FEATURE SELECTION FOR IDENTIFYING SKILLS AND PERSONAL AFFINITIES IN TUTORS AND STUDENTS

Argelia B. Urbina-Nájera[1], Jorge de la Calleja[2]*, Ma. Auxilio Medina[3]

[1]Dra., Universidad Politécnica de Puebla, MÉXICO, argelia.urbina@uppuebla.edu.mx

[2]Dr., Universidad Politécnica de Puebla, MÉXICO, jorge.delacalleja@uppuebla.edu.mx

[3]Dra., Universidad Politécnica de Puebla, MÉXICO, maria.medina@uppuebla.edu.mx

*Corresponding Author

Abstract

In this paper we present an experimental study to identify skills of tutors and students, as well as affinities between them, applying three machine learning feature selection methods for academic tutoring. Academic tutoring plays an important role, supporting students with the purpose of improving their academic performance. The feature selection methods used for experiments evaluates the worth of an attribute respect others, applying different metrics, thus we used GainRatioAttributeEval, InfoGainAttributeEval and SymmetricalUncertAttributeSetEval. Data set was collected applying a measure instrument consisting of 18 items that estimates skills such as interpersonal, intrapersonal, autonomy, self-direction, digital and communication; 16 items to identify the affinities and 2 demographic variables (sex, age). Experiments were performed using two data sets: The first one was obtained by a simple random probability sampling to select students, whose population includes 1199, with a confidence level of 90% and an acceptable error of 5%, obtaining a sample of 277 students, composed by 143 men and 134 women. The second data set was created by the criterion of voluntary participation, with 35 people; surveyed only 19 tutors, consisting of 6 women and 13 men. According to experimental results, we can conclude that to form groups of students, affinity between them is important, as well as the same skills, which according to its importance would be: intrapersonal, interpersonal, self-direction, digital and communication. In the case of tutors, it has been identified to meet the challenges, also it is important to assign them according to the similar skills to be with the group of students; and in this case, irrelevant attributes are: sex, age and affinities.

Keywords: Academic tutoring, machine learning, tutors, attribute selection, groups.

1 INTRODUCCIÓN

The current society globalized, contributes to the ongoing transformation of education demanding autonomous persons, creative, proactive, helpful and with values that benefit the growth of the same society.

In this sense, academic tutoring is an inherent element of teaching with a holistic view of education. It is a personal relationship with the student in the structure of the teaching-learning and the promotion of attitudes, skills, knowledge and interests (Castillo Arredondo, Torres González & González Polanco, 2009). That is why, it is necessary to enhance and promote a specific profile of the tutor to enable him to guide and accompany students for making decision so accurate regarding their academic process (Ruiz Martinez, & Valladares, 2012).

In this regard, authors such as Castillo Arredondo, Torres González & González Polanco (2009) suggest a profile of the tutor including the need for ethical commitment, training in all kinds of knowledge and teaching techniques, as well as an own teaching style and innate. On the other hand, the Iowa State University (2011), indicates that a tutor should have knowledge of subject matter, ability to effectively communicate, patience, respect toward others, respect for different learning styles, interest in teaching and helping others learn in small groups, ability to listen and answer questions during tutoring sessions and good time management skills. Likewise, Ruiz Martinez, & Valladares (2012) propose a profile for the tutor, which should include skills such as teaching skills in research, communication, forecasting and socialization. While Polson & Richardson (2013), suggest that tutor must exercise some control curriculum, that is, the selection and sequencing of the material to be presented, must be able to respond to students questions about the subject matter and must be able to determine when students need help in the course or practicing a skill and what kind of help is needed. Similarly, Santalucía & Cisi Solari (2014), suggest that a tutor should have human qualities such as empathy, authenticity, maturity, responsibility, sociability, scientific qualities relating to the mastery of the subject, concerning technical qualities and technical skills the study area.

Because of the different proposals mentioned above with respect to the tutor profile, in this study we used the instrument proposed by Urbina Najera, de la Calleja, Vega Lebrún, Lopez Maldonado, & Pico Gonzalez (2014), since it includes different skills that encompass important features such as: interpersonal, intrapersonal, communication, digital, autonomy in the learning process and autonomy, and includes a number of attributes that identify the personal affinities like hobbies, physical exercise, kind of reading, among others.

Competences can be defined as variables that intend to estimate the level of skills that individuals have in terms of personal skills by the categories mentioned in common for Cruz Bejarano, 2010; Sanz de Acedo Lizarraga, 2010; Repetto & Beltran, 2009; Vaello Orts, 2009; Bautista & Gonzalez F. Cabrera 2006; Bautista, Borgues, & Forés, 2006; namely: interpersonal, intrapersonal skills, communication skills, self-direction skills and digital competences. In another sense, autonomy in the learning process can be seen according to Rue (2009) as the act of students respond to specific demands of knowledge, choosing for himself only those conditions it deems necessary for the response. Finally, the affinities are the attraction or adequacy of characters, opinions, likes, etc. between two or more people (Real Academia Española, 2013), while personal interests consist tastes or inclinations activities, persons or objects; depending on social, cultural, academic and even own age factors (Nuevas Tecnologías, 2011).

On the other hand, machine learning (ML) is the branch of artificial intelligence that is dedicated to the study of agents or learning programs to evolve based on their experience, and to perform a particular task getting better. The main goal of any learning process is to use the evidence known to create a hypothesis able to respond to new situations (Mitchell, 1997). In order to create automatic programs, several processes can be performed. In particular, feature selection algorithms used tools and techniques to reduce inputs to an appropriate size for their processing and analysis. The selection of attributes not only involves reducing cardinality, that is, the imposition of an arbitrary or predefined limit on the number of attributes that can be considered to create a model, but also the choice of attributes, so that the tool modeling (Weka[1], Orange[2], RapidMiner[3], KNIME[4]) must actively select or discard the attributes in terms of their usefulness for analysis. In short, feature selection works by calculating a score for each attribute and selecting only the attributes that have earned top scores (Microsoft, 2014).

According to (Alvarez & Zwir, 2001), (Shawe-Taylor, 2006), (Rodriguez & Mislej, 2013), (Hamilton, 2014), machine learning technology has had an impact on various areas of knowledge such as information theory, statistics, neurobiology, molecular biology, economics, marketing, finance, robotics, medicine, veterinary, human resources, among others.

The aim of this work is to present an experimental study to identify skills of tutors and students, as well as affinities between them; applying three machine learning feature selection methods for academic tutoring.

[1] The University of Waikato. Weka 3: Data Mining Software in Java. http://www.cs.waikato.ac.nz/~ml/weka/
[2] University of Ljubljana. Data Mining-Fruitful and fun. Orange. http://orange.biolab.si/
[3] RapidMiner. Predictive Analytics Reimagined. https://rapidminer.com/
[4] KNIME.com AG. Open for innovation. http://www.knime.org/

2 METHODOLOGY

This section describes the methodology used for the characterization of students and tutors based on their skills and personal affinities with each other. In the first part, we detail the procedure to obtain the data set and the second part describes the instrument for data collection.

2.1 Data collection

Dataset was taken of the population of engineering students from the Polytechnic University of Puebla registered in January-April 2014 and tutors of all engineering's in the same period.

For selecting students, simple random sampling was used considering a population of 1199. However, only 277 were selecting according to confidence level of 90% and an acceptable error of 5%, therefore 143 men and 134 are women was considered, all between 18 and 28 years old.

Regarding the tutors they were selected under the criterion of voluntary participation, whose population is composed of 35 people. They were surveyed 19 tutors, (6 women and 13 men), all between 31 and 45 years old.

2.2 Instrument to collect information

As mentioned above, the instrument used to collect data, was proposed by Urbina Nájera, de la Calleja, Vega Lebrún, Lopez Maldonado & Pico González (2014), offering a range of skills tutor and student should have and that also considers personal common affinities, which is an important factor for group living.

The original instrument consisted of 38 items divided into three groups: 18 items to identify competencies, 16 items to identify affinities and 4 items that identify the demographic characteristics. In the later instrument two demographic variables were eliminated (working and marital status). The final version of the instrument applied is shown in Figure 1.

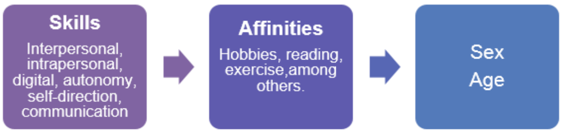

Figure 1. Attributes used in the instrument for data collection

3 ATTRIBUTE SELECTION METHODS

Weka is a software package that implements machine learning algorithms for data mining tasks. These algorithms can be applied to datasets or using JAVA code in own software applications. In addition, Weka contains tools for performing several tasks such as data pre-processing, classification, regression, clustering, association, visualization and selecting attributes (The University of Waikato, 2013).

In this case, we have used Weka for feature selection. Weka provides an attribute selection tool, thus the process is separated into two parts: choosing the attribute evaluator and the search method. The attribute evaluator is the method by which subsets of attributes are assessed. The search method permits specify the way in which the search space of possible attribute subsets is guided based on the subset evaluation. Attribute selection is normally done by searching the space of attribute subsets, evaluating each one. This is achieved by combining 1 of the 6 attribute subset evaluators with 1 of the 10 search methods. Single-attribute evaluators are used with Ranker search method to generate a ranker list from Ranker discard a given number. They can also be used in the RankSearch method (Witten, Frank, & Hall, 2011).

In this study we only used three attribute subset evaluators: GainRatioAttributeEval, InfoGainAttributeEval and SymmetricalUncerAttributeEval.

GainRatioAttributeEval*:* Evaluates the worth of an attribute by measuring the gain ratio with respect to the class, according to equation (1) (Hall, 2011).

$$\text{GainR(Class, Attribute)} = (\text{H(Class)} - \text{H(Class} \mid \text{Attribute)})/\text{H(Attribute)} \quad \ldots (1)$$

InfoGainAttributeEval: Evaluates the worth of an attribute by measuring the information gain with respect to the class, according to equation (2) (Hall, 2011).

$$\text{InfoGain(Class,Attribute)} = \text{H(Class)} - \text{H(Class} \mid \text{Attribute)} \quad \ldots (2)$$

SymmetricalUncerAttributeEval: Evaluates the worth of a set attributes by measuring the symmetrical uncertainty with respect to another set of attributes, according equation (3) (Zhao, 2003).

SymmU (AttributeSet2, AttributeSet1) =2*(H(AttributeSet2)-H(AttributeSet1 | AttributeSet2)) / H(AttributeSet2) + H(AttributeSet1) … (3)

4 EXPERIMENTAL RESULTS

In this section the results obtained are presented as follows: In the first part we present the experiment and results to identify those relevant features of tutors, considering demographics (gender and age) and experimentation without them. In the second part we show the application of the same conditions of the experiment for students.

It is noteworthy that the same instrument was used to characterize tutors and students because it seeks to identify those skills and common affinities between them to achieve homogeneity in the groups and closeness tutors with students to strengthen the process of mentoring and thus, minimize, among other situations, truancy and poor academic performance.

4.1 Attribute selection for tutors

Figure 2 shows the results obtained using the feature selection algorithms for the data set of tutors. It is noted that, with demographic variables, the most important attribute is the variable sex. As mentioned above, the data set consists of 13 men, which determines that the dominant sex is male. Now, by doing the experimentation without demographic variables, affinities are the principal attributes for selection. These results may suppose that personal affinities between students and tutors are very important to ensure integration between both of them and perhaps this helps to provide a better track students.

GainRatioAttributeEval		InfoGainAttributeEval		SymmetricalUncerAttributeEval	
With demographic variables (sex/age)	No demographic variables	With demographic variables (sex/age)	No demographic variables	With demographic variables (sex/age)	No demographic variables
1. Sex	1. Affinities	1. Sex	1. Affinities	1. Sex	1. Affinities
2. Intrapersonal	2. Self-Direction	2. Intrapersonal	2. Self-Direction	2. Intrapersonal	2. Self-Direction
3. Interpersonal	3. Interpersonal	3. Interpersonal	3. Interpersonal	3. Interpersonal	3. Interpersonal
4. Self-Direction	4. Intrapersonal	4. Self-Direction	4. Intrapersonal	4. Self-Direction	4. Intrapersonal
5. Digital	5. Autonomy	5. Digital	5. Autonomy	5. Digital	5. Autonomy
6. Age	6. Digital	6. Age	6. Digital	6. Age	6. Digital
7. Autonomy	7. Communication	7. Autonomy	7. Communication	7. Autonomy	7. Communication
8. Affinities		8. Affinities		8. Affinities	
9. Communication		9. Communication		9. Communication	

Figure 2. Results for tutors

4.2 Attribute selection for students

Figure 3 shows the results obtained for the case of students. It is observed that in all three cases the results were the same. When experimentation was done with demographic variables (sex and age) the most important attribute of the list is the sex variable, which means that for creating homogenous groups and achieve significant academic stability, it is required to consider the sex of the individual. As mentioned above, the data set consists of 143 men, which determines that the dominant sex is male. However when experimentation was performed without demographic variables, the personal affinities variable is the most important to form groups. In both types of experimentation (with and without demographics variables), the communication skills variable is less important for that. It was expected that communicative competence be at the top, because it is an indispensable tool in achieving companionship among individuals of a group.

ISBN: 978-605-64453-4-7

GainRatioAttributeEval		InfoGainAttributeEval		SymmetricalUncerAttributeEval	
With demographic variables (sex/age)	No demographic variables	With demographic variables (sex/age)	No demographic variables	With demographic variables (sex/age)	No demographic variables
1. Sex	1. Affinities	1. Sex	1. Affinities	1. Sex	1. Affinities
2. Intrapersonal	2. Self-Direction	2. Intrapersonal	2. Self-Direction	2. Intrapersonal	2. Self-Direction
3. Interpersonal	3. Interpersonal	3. Interpersonal	3. Interpersonal	3. Interpersonal	3. Interpersonal
4. Self-Direction	4. Intrapersonal	4. Self-Direction	4. Intrapersonal	4. Self-Direction	4. Intrapersonal
5. Digital	5. Autonomy	5. Digital	5. Autonomy	5. Digital	5. Autonomy
6. Age	6. Digital	6. Age	6. Digital	6. Age	6. Digital
7. Autonomy	7. Communication	7. Autonomy	7. Communication	7. Autonomy	7. Communication
8. Affinities		8. Affinities		8. Affinities	
9. Communication		9. Communication		9. Communication	

Figure 3. Results for students

5 CONCLUSIONS

Attribute selection algorithms help to enlist those most important characteristics of a data set; in this case, those skills should have a tutor and a student, to conform in a group.

Logically, if the sample has more men, the predominant sex in the formation of groups would be male. It is suggested that the data set is homogeneous, considering the same number of men and women and thus verify if the male gender would remain predominant.

When groups of students are created based on their personal affinities, it is believed that increase the sense of belonging to increase motivation among members. At the same time, feelings, ideas, successes and failures that may help to solve conflicts better and find better or new solutions and ideas for achieving the desired objectives are shared. In this regard, the role of tutors play an important role as having similar affinities, because they can help steer these solutions, foster a sense of belonging, enhance mutual trust and complete the objectives pursued in higher education.

REFERENCE LIST

Álvarez, J., & Zwir, I. (2001). *Aprendizaje Automático (AA).* Obtenido de brains and machines: http://www-2.dc.uba.ar/materias/aa/aa.html#Qu%C3%A9%20significa

Bautista, G., Borgues, F., & Forés, A. (2006). *Didáctica universitaria en entornos virtuales de enseñanza-aprendizaje.* Madrid, España: Narcea.

Bishop, C. M. (2007). *Pattern recognition and Machine Learning.* Singapore: Springer.

Cabrera Dokú, K., & González F., L. E. (2006). *Currículo universitario basado en competencias.* Barranquilla, Colombia: Ediciones Uninorte.

Castillo Arredondo, S., Torres González, J. A., & Polanco González, L. (2009). *Tutoría en la enseñanza, la universidad y la empresa.* Madrid: España: Pearson Education S.A.

Cruz Bejarán, R. M. (2010). *Espacio Docente:Portal de recursos para el docente innovador*. From El estudiante virtual: http://docentes.unibe.edu.do/noticias_x.aspx?idx=81

Hall, M. (2011). *Class GainRatioAttributeEval*. Obtenido de weka.sourceforge.net: http://weka.sourceforge.net/doc.dev/weka/attributeSelection/GainRatioAttributeEval.html

Hall, M. (2011). *Class InfoGainAttributeEval*. Obtenido de weka.attributeSelection: http://weka.sourceforge.net/doc.dev/weka/attributeSelection/InfoGainAttributeEval.html

Hamilton, L. (2014). *Six Novel Machine Learning Applications.* Obtenido de Forbes: http://www.forbes.com/sites/85broads/2014/01/06/six-novel-machine-learning-applications/

Iowa State University. (2011). *Academic Success Center*. Recuperado el 2015, de Tutoring services: http://www.dso.iastate.edu/asc/tutoring/becomeatutor

Linás González, E. C. (2009). *La orientación académica desde el bienestar universitario.* Universidad del Norte.

Microsoft. (2014). *Feature selection (Data Mining)*. Obtenido de Microsoft: https://msdn.microsoft.com/es-

es/library/ms175382%28v=sql.120%29.aspx

Mitchell, T. M. (1997). *Machine Learning.* Singapore: McGraw-Hill.

Nuevas Tecnologías. (2011). *Fernández Editores*. Obtenido de La elección de una carrera o un trabajo: http://www.tareasya.com.mx/index.php/tareas-ya/secundaria/formacion-civica-y-etica/el-individuo/1728-Habilidades,-aptitudes-e-intereses.html

Polson, M. C., & Richardson , J. J. (2013). *Foundations of Intelligent Tutoring Systems.* USA: Psychology Press.

Real Academia Española. (2013). *Diccionario de la Lengua Española*. Obtenido de afinidad: http://lema.rae.es/drae/?val=afinidad

Repetto, E., & Beltrán, S. (2009). *Formación en competencias socioemocionales.* Madrid, España: La Muralla.

Rodríguez, R., & Mislej, E. (2013). *Aprendizaje Automático - Machine Learning.* Obtenido de Departamento de Computación- Universidad de Buenos Aires: http://www.dc.uba.ar/materias/aa/2013/cuat1

Rué, J. (2009). *El aprendizaje autónomo en educación superior.* Barcelona, España: Narcea Ediciones.

Ruíz, R., Martínez, R., & Valladares, L. (2012). *Innovación en la educación superior: Hacia las sociedades del conocimiento (Google eBook).* México, D.F.: Fondo de Cultura Económica.

Sánchez Miguel, E., Paniagua Valle, G., Huguet Comelles, T., Bassedas i Ballús, E., Rodríguez Muñz, V. M., Velaz de Medrano Ureta, C.,

……... Babio Galán, M. (2011). *Orientación educativa. Modelos y estrategias de intervención.* Grao.

Santalucía, P., & Cisi de Solari, O. (2014). *Herramientas para la acción tutorial.* Buenos Aires, Argentina: Editorial Dunken.

Sanz de Acedo Lizarraga, M. L. (2010). *Competencias congnitivas en educación superior.* Madrid, España: Narcea.

Shawe-Taylor, J. (2006). *Machine learning research topics and application field.* Obtenido de Centre for Computational Statistics and Machine Learning: http://www.csml.ucl.ac.uk/courses/msc_ml/?q=node/37

The University of Waikato. (2013). *Weka 3: Data Mining Software in Java*. Obtenido de Machine Learning Group at the University of Waikato: http://www.cs.waikato.ac.nz/ml/weka/

Urbina Nájera, A. B., de la Calleja, J., Vega Lebrún, C. A., López Maldonado, N., & Pico González, B. (2014). Desarrollo y validación de un instrumento para identificar perfiles de tutorados y tutores de la modalidad virtual. *CAFVIR 2014* (págs. 227-234). Antigua Guatemala: ESVIAL.

Vaello Orts, J. (2009). *El profesor emocionalmente competente: Un puente sobre "aulas" turbulentas.* Barcelona, España: Grao.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques.* USA: MK Morgan Kaufmann.

Zhao, Z. (2003). *Class SymmetricalUncertAttributeSetEval*. Obtenido de weka.attributeSelection: http://weka.sourceforge.net/doc.packages/fastCorrBasedFS/weka/attributeSelection/SymmetricalUncertAttributeSetEval.html