another disadvantage to the use of the speech acts, that is the absence of a standardization of speech acts for determining the social behaviors of learners. It makes the task of analysis and classification more difficult to get good results. In literature, the current analysis tools are essentially based on quantitative analysis of interactions [16]. These tools do not address the problems of lexical and syntactical analysis of conversations and correction of spelling errors in textual conversations. Besides, reactions to the interventions are not considered to refine the calculations of interactions. The analytical difficulties presented above raised by fo-

rums are not solved, especially if one wants to analyze automatically large volumes of contributions. This is the case of non-semi-structured forums.

2.2 Behavioral profiles of students in groups

Robert Pléty has studied the behavior of students working in groups in a class. The study focus was on interactions between students working in groups of four for solving Algebra problems [20]. From this experience [20], Pléty defined behavior patterns that characterize the roles that learners play in the group during a period of time.

In order to determine these behavioral profiles, four kinds of observations are made for each student: the volume (number) of interventions, the different types of interventions, the communication gesture types (look and movement) and the reactions of other participants (consequences of behaviors). As we are using text messages exchanged on forums, speech acts are of interest for defining the intervention types. For our study, among the intervention indicators identified by Pléty (column "type of intervention" in the Table 1), we cite interrogation, proposal, reaction, response and evaluation, etc. These types are defined by the teacher's expertise.

These behavioral profiles generalize behavior patterns and are called profiles in the rest of the paper. Pléty identified four different profiles: Animator, Checker, Seeker and Independent. The characteristics of these four profiles are summarized in Table 1.

Table 1. Behavioral profiles of students working in groups [20].

Name	Volume of Intervention	Type of Interventions	Entrained Reaction
Animator	Important	Question or proposal	Followed by positive reactions
Checker	Enough important	Réaction, reponse and evaluation	No monitoring reactions
Seeker	Little important	(Very oubtful (question))	Questions are well accepted
Independent	Low	Little or no proposal or evaluation.	Interventions remain unresolved

Building on the work of Pléty [3], we have been able to find the same profiles among learners working in a group through a forum. Animator, Checker, Independent and Seeker are almost always present in all groups, showing that the characteristics of the classical groups are also found for those working remotely.

We believe that it is useful to have a computer system capable of automatically analyzing these behaviors. Particularly, in particular, in a context of online training, it is interesting to try to determine automatically patterns of social behavior within groups of learners, because they can be useful both for the tutor and for the learners themselves.

2.3 System

It is possible to identify profiles by manually analyzing the interactions and conversations, stored within forums for example, in online collaborative learning. Our goal is to automate the determination of these profiles by designing a system for anal-

ysis of asynchronous textual conversations. In this section, we present the existing work on which we rely to achieve our goal.

The existing system, described in [4, 21], is based on multi-agent systems (Figure.2). The general principle consists in the analysis of asynchronous messages exchanges between learners in LMS (Learning Management System). Each of these messages undergoes a sequence of treatments which extracts a profile. The determined profiles are inspired by the work of Pléty [20]. Figure 1 presents Below is the organization specifying the roles and interactions of the system. The notation used is the one defined in the ASPECS methodology [22, 23, 24].

The ASPECS methodology was chosen because of it completeness and the ease of deployment with its associated development platform: Janus. In Figure 2, we describe how a Behavior Analysis community is organized in Janus.

Preprocessing of messages. In order to determine each profile, each role described in figure (1) performs a specific treatment. The extraction of messages is followed by a step of spelling grammar correction. Indeed, spelling mistakes in texts frequently cause analysis problems. Hence the need for a corrector that first uses a word in the dictionary (corpus) associated with an algorithm that takes into account the variations of the language (verb conjugation, agreement of nouns and adjectives). Then, the corrector compares the words in the text to the corpus and suggestions by considering the context of the sentence. After message extraction, the pre-filtering treatment automatically deletes the words that do not contain information. This is done by using a "stop-words" list specific to each language. According to Zipf law [25], their removal during message preprocessing allows saving time during the modeling and the analysis of the message. The fourth treatment is based on a classification of messages by a supervised learning algorithm. Finally, indicators are computed and the last role uses the computed results and fuzzy logic technique to determine for each learner the evolution of its behavioral profiles.
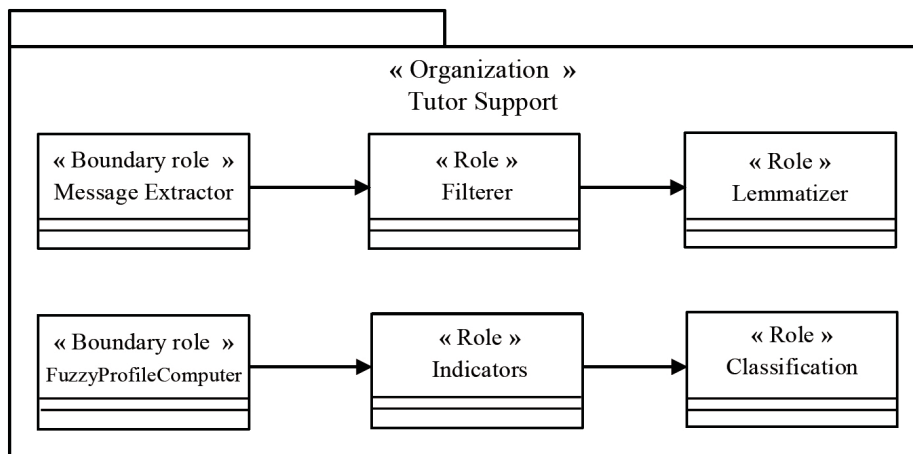

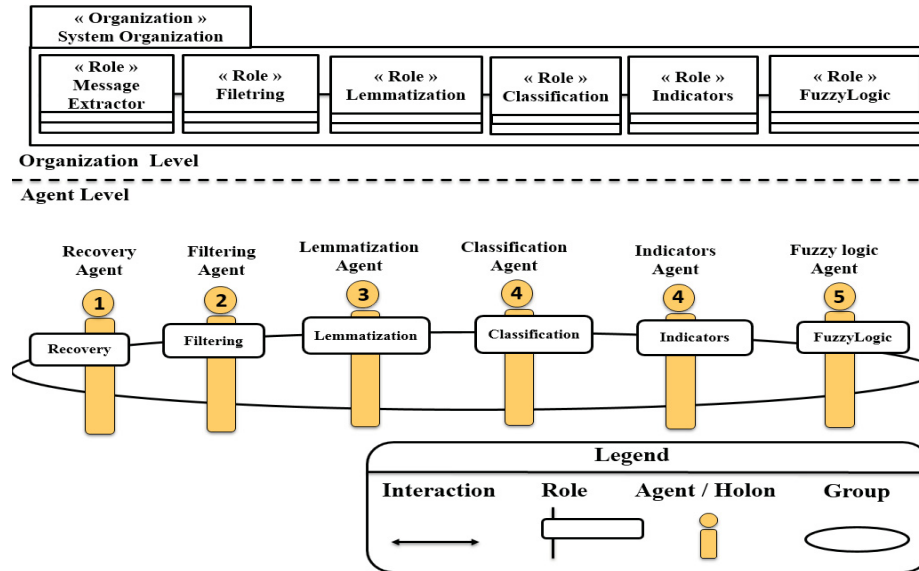**Fig. 1.** The ASPECS methodology of our system [21]

Fig. 2. The organization and group of a Behavior Analysis community in Janus [21]

Indicators Agent. Based on the above specifications for the design of our system, we are able to design an automatic analysis system to identify four behavior profiles identified by Pléty: Animator, Checker, Seeker and Independent. Our objective here is not to obtain a complete model of the learner as is the case of intelligent tutors to model knowledge, but we want to have an image of social behavior during conversations. We present the formulas used by indicator agent to analyze the discussions in collaborative works. These heuristics formulas were determined from the work of Pléty and were refined in experiments.

Volume of interventions: The following formula (Equation 1) calculates the ratio of participation of a learner by dividing nbMsgLearner(p) which is the number of messages sent by learner P, by NbrTotalMessagesGroup(x) that is the number of messages sent by students of the same group.

This ratio refers to the volume of intervention "VI" for a learner (p) belonging to a group x:

$$VI = \frac{nbMsgLearner(P)}{NbrTotalMessagesGroup(x)} * 100 \tag{1}$$

Type of interventions: Four expressions are used to calculate the speech acts' ratios in messages. The presence of speech acts in a message sent by a learner P is calculated as follows (Equation 2):

$$SpeechActAnimator = \frac{SpeechActAnimator(P)}{SpeechAct(A,C,S,I)} * 100 \tag{2}$$

In this formula, SpeechActAnimator(p) is the number of speech acts of category animator (for Example propose, encourage,…etc) sent by learner (p). SpeechAct(A,C,S,I) is the total number of speech acts (respectively Animator,

Checker , Independent and Seeker) sent by the learner. The Calculations of other ratios types (checker, seeker and independent) are obtained similarly.

Entrained Reaction: According to the characteristics of the defined profiles (Table 1), the volume of reactions triggered by a message allows to characterize a behavioral profile. For example, an animator profile requires a very large monitoring of reactions compared to that of a checker. We calculate, for each message, direct reactions (Equation 3) (first reaction to a message) and indirect reactions (Equation 4) (number of interventions after the creation of the message). According to the tree structure defined for messages, the nodes represent the identifiers of messages sent by the learners and the size of this tree is equivalent to the number of interventions made after the creation of the topic. Two expressions, using the n-ary tree structure of the messages, are used to calculate subsequent reactions of each message:

The direct reaction is the number of direct responses to the messages of the learner divided by the total number of direct answers on posted messages by learners in the group.

$$Reaction_{Direct} = \frac{\sum_{i=1}^{m} ReplyToMessage_i(Learner)}{TotalOfReponse(Group)} * 100 \qquad (3)$$

Indirect reaction is the depth of discussion minus the number of direct reaction divided by the sum of the depths of the subjects send by learners.

$$Reaction_{Indirect} = \frac{Depth-Reponses}{TotalDepth} * 100 \qquad (4)$$

Fuzzy Logic Agent. Once all previous treatments are realized, the MAS needs to determine a profile. This kind of reasoning is done through fuzzy logic techniques as it is adapted to deal with incomplete and imperfect information [26, 27]. The fuzzy model represents the knowledge of experts-teachers under a linguistic form and includes characteristics of the learner. That is to say, a decision is made by a combination of a set of fuzzy facts, each contributing at some point, to some degree to a fuzzy relation and to the final decision. Based on the defined learning profiles characteristics (see Table 1), an analysis is performed to allow the tutor to associate a profile for each learner [21]. After applying the fuzzification methods, inference engine and defuzzification method [21], the resulting values give an estimation, through a percentage, of the behavioral level of each learning profile (Animator, Checker, Seeker and Independent).

3 Classification Approach

In the literature, many solutions for text classification are proposed [28, 29, 30, 31]. In this work, we propose to use machine learning for Text Classification and specifically a naïve Bayes based technique. Based on a comparison of different techniques for automatic classification [32, 33, 34, 35], we hope that the naïve Bayes classifier will be adapted to the problem addressed in this paper.

In order to understand the classification techniques that we will present, the concept of "message" must be examined. According to [36], statistical approaches usually use a bag of words model for texts. «The bag-of-words model is a simplifying representation used in natural language processing and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity [37]». The associated text representation is then a vector, which can be:

Binary: One retains only the presence or the absence of a word, regardless of the number of occurrences. This is the oldest approach and the simpler;

Frequency: This is an extension of the binary model, where the terms occurrences are counted. Its normalized version can take into account different sizes of text. Each component of the vector is weighted by the size of the document;

TF-IDF: This acronym, for Term Frequency–Inverse Document Frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus [38]. It is often used as a weighting factor in information retrieval and text mining. The TF-IDF value increases proportionally to the number of times a word appears in the document, but it is balanced by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general [39, 31].

The representations of the documents are considered sequentially here. They are mainly used in a Bayesian network. These learning based approaches can be implemented on a large corpus, and give good results in tasks such as opinion analysis and text classification.

3.1 Naïve Bayes text classification

A naive Bayes classifier is a probabilistic classifier that estimates conditional probabilities of the dependent variable from training data and uses them for classification of new data instances. The algorithm is very fast for discrete features but runs slower for continuous features. We can summarize its use when applied to text classification as follows:

— Classification that maximizes the probability of observing the words that were actually found in the example messages.
— During the training phase, the classifier calculates the probability that a new message belongs to such category from the proportion of training messages belonging to this category. It also calculates the probability that a word is present in a text, knowing that this text belongs to this category.
— Thereafter, when a new message is to be classified, the probability that it belongs to each category using the Bayes rule.

The estimation of the probability is: P $(c_j \mid a_1, a_2, a_3, ..., a_n)$ where:

· c_j is a category
· a_j is an attribute

The probability that a word appears in a text is independent of the presence of other words in the text. We know that this is generally false. For example, the probability of occurrence of the word "artificial" may depend partly on the presence of the word "intelligence." However, this assumption does not prevent such a classifier to present satisfactory results. Besides, the most important point is that it greatly reduces the necessary calculations. Without it, the classifier would take into account all possible combinations of words in a text, which on the one hand involves a large number of calculations, but also reduce the statistical quality of the estimation. This is because the occurrence frequency of each combination would be lower than the occurrence frequency of words separately.

3.2 Description of Bayesian model

Text classifiers often don't use any kind of deep representation of language: a message is often represented as a bag of words which is a set that allows elements repetition. This is an extremely simple representation: the classifier only knows which words are included in the message (and how many times each word occurs), and throws away the word order!

Consider a message M, whose class is given by C. In our case, there are four classes C = A (Animator), C = V (Checker), C = Q (Seeker) and C = I (Independent). We classify M as the class which has the highest posterior probability P (C | M) that can be re-expressed using the Bayes Theorem (Equation 5):

$$P(C|M) = \frac{P(M|C)P(C)}{P(M)} \tag{5}$$

We must look at four probabilistic models of messages, each of them represent messages as a bag of words, using the Naïve Bayes assumption.

A model represents messages using feature vectors whose components correspond to words types. If we have a vocabulary V, containing |V| word types, then the feature vector dimension is d = |V|.

3.3 The Bernoulli message model

In the In the Bernoulli model [40], a message is represented by a features vector with binary elements taking value 1 if the corresponding word is present in the message and 0 if the word is not present. This message is represented by a binary vector, which represents a point in the space of words. If we have a vocabulary V containing a set of |V| words, then the tth dimension of a message vector corresponds to word w_t in the vocabulary. Le b_i be the feature vector for the ith message M^i, the tth element of b_i, written b_{it}, is either 0 or 1 representing the absence or presence of word W_t in the ith message. Let $P(W_t|C)$ be the probability of word w_t occurring in a message of class C. The probability of W_t not occurring in a message of this class is given by $1 - P(W_t|C)$. If we make the Naive Bayes assumption, that the probability of each word occurring in the message is independent of the occurrences of the other words,

then we can write the message likelihood $P(M^i|C)$ in terms of the individual word likelihoods $P(W_t|C)$ (Equation 6) :

$$P(M^i|C) \approx P(b_i|C) = \prod_{t=1}^{|V|}[b_{it}P(w_t|C) + (1 - b_{it})(1 - P(w_t|C))] \quad (6)$$

This product goes over all words in the vocabulary. If word w_t is present, then $b_{it} = 1$ and the required probability is $P(W_t|C)$. If the word W_t is not present, then $b_{it} = 0$ and the required probability is $1 - P(W_t|C)$.

We can imagine this as a model for generating message feature vectors of class C. In this model, each message feature vector is modeled as a collection of dimension d=|V|, the tth element having a probability of success equal to $P(W_t|C)$. The likelihood parameters are the probabilities of each word given the message class $P(W_t|C)$ (Equation 7). The model is also parameterized by the prior probabilities P(C). We can learn (estimate) these parameters from a training set of messages labeled with class C = k. Let $n_k(w_t)$ be the number of messages of the class C = k in which w_t is observed, and let N_k be the total number of messages of this class. Then we can estimate the parameters of the word likelihoods as, the relative frequency of messages of class C = k that contain word W_t

$$P(w_t|C = k) = \frac{n_k(w_t)}{N_k} \quad (7)$$

If there are N messages in the training set, then the prior probability of the class C = k may be estimated as the relative frequency of messages of this class C = k (Equation 8):

$$P(C = k) = \frac{N_k}{N} \quad (8)$$

Thus given a training set of messages (each labeled with a class) and a set of K classes, we can estimate a Bernoulli text classification model as follows:

1. Define the vocabulary V. The number of words in the vocabulary defines the dimension of the feature vectors.
2. Count the following in the training set:
   (a) N the total number of messages;
   (b) N_k the number of messages labeled with class C = k, for k = 1,…, K;
   (c) $n_k(w_t)$ the number of messages of class C = k containing word w_t for every class and for each word in the vocabulary.
3. Estimate the likelihoods $P(W_t|C = k)$. using equation (7).
4. Estimate the priors P (C = k) using equation (8).

To classify an unlabeled message M^J, we estimate the posterior probability for each class (equation 9) combining equations (5) and (6):

$$\begin{aligned}
P(C|M^J) &= P(C|b_j) \\
&\approx P(b_j|C)P(C) \\
&\approx P(C) \prod_{t=1}^{|V|}[b_{jt}P(w_t|C) + (1 - b_{jt})(1 - P(w_t|C))]
\end{aligned} \quad (9)$$

3.4 Problem of matching categories

The computed probabilities by the Naïve Bayes classifier are generally not sufficient to determine messages categories. In order to obtain a more reliable classification, the probabilities and the results provided by the indicator agent are used by the fuzzy logic agent. From Table 1 (columns: type of intervention and entrained reaction), the animator messages are followed by important reaction. For example, if the probabilities of a message in categories (Animator, Checker, Seeker and Independent) are close and if the direct and indirect reaction are important, then the message is classed in the animator category. This problem is solved by the technique of fuzzy logic.

Now that we have seen how the system calculates the probability for each class, their analysis in light of the characteristics of defined messages (see table 1, columns type of intervention and entrained reaction) is performed to allow the tutor to associate a message to a class. In our case, we developed a fuzzy system with 6 input variables and 4 output variables. The input variables correspond to the indicators calculated in section 3 and 2.2.2 (for each message we calculated the direct and indirect reaction). These input variables and their subdivisions in terms of fuzzy sets are given as follows:

— Probability class animator (in %), the Fuzzy sets is Low, Medium, High
— Probability class checker (in %), the Fuzzy sets is Low, Medium, High
— Probability class seeker (in %), the Fuzzy sets is Low, Medium, High
— Probability class independent (in %), the Fuzzy sets is Low, Medium, High
— Direct reactions (in %), the Fuzzy sets is Low, Medium, High
— Indirect reactions (in %), the Fuzzy sets is Low, Medium, High

The output variables, representing the behavioral level for the different profiles (Animator, Checker, Seeker and Independent), and their fuzzy set subdivisions are given as follows:

— Category animator (in %), the Fuzzy sets is Insufficient, Medium, Good, Excellent
— Category checker (in %), the Fuzzy sets is Insufficient, Medium, Good, Excellent
— Category seeker (in %), the Fuzzy sets is Insufficient, Medium, Good, Excellent
— Category independent (in %), the Fuzzy sets is Insufficient, Medium, Good, Excellent

Now that we have defined the fuzzy variables, we are able to run the inference engine. By using data corresponding to the various inputs and outputs, the tutor-experts supply series of combinations based on the conditions of Pléty [3] which characterize every message sent by the learners (Animator, Checker, Seeker and Independent). Each rule is written in the form: "if condition then action/conclusion".

```
IF  (Probability class animator IS High) AND
    (Direct Reaction IS High) AND
    (Indirect Reaction IS High)
THEN
    (Category animator IS Excellent)
```

Once the base of rules is established and given values for all the input variables, we infer the rules in order to obtain a modified membership function for each output variable. For that, we have used the MAX-MIN inference method [41, 42]. The last stage of the fuzzy process is defuzzification. There are several existing methods to perform defuzzification. In our case, we opted for the "gravity center" method consisting in retaining the value corresponding to the gravity center of the surface defined by the modified membership function. After applying the defuzzification process to all the output variables, the fuzzy logic reasoning module provides a confidence degree of each class (Animator, Checker, Seeker and Independent). The estimated class is the one with the highest confidence. Four expressions are used to calculate the type of intervention corresponding to the four classes: Animator, Checker, Seeker and Independent. For example, the type of animation intervention sent by a learner P is calculated as follows (Equation 10):

$$AnimatorIntervention = \frac{NbMessagesAnimator(P)}{TotalMessagesLearner(A,C,S,I)} * 100 \qquad (10)$$

In this formula, NbMessagesAnimator(P) is the number of messages of category animator (for Example: propose, encourage, etc.) sent by learner (p). TotalMessagesLearner(A,C,S,I) is the total number of messages, respectively Animator, Checker, Seeker and Independent, sent by learner. The Calculations of other ratios types (Checker, Seeker and Independent) are obtained similarly.

4 Experimentation

The purpose of our work is to support tutors during the monitoring of learning activities involved in Virtual Learning Environments (VLE). In order to evaluate the pertinence of the proposed system, this section proposes a comparative study between a "classical" human-only evaluation, and the results produced by the MAS.

4.1 Evaluation of human interventions

We asked students of the Software Quality Master Program to set up a work as part of mobile development course. The goal is to develop and implement mobile applications. This work should be done exclusively online, using the forum that we created on the Moodle platform. This work lasted four-month and was organized into six phases, each corresponding to different tasks. For each phase, the number of contributions exchanged (messages) by the learners in each group was greater than one hundred.

Students who participated in this study were first year Master Program students at the University Ibn Tofail in Kenitra, Morocco. Their number of these students was approximately 40 students. They had to work by a group of 4. Besides, they have a good knowledge of ICT and in particular of the use of the discussion forums.

The experimentation is made of three parts: a first part consists in human tutor's analysis of profiles. The results of this analysis act as a reference. A second part results from the use of a previous version of the system within which speech acts

were predetermined for a specific profile [4]. Eventually, the third part is the output of the system presented in this paper.

4.2   Results and Discussions

The first set of results from the analysis on forums of learners was monitored by three human tutors. Each tutor analysis was done in two steps. The first step consisted in assigning a profile to each student according to it behavior during classroom exercises. The second step consisted in analyzing the messages exchanged by students. The tutors were asked to classify the messages of learners into four profile categories (Animator, Checker, Seeker and Independent) by analyzing their content, i.e. identifying the speech acts that characterize messages.

The characterization of each profile is based on the Pléty's behavioral profiles analysis summarized in Table 1. Tables 2 and 3 illustrate these results (two steps analysis) for the first two phases of the project. The learner profile column is the result of the first step. The group of columns: intuitive message classification, volume intervention and entrained reaction are the result of the second step.

Table 2. Results of the intuitive analysis for group 1, phase 1 Requirement

	Profil	Intuitive result analysis for Group N°1 Phase Requirement						
		Intuitive message classification (Type intervention)				Entrained Reaction		Volume Intervention
		Number of Messages category "Animator"	Number of Messages category "Checker"	Number of Messages category "Seeker"	Number of Messages category "Independent"	Direct reaction	Indirect reaction	
Learner 1	Animator	15 (34,88 %)	21 (48,83 %)	07 (16,27 %)	00 (00,00 %)	36,53 %	38,70 %	Important (44,33 %)
Learner 2	Animator	11 (47,82 %)	11 (47,82 %)	01 (04, 34 %)	00 (00,00 %)	26,92 %	35,23 %	Important (23,71 %)
Learner 3	Independent	3 (50,00 %)	3 (50,00 %)	00 (00,00 %)	00 (00,00 %)	04,80 %	09,25 %	Low (06,19 %)
Learner 4	Animator	13 (52,00 %)	11 (52,00 %)	01 (04,00 %)	00 (00,00 %)	28,84 %	39,50 %	Important (25,77 %)

Table 3. Results of the intuitive analysis for group 1, phase 2 Analysis and Design

	Profil	Intuitive result analysis for Group N°1 Phase Analysis and Design						
		Intuitive message classification (Type intervention)				Entrained Reaction		Volume Intervention
		Number of Messages category "Animator"	Number of Messages category "Checker"	Number of Messages category "Seeker"	Number of Messages category "Independent"	Direct reaction	Indirect reaction	
Learner 1	Animator	22 (53,65 %)	12 (29,26 %)	05 (12,19 %)	02 (04,87 %)	42,59 %	23,11 %	Important (39,05%)
Learner 2	Checker	03 (15,78 %)	13 (68,42 %)	02 (10, 52 %)	01 (05,26 %)	20,37 %	15,16 %	Enough Important (18,10 %)
Learner 3	Independent	1 (16,66 %)	3 (50,00 %)	00 (00,00 %)	02 (33,33 %)	00,92 %	02,88 %	Low (05,71 %)
Learner 4	Animator	16 (48,48 %)	12 (36,36 %)	03 (09,09 %)	02 (06,06 %)	27,77 %	20,21 %	Important (31,43 %)

The same forums were used to automatically estimate the profile of each learner. So, two automatic methods were used. The first method is based on a previous version of the system presented in this paper. In this previous version, speech acts were predetermined for a specific profile [4, 21]. The second method is the one proposed in this paper.

As described before, the methods start by extracting the qualitative and quantitative indicators that characterize learner profiles. For illustration figure 3 shows all the indicators calculated by our system for each learner of the groups 1 to 4 during the first phase of the project (Phase 1: Requirement). These indicators are: the volume of interventions, the percentage of direct reactions, the percentage of indirect reactions and type of intervention (for Animator, Checker, Independent and Seeker) using both versions of the system. The first version is predetermined speech acts (see equation "E1" and see column Intervention type "Speech Acts" of figure 3). The second version is Naïve Bayes classification (see equation "E2" and see column Intervention type "Nave Bayes" of figure 3).

Result Analysis System													
Group	Learner	Volume Intervention	Entrained Reaction		Intervention Type « Naive Bayes »				Intervention Type « Speech Acts »			
			Direct	Indirect	Animator	Checker	Seeker	Independent	Animator	Checker	Seeker	Independent
Group 1	Learner 1	44,33 %	36,53 %	38,70 %	33,45 %	47,03 %	15,98 %	00,00 %	30,00 %	43,83 %	14,20 %	00,00 %
	Learner 2	23,71 %	26,92 %	35,23 %	46,82 %	46,12 %	03,33 %	00,00 %	45,11 %	46,67 %	02,89 %	00,00 %
	Learner 3	06,19 %	04,80 %	09,25 %	48,00 %	48,88 %	00,00 %	00,00 %	45,00 %	46,90 %	00,00 %	00,00 %
	Learner 4	25,77 %	28,84 %	39,50 %	49,9 %	43,00 %	03,70 %	00,00 %	49,12 %	43,11 %	03,11 %	00,00 %
Group 2	Learner 5	14,81 %	11,11 %	20,00 %	05,17 %	33,33 %	00,00 %	20,00 %	11,11 %	33,33 %	00,00 %	20,00 %
	Learner 6	29,63 %	14,81 %	39,75 %	00,00 %	00,00 %	00,00 %	39,75 %	14,81 %	25,00 %	50,00 %	39,75 %
	Learner 7	29,63 %	48,14 %	36,14 %	21,27 %	00,00 %	40,00 %	36,14 %	48,14 %	00,00 %	00,00 %	36,14 %
	Learner 8	11,11 %	07,40 %	26,50 %	36,17 %	18,18 %	09,09 %	26,50 %	07,40 %	00,00 %	00,00 %	26,50 %
Group 3	Learner 9	05,00 %	06,89 %	03,65 %	06,89 %	00,00 %	99,00 %	03,65 %	06,89 %	00,00 %	99,00 %	03,65 %
	Learner 10	38,75 %	55,17 %	53,20 %	55,17 %	10,00 %	35,00 %	53,29 %	55,17 %	10,00 %	35,00 %	53,29 %
	Learner 11	07,50 %	10,34 %	00,00 %	10,43 %	20,00 %	00,00 %	00,00 %	10,43 %	20,00 %	00,00 %	00,00 %
	Learner 12	18,18 %	05,17 %	13,95 %	05,17 %	33,33 %	00,00 %	13,95 %	05,17 %	33,33 %	00,00 %	13,95 %
Group 4	Learner 13	00,00 %	00,00 %	00,00 %	25,17 %	33,33 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %
	Learner 14	22,92 %	21,27 %	20,08 %	00,00 %	00,00 %	00,00 %	20,08 %	21,27 %	00,00 %	40,00 %	20,08 %
	Learner 15	37,50 %	36,17 %	47,13 %	05,17 %	33,33 %	00,00 %	47,13 %	36,17 %	18,18 %	09,09 %	47,13 %
	Learner 16	18,75 %	23,40 %	19,67 %	23,40 %	00,00 %	57,14 %	19,67 %	23,40 %	00,00 %	57,14 %	19,67 %

Fig. 3. Indicators calculated by our system for each learner of the groups 1 to 4 during the first phase of the project.

By handling the previous indicators and using the fuzzy logic reasoning module, the method based on Naive Bayes classifier can estimate profiles for each learner. The Tables 4 and 5 present respectively the estimated profiles of the learners of group 1, during the first two phases of the project (Phase 1: Requirement and Phase 2: Analysis and Design). As previously presented, the fuzzy logic reasoning module provides a confidence degree for each profile. The resulting estimated profile is the one with the highest confidence degree. As one can see, the profiles estimated by our method are identical to the ones provided by the intuitive analysis (see tables 2 and 3).

In order to compare the two versions of the system, we provided the same set of data (the one used to produce tables 4 and 5) to the previous version of the system based on predetermined speech acts. The results are given in tables 6 and 7.

Table 4. Results from our system based on Naïve Bayes classifier for group 1 (phase 1 Requirement).

	Volume Intervention	Type intervention				Entrained Reaction		Profile
		Animator	*Checker*	*Seeker*	*Independent*	*Direct*	*Indirect*	
Learner 1	44,33 %	33,45 %	47,03 %	15,98 %	00,00 %	36,53 %	38,70 %	Animator
Learner 2	23,71 %	46,82 %	46,12 %	03,33 %	00,00 %	26,92 %	35,23 %	Animator
Learner 3	06,19 %	48,00 %	48,88 %	00,00 %	00,00 %	04,80 %	09,25 %	Independent
Learner 4	25,77 %	49,9 %	43,00 %	03,70 %	00,00 %	28,84 %	39,50 %	Animator

Result Analysis System for Group N°1 Phase Requirement

Table 5. Results from our system based on Naïve Bayes classifier for group 1 (phase 2: Analysis and Design)

	Volume Intervention	Type intervention				Entrained Reaction		Profile
		Animator	*Checker*	*Seeker*	*Independent*	*Direct*	*Indirect*	
Learner 1	39,05 %	52,30 %	28,27 %	12,03 %	04,50 %	42,59 %	23,11 %	Animator
Learner 2	18,10 %	14,80 %	67,80 %	10,00 %	05,12 %	20,37 %	15,16 %	Checker
Learner 3	05,71 %	15,00 %	50,00 %	00,00 %	32,21 %	00,92 %	02,88 %	Independent
Learner 4	31,43 %	31,43 %	35,00 %	08,60 %	05,43 %	27,77 %	20,21 %	Animator

Result Analysis System for Group N°1 Phase Analysis and Design

Table 6. Results from our system based on speech acts for group 1 (phase 1: Requirement)

	Volume Intervention	Type intervention				Entrained Reaction		Profile
		Animator	*Checker*	*Seeker*	*Independent*	*Direct*	*Indirect*	
Learner 1	44,33 %	30,00 %	43,83 %	14,20 %	00,00 %	36,53 %	38,70 %	Animator
Learner 2	23,71 %	45,11 %	46,67%	02,89 %	00,00 %	26,92 %	35,23 %	Animator
Learner 3	06,19 %	45,00 %	46,90 %	00,00 %	00,00 %	04,80 %	09,25 %	Independent
Learner 4	25,77 %	49,12 %	43,11 %	03,11 %	00,00 %	28,84 %	39,50 %	Animator

Result Analysis System for Group N°1 Phase Requirement

Table 7. Results from our system based on speech acts for group 1 (phase 2: Analysis and Design)

	Volume Intervention	Type intervention				Entrained Reaction		Profile
		Animator	*Checker*	*Seeker*	*Independent*	*Direct*	*Indirect*	
Learner 1	39,05 %	40,00 %	27,83 %	11,20 %	04,00 %	42,59 %	23,11 %	Animator
Learner 2	18,10 %	14,11 %	62,67 %	09,89 %	05,00 %	20,37 %	15,16 %	Checker
Learner 3	05,71 %	15,55 %	49,09 %	00,00 %	30,44 %	00,92 %	02,88 %	Independent
Learner 4	31,43 %	47,58 %	36,11 %	08,60 %	05,00 %	27,77 %	20,21 %	Animator

Result Analysis System for Group N°1 Phase Analysis and Design

As one can see, the estimated profiles by our methods (see tables 4 and 5 and tables 6 and 7) are identical to the ones provided by the intuitive analysis (see tables 2 and 3). These results tend to show that both methods are robust and that the used

criterions are not sufficient to distinguish between the two automatic methods. All these results allowed us to verify that the estimations of the tutors are the same as those based on these two methods, with a margin of error that we will calculate in the following (see Tables 8 and 9). We remind that the both indicators, the volume of intervention and the entrained reaction are quantitative. The results obtained with both indicators allowed us to verify that the estimations of the tutors are the same as those based on these two methods. The idea is to compare the types of intervention which are qualitative. The results are presented in brief in tables 8 and 9. They illustrate the results of the error margin calculation between the intuitive analysis (by a tutor) and the analysis performed by our system using both classifications.

The error margin is computed as follows (Equation 11):

$$Errormargin = \frac{\text{Results using predetermined Speech Acts}}{\text{Results of the tutors intuitiveanalysis}} * 100 \qquad (11)$$

Table 8. Comparison between predetermined speech acts and intuitive analysis.

	Comparison between predetermined speech acts and intuitive analysis				Result
	Animator	*Checker*	*Seeker*	*Independent*	
Learner 1	86,00 %	89,76 %	87,27 %	100 %	90,76 %
Learner 2	94,33 %	97,59 %	66,58 %	100 %	89,62 %
Learner 3	90,00 %	93,80 %	100 %	100 %	95,95 %
Learner 4	94,46 %	97,97 %	77,75 %	100 %	92,54 %
Total results :					**92,22 %**

Table 8 shows the margin of error between the intuitive analysis of tutors and the system analysis based on predetermined speech acts. This error margin gives a confidence degree for results validation. A result of 100% means that the system is perfectly aligned with the human tutor. In the example below, from the intuitive analysis the learner 1 emerge as animator profile with 34,88 % (see table 2) and 30,00 % according to the results of the system, compared to group learners (see table 4). We have considered the result of intuitive analysis tutors as a reference; we can see that the error rate is 14 %. All these results are averaged for every learner and profile to produce a single indicator called "Total results". This later allowed us to verify that the system has an error rate of 7.78 % compared with the intuitive analysis.

Table 9. Comparison between Naïve Bayes classification and intuitive analysis.

	Comparison between Naïve Bayes classification and intuitive analysis.				Result
	Animator	*Checker*	*Seeker*	*Independent*	
Learner 1	95,90 %	96,31%	98,21 %	100 %	97,60 %
Learner 2	97,90 %	96,44 %	76,72 %	100 %	92,77 %
Learner 3	96,00 %	97,76 %	100 %	100 %	98,44 %
Learner 4	95,96 %	97,72 %	92,50 %	100 %	96,54 %
Total results :					**96,34 %**

This table 9 shows the error margin calculation between intuitive analysis and the system based on Naive Bayes. For this version of the system, one can note that the "Total results" indicator shows that the Naïve Bayes classifier provides a significant improvement to the results. Indeed, it reduces the error margin to 3, 66 %.

5 Conclusion

The work presented in this paper proposes a system based on the automatic analysis of asynchronous textual conversations. The automatic analysis of interactions is the basic building block for the support of tutors that play a remote tutoring role in this context. More specifically, our MAS realizes a complete sequence of treatments in order to determine the sociological profiles of learners within the framework of remote collaborative process. The implementation of this type of systems often addresses problems of representation and manipulation of imperfect information. This implies the need for providing MAS with techniques of representation and knowledge manipulation, allowing them to take into consideration this imperfection. The approach we have chosen in order to remedy to this problem is to provide MAS with techniques of representation and manipulation of imperfect information [43]. In fact, the purpose of our approach is to provide a system for automatic analysis of asynchronous textual conversations in order to determine sociological profiles of learners. Therefore, we proposed a MAS based on Naive Bayes and fuzzy logic techniques, in order to resolve the problems that many of the complex systems are facing that is processing information of imperfect nature. This analysis procedure consists of four stages: recovery, filtering, lemmatization and message classification. The Naive Bayes classifier has proved to be very useful in practice. Indeed, it is suited to problems of messages categorization and has the advantage of being extremely efficient in terms of processing power in the absence of standardization of speech acts for determining social behaviors of learners. Besides, we have shown that Naïve Bayes approximation gives better results than text classification using speech acts. We have shown how the approximation Naive Bayes can be used for classification of messages and gives better results than text classification using speech acts. We have calculated indicators from the profiles defined and adapted by the works of Pléty [3]. These indicators, coupled with the classifications of messages described above, allow allocating a sociological profile to every learner. The evaluation process is influenced by the indistinctness due to the errors and to the approximations implied during the collection of the information. To treat these imperfections and to incorporate the expertise of the teacher in the evaluation of the learner, we use an approach based on the theory of fuzzy sets.

Our approach has been tested in a real situation and showed a perfect concordance between the results observed by human expert tutors and those determined automatically by our system. In the current state of this study, we plan several enhancements in order to enrich our system by semantic analysis of the text. A further research direction to follow-up our work involves a deeper study of algorithms which use semantic similarity measures to assign the appropriate senses to words depending on their

context. This will be done at the scale of a text or a corpus for automatic classification and categorization of textual conversation, in order to determine the behavioral (sociological) profile for each student.

6 References

[1] Sujana, J., Claire, M., & John, K. (2012). A visualisation tool to aid exploration of students' interactions in asynchronous online communication. Computers & Education, Volume 58, Issue 1, January 2012, (pp. 30-42).

[2] Mbala A., Reffay C., Chanier T., «SIGFAD : un système multi-agents pour soutenir les utilisateurs en formation à distance». In Actes de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH'2003), Strasbourg, France, 2003.

[3] Pléty, R. (1998). Comment apprendre et se former en groupe (Retz, 1998).

[4] Chaabi.Y, al . (2014), «An automatic system for the determination of learner's sociological behavior from textual asynchronous conversations analysis in online collaborative learning". International Conference on Intelligent Systems: Theories and Applications (SITA-14) 7-8 May 2014, IEEE.

[5] Qiang.G, al . 2010. "An Effective Algorithm for Improving the Performance of Naive Bayes for Text Classification", Second International Conference on Computer Research and Development,IEEE.

[6] Meena, M.J, al. 2009. "Naïve Bayes text classification with positive features selected by statistical method". International Conference on Advanced Computing, 2009. ICAC 2009, IEEE.

[7] Zheng.G,al. 2010. "Chinese Web Text Classification System Model Based on Naive Bayes". International Conference on E-Product E-Service and E-Entertainment (ICEEE), 2010, IEEE

[8] Huang, J. & Morgan, G. (2003). A Functional Approach to Evaluating Content Knowledge and Language Development in ESL Students' Science Classification Texts. International Journal of Applied Linguistics 13, (2), pp.234-262. https://doi.org/10.1111/1473-4192.00046

[9] S. Herring (ed.), (1996), «Posting in a different voice: Gender and ethics in computer-mediated communication». In C. Ess (ed.), Philosophical Perspectives on Computer-Mediated Communication, 115-45. Albany: SUNY Press. Reprinted (1999) in P. Mayer (ed.), Computer Media and Communication: A Reader, 241-65. New York: Oxford University Press.

[10] Pemberton, R. Edward, S.L., Or, W.W.F., and Pierson, H.D. (Eds.). (1996). «Taking Control: Autonomy in Language Learning». Hong Kong: Hong Kong University Press.

[11] Erickson T., «Social Interaction on the Net : Virtual Community as a Participatory Genre», In. Proceedings of the Thirtieth Hawaii International Conference on System Sciences, 1997, document en ligne : www. pliant. org/ personal/ Tom_Erickson/ VC_as_Genre. html

[12] Hutchby, I, (2001), «Conversation and Technology: From the Telephone to the Internet». Cambridge:Polity Press.

[13] Fornel, M.de (1989): «Une situation interactionnelle négligée : la messagerie télématique», Réseaux 38:31-48. https://doi.org/10.3406/reso.1989.1344

[14] Marcoccia, M. (2000a) :«La représentation du non verbal dans la communication écrite médiatisée par ordinateur», Communication & Organisation n° 18 : 265-274.

[15] GEORGE.S, Cécile BOTHOREL, «Conception d'outils de communication spécifiques au contexte éducatif», Revue STICEF, Volume 13, 2006, ISSN : 1764-7223, mis en ligne le 23/02/2007.

[16] Oumaira.I, al (2011). «Instrumentation des activités des tuteurs à l'aide d'un système multi-agents d'analyse automatique des interactions » A R I M A – Volume 14 pp. 125-147 – 2011.

[17] George, S (2003). «Analyse automatique de conversations textuelles synchrones d'apprenants pour la détermination de comportements sociaux», Revue STICEF, numéro spécial « technologies et formation à distance », INRP, Vol. 10, 2003, (p. 165-193)

[18] Luzzati, D. (1995). Le dialogue verbal homme-machine. Paris, Masson.

[19] Self, J. A. (1988). Bypassing the Intractable Problem of Student Modelling. Proceedings of the 1st International Conference on Intelligent Tutoring System, Montréal, Canada, (june 1-3), 18-24.

[20] Pléty, R. (1996). L'apprentissage coopérant. Lyon, ARCI Presse Universitaire, Collection Ethologie et psychologie des communications.

[21] Chaabi, Y., Messoussi, R., Hilaire, V., Ruichek, Y., Lekdioui, K., Touahni, R., Design of an Intelligent System to Support Tutors in Learning Communities Using Multi-Agent Systems and Fuzzy Logic, (2015) International Review on Computers and Software (IRECOS), 10(8), pp. 845-855. https://doi.org/10.15866/irecos.v10i8.7015

[22] Rodriguez, S., Hilaire, V., Gaud, N., Galland, S., & Koukam, A. (2011). Chapter in Self-organizing Software: From Natural to Artificial Adaptation, (first edition), Self-Organising Software From Natural to Artificial Adaptation - Natural Computing series, chapter 11, (pp. 238- 263), Springer. ISBN: 978-3642173479, 2011.

[23] Cossentino, M., Gaud, N., Hilaire, V., Galland, S., & Koukam, A. (2010). In Journal for Autonomous Agents and Multi-Agent Systems (JAAMAS). Vol. 20(2). Springer Netherlands. (pp.260-304). ISSN 1387-253. Mars, 2010.

[24] Rodriguez, S., Gaud, N., Hilaire, V., Galland, S., & Koukam, A. (2014). In the Fourth International Workshop on Engineering Self-Organizing Applications (ESOA'06). Eds. Sven Brueckner, Salima Hassas, Mark Jelasity, Daniel Yamins. (pp. 62-75). Future University-Hakodate, Japan, May, 2014.

[25] Zipf G. K. (1949) 'Human Behaviour and the Principles of Least Effort`, Addison-Wesley, 1949.

[26] Serin, S., Morova, N., Saltan, M., Terzi, S., & Karaşahin, M. (2014). Prediction of the marshall stability of reinforced asphalt concrete with steel fiber using fuzzy logic. Journal of Intelligent & Fuzzy Systems, 26(4), 1943-1950.

[27] Karakasidis, T. E., & Georgiou, D. N. (2004). Partitioning elements of the Periodic Table via fuzzy clustering technique. Soft Computing, 8(3), 231-236. https://doi.org/10.1007/s00500-003-0301-3

[28] Li.D, all. (2014). "Automatic text categorization using a system of high-precision and high-recall models". Symposium on Computational Intelligence and Data Mining (CIDM), 2014 IEEE. https://doi.org/10.1109/CIDM.2014.7008692

[29] Wang.Z, all. (2008). "Text Categorization Based on LDA and SVM". International Conference on (Volume:1) Computer Science and Software Engineering, 2008 IEEE .

[30] Yu.F, all. (2004). "Intelligence text categorization based on Bayes algorithm". International Conference on Information Acquisition, 2004. Proceedings 2004 IEEE. https://doi.org/10.1109/ICIA.2004.1373386

[31] Hassan, S, all. (2011). "Comparing SVM and naïve Bayes classifiers for text categorization with Wikitology as knowledge enrichment". 14th International Multitopic Conference (INMIC), 2011 IEEE

[32] Joachims, T. (1998). "Text categorization with support vector machines: Learning with many relevant features". European Conference on Machine Learning (ECML), 1998. https://doi.org/10.1007/BFb0026683

[33] Vinot, R., & Yvon, F. (2002). « Quand simplicit rime avec efficacit: Analyse d'un catgoriseur de textes ». Colloque International sur la Fouille de Texte CIFT'02)(pp. 17–26). Tunisie.

[34] Yang, S. P. (2002). Problem-based learning on the World Wide Web in an undergraduate kinesiology class: An integrative approach to education. Unpublished doctoral dissertation, University of New Brunswick (UMI No. MQ76421).

[35] Ritschard, G., S. Marcellin et D.A. Zighed (2009), Arbre de décision pour données déséquilibrées : sur la complémentarité de l'intensité d'implication et de l'entropie décentrée, in R. Gras, J.-C. Régnier et F. Guillet, (eds) Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités, Vol. E-16 de Revue des Nouvelles Technologies de l'Information, 207-221.

[36] Denoyer L., Apprentissage et Inférence statistique dans les bases de documents structurés: Application aux corpus de documents textuels, Thèse de Doctorat Paris 6, 2004.

[37] Sivic, Josef (April 2009). "Efficient visual search of videos cast as text retrieval". IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 31, No. 4. IEEE. pp. 591–605.

[38] Rajaraman, A.; Ullman, J. D. (2011). "Data Mining". "Mining of Massive Datasets". pp. 1–17. https://doi.org/10.1017/CBO9781139058452.002

[39] Salton, G.; Buckley, C. (1988). "Term-weighting approaches in automatic text retrieval". Information Processing & Management 24 (5): 513–523. doi:10.1016/0306-4573(88)90021-0. https://doi.org/10.1016/0306-4573(88)90021-0

[40] A. McCallum and K. Nigam. (1998), A comparison of event models for Nave Bayes text classification. Proc. Of AAAI/ICML-98 Workshop on Learning for Text Categorization, 1998, pp. 41-48.

[41] L.A. Zadeh, (1965), Fuzzy sets. Information and Control. 8, 338-353, June 1965. https://doi.org/10.1016/S0019-9958(65)90241-X

[42] Driankov; D., Hellendoorn, H. and Reinfrank, M., (1993), An Introduction to Fuzzy Control, Springer, Berlin. https://doi.org/10.1007/978-3-662-11131-4

[43] Lee, C.C., (1990), Fuzzy logic in control systems: fuzzy logic controllers—parts I and II, IEEE Trans. Syst. Man. Cybernet. 20:404-435. https://doi.org/10.1109/21.52551

7   Authors

Youness Chaabi, **Lekdioui Khadija,** and **Rochdi Messoussi** are with Telecommunication Systems and Decision Engineering Laboratory (LASTID), Ibn Tofaïl University, Kenitra, Morocco.

Fatine Jebbor is with Analysis, Modeling and Integration of Processes and Systems Team (AMIPS), Mohammed V University in Rabat, School Mohammadia of Engineers, Morocco.