

Language Learning Assisted by Group Profiling in Social Networks

<http://dx.doi.org/10.3991/ijet.v8i3.2684>

C. Troussas¹, M. Virvou¹, J. Caro² and K. J. Espinosa²

¹ University of Piraeus, Piraeus, Greece

² University of the Philippines, Quezon City, Philippines

Abstract—This paper describes an implementation for group profiling assisting language learning in social networking sites and particularly in Facebook. In this study, we have developed an educational application in Facebook for grammar learning, while the whole procedure is assisted by group profiling. We applied K-means clustering algorithm to determine the groups of students with the same learning styles and abilities, creating a more sophisticated and dynamic model. The algorithm takes as input, to initialize the process, two important students' characteristics – age and knowledge. Our study exploits the fact that educational systems have a large number of users and we use a machine learning reasoning mechanism that is based on recognized similarities between them. Future plans include evaluating our Facebook application in order to examine the degree of usefulness of the group profiling for language learning.

Index Terms—Facebook, Group profiling, Social networking services, User clustering.

I. INTRODUCTION

Currently, social networks are being adopted rapidly by millions of users worldwide, most of whom are students [5]. Social network tools support educational activities by making feasible the interaction, collaboration, active participation, information and resource sharing, and critical thinking between users [1]. Using social networks, such as Facebook, in educational and instructional contexts can be considered as a potentially powerful idea simply because students spend anyway a lot of their spare time on these online networking activities [6].

On the other hand, globalization is a phenomenon responsible for joining different cultures from all over the world, and therefore learning foreign languages has been accentuated by the emerging needs of modern life. According to the Greek National Statistical Service, there were 46465 Filipinos in Greece as of 2006 [15]. This number corresponds to the 9% of the total authorized immigrants in Greece [4]. There are also estimations that there are many more Filipinos immigrants in Greece, namely more than 50000 Filipinos [14]. For this reason, the existence of educational applications, which are oriented to the learning of the Greek language from Filipinos and vice versa, is quite significant.

Group interaction has significant educational implications in collaborative learning that support the learners' personal relationships and social interaction with their classmates. Therefore, the support of collaboration in language learning promotes the learning process. When adaptive personalized e-learning systems could accelerate the

learning process by revealing the strengths and weaknesses of each student in a collaborative environment, they could dynamically plan lessons and personalize the communication and didactic strategy. Machine learning techniques can be used for acquiring models of individual users interacting with educational systems and group them into communities or stereotypes with common interests, so that the student reap the benefits of collaboration. Group profiling concerns the construction and application of correlated data sets, based on searching large databases. The correlations and patterns that emerge can highlight specific characteristics of an already existing group or they can form the construction of the group/cluster/category itself.

Many researchers investigated the use of SNSs such as Facebook in the context of learning, specifically language learning such as in [8], [3], [7], [2], [13] and [9]. For example, in [9], the author has found out that simple activities in Facebook have helped the less language-proficient students to become more actively involved in the language learning process. In all the aforementioned scientific literature, it can be assumed that SNSs possibly played a crucial role in the field of language learning because it provides a good interactive environment for learning.

However, after a thorough investigation in the related scientific literature, we came up with the result that there is not any application in social networks and specifically in Facebook that incorporates group profiling for language learning. Hence, we implemented a prototype application, which creates clusters of students so that the educational procedure is ameliorated.

In view of the above, we have developed an educational application in Facebook for learning the grammatical phenomenon of conditionals. The users are modeled so that they can receive advice from the system. The procedure of personalized profiling is further improved by the incorporation of machine learning techniques for clustering, namely for group profiling. The prototype application combines an attractive multimedia interface in Facebook and adaptivity to individual student needs in the social network. The communication between the application and its potential users as students is accomplished through the use of web services.

II. GENERAL ARCHITECTURE

Group profiling is the construction of a descriptive profile for a provided group. In this section, we motivate this task, define the problem and explain how it is used in our application. Group profiles are computed using the rec-

ordered consumptions of the group and used as a reference for comparing the different profile aggregation strategies.

A. Aggregation-based Group Profiling (AGP)

Since group profiling aims to find features that are shared by the whole group, a natural and straightforward approach is to find attributes that are most likely to occur within the group. This aggregation-based group profiling (AGP) essentially solves the problem below [10]:

$$\max_{\{A_i\}_{i=1}^k} \sum_{i=1}^k P(A_i|+) \tag{1}$$

In our application, individual attributes in the group can be simply aggregated and the top-k most frequent features can be picked. This aggregation-based profiling is used in our system in the form of modeling groups in order to ameliorate the educational process by collaboration.

B. K-means clustering algorithm

The K-mean clustering algorithm is an iterative procedure, which is used in our application to promote group profiling (Fig. 2). In general, the algorithm assigns first an arbitrary initial cluster vector. The second step classifies each vector to the closest cluster. In the third step the new cluster mean vectors are calculated based on all the vectors in this cluster. The second and third steps are repeated until the "change" between the iteration is either quite small or zeroed. The "change" can be defined in several different ways, either by measuring the distances the mean cluster vector have changed from one iteration to another or by the percentage of vectors that have changed between iterations. The objective of the k-means algorithm is to minimize the within cluster variability. The objective function (which is to be minimized) is the sums of squares distances (errors) between each vector and its assigned cluster center.

$$SS_{\text{distances}} = \sum_{\forall x} [x - C(x)]^2 \tag{2}$$

where C(x) is the mean of the cluster that vector x is assigned to.

Minimizing the SS_{distances} is equivalent to minimizing the Mean Squared Error (MSE). The MSE is a measure of the within cluster variability [12]:

$$MSE = \frac{\sum_{\forall x} [x - C(x)]^2}{(N - c)b} = \frac{SS_{\text{distances}}}{(N - c)b} \tag{3}$$

where N is the number of vectors, c indicates the number of clusters, and b is the number of spectral bands.

K-means is very sensitive to initial starting values. For two classifications with different initial values and resulting different classification one could choose the classification with the smallest MSE (since this is the objective function to be minimized). However, for two different initial values the differences in respects to the MSE are often very small while the classifications are very different [12]. Visually it is often not clear that the classification with the smaller MSE is truly the better classification. From a statistical viewpoint, the clusters obtained by k-mean can be interpreted as the Maximum Likelihood Es-

timates (MLE) for the cluster means if we assume that each cluster comes from a spherical Normal distribution with different means but identical variance (and zero covariance).

The aim of clustering analysis is to group data in such a way that similar objects are in one cluster and objects of different clusters are dissimilar. The k-Means algorithm basically consists of the three aforementioned steps [12]:

1. an initial set of 'k' (where 'k' is the number of clusters) so-called centroids, i.e. virtual points in the data space is randomly created,
2. every point of the data set is assigned to its nearest centroid and
3. the position of the centroid is updated by the means of the data points assigned to that cluster. In other words, the centroid is moved toward the center of its assigned points.

This is done until no centroid was shifted in one iteration resulting in 'k' subsets/cluster.

The assignment of points to centroids results in a partitioning of the data space. It results in virtual borders between each two centroids, where the distance is equal on each side of the border to these centroids. This kind of partitioning is also known as a Voronoi tessellation.

In this small animation a grid of points is created with each vector as one point and 'k' randomly created centroids then the k-Means runs until convergence. After convergence the centroids are again randomly initialized and the k-Means runs again and so on. Since every possible point of the input space is assigned to one centroid the result is a Voronoi tessellation of the input space.

For the incorporation of the algorithm into the resulting application (Fig. 1) we may observe the following basic steps [12]:

For the initialization of the application the k-means algorithm receives as input, pre-stored data or data from empirical studies. It uses two fundamental characteristics which tend to influence the educational procedure, namely, the age of students along with their level of knowledge in the English language. These characteristics have been found quite significant in past language learning applications [11].

Machine learning techniques are used as a next step in order to describe efficiently the cognitive processes that underlie the student's actions along with the student's behavioral patterns and preferences.

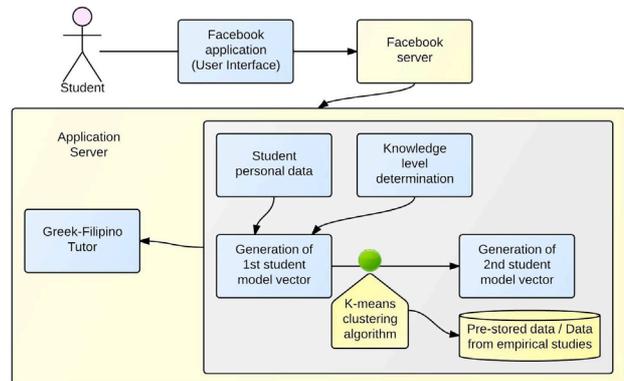


Figure 1. General architecture

Based on the aforementioned characteristics, clusters of the already existing students are creating. These clusters contain valuable information about their members, considering their behavior, their preferences and generally their interaction with the application.

Fig. 2 illustrates how the means m_1 , m_2 , m_3 and m_4 move into the centers of the clusters.

Finally, when a new user is added to the application, by registration, the student's initial characteristics (age, level of knowledge) are processed by the machine learning mechanisms to classify the student in an existing student cluster. In this way, it can offer personalized advice and error proneness, concerning the language learning, along with adaptation to students' needs.

I. GENERAL OVERVIEW OF OUR APPLICATION

Fig. 3 illustrates an indicative snapshot of the prototype educational application, where a student is asked to study the curriculum being taught. The domain model is composed of lessons about the use of conditionals. Filipino students who want to learn the conditional structures in Greek take the Greek lessons and the same way conversely. The structure includes the theory and examples for each conditional.

I. CONCLUSIONS AND FUTURE WORK

In this paper, we have described our implementation for group profiling assisting language learning in social networks, such as Facebook. Our application combines attractiveness and user-friendliness along with a framework that addresses the problem of clustering students in order to provide sophisticated and dynamic models. Clustering is conducted by the k-means algorithm, which takes as input, to initialize the process, two important students' characteristics. Our approach to student modeling exploits the fact that educational systems have a large number of users and we use a machine learning reasoning mechanism that is based on recognized similarities between them.

It is in our future plans to evaluate our Facebook application in order to examine the degree of usefulness of the group profiling for language learning. Moreover, we are planning to further evaluate the system in terms of its usefulness of the language learning support along with the efficiency of k-means clustering in the group profiling process.

REFERENCES

[1] H. Ajjan, R. Hartshorne, "Investigating faculty decisions to adopt Web 2.0 technologies: Theory and empirical tests, in *Internet and Higher Education*, Volume 11, Issue 2, 2008, pp. 71-80. <http://dx.doi.org/10.1016/j.iheduc.2008.05.002>

[2] W. Hiew, "English Language Teaching and Learning Issues in Malaysia: Learners' Perceptions via Facebook Dialogue Journal." *Journal of Arts, Science & Commerce (ResearchWorld)*, 2011: 11-19.

[3] I. Ho-Abdullah, S. H. Ruzy, J. Azhar and I. Rosnani, "Enhancing Opportunities for Language Use through Web-based Social Networking." 2011 International Conference on Social Science and Humanity. Singapore: IACSIT Press, 2011. 136-139.

[4] C. Kasimis, "Greece: Illegal Immigration in the Midst of Crisis", in <http://www.migrationinformation.org>, 2012.

[5] A. Lenhart, M. Madden, "Teens, privacy, and online social networks", Pew Internet and American Life Project Report, 2007.

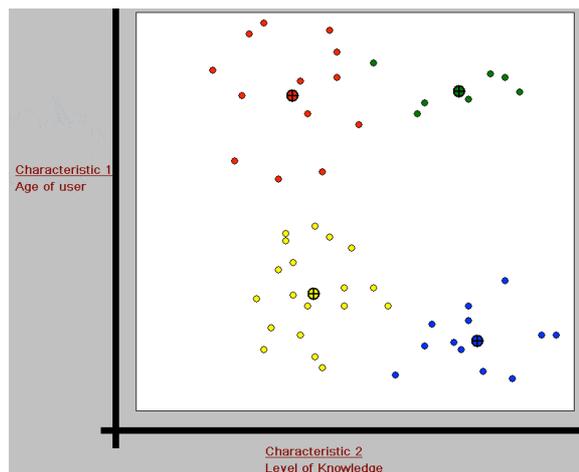


Figure 2. Means by K-means clustering algorithm

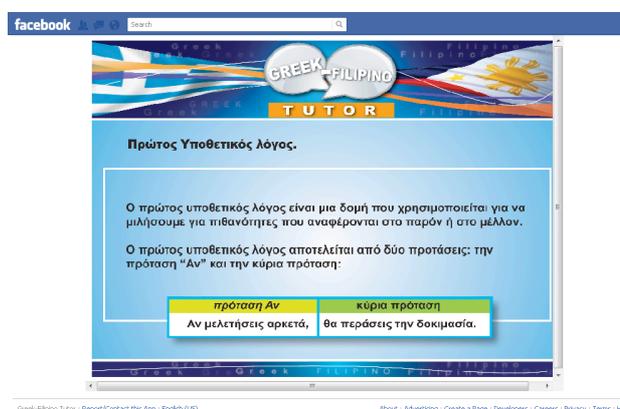


Figure 3. Part of Domain Knowledge in Greek

[6] S. Mazman, Y. Usluel, "Modeling educational usage of Facebook", *Computers and Education*, 55(2), 2010, pp. 444-453. <http://dx.doi.org/10.1016/j.compedu.2010.02.008>

[7] F. Ota, "A Study of Social Networking Sites for Learners of Japanese." Sydney: Japan Foundation Sydney, 2011.

[8] Y. Piriyaasilpa, "See you in Facebook: The Effects of Incorporating Online Social Networking in the Language Classroom." *Journal of Global Management Research*, 2011: 67-80.

[9] L. Promnitz-Hayashi, "A Learning Success Story Using Facebook." *Studies in Self-Access Learning Journal* 2, no. 4 (December 2011): 309-316.

[10] L. Tang, W. Xuwei, L. Huan, "Group profiling for understanding social structures", *ACM Transactions on Intelligent Systems and Technology*, Volume 3, Issue 1, 2011.

[11] V. Tsiriga, M. Virvou, "A Framework For The Initialization Of Student Models In Web-Based Intelligent Tutoring Systems", in *User Modelling and User-Adapted Interaction* Volume 14, Issue 4, pp. 289-316, 2004.

[12] M. Virvou, C. Troussas, E. Alepis, "Machine learning for user modeling in a multilingual learning system", *International Conference on Information Society, i-Society 2012*, Article number 6284978, 2012, pp. 292-297.

[13] M. Virvou, C. Troussas, J. Caro, K. J. Espinosa, "User modeling for language learning in Facebook", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 15th International Conference on Text, Speech and Dialogue, Volume 7499 LNAI, 2012, pp. 345-352. http://dx.doi.org/10.1007/978-3-642-32790-2_42

[14] Athens News: <http://www.athensnews.gr>

[15] Embassy of The Philippines in Greece: <http://www.philembathens.gr>

SHORT PAPER
LANGUAGE LEARNING ASSISTED BY GROUP PROFILING IN SOCIAL NETWORKS

AUTHORS

C. Troussas is a Ph.D. Candidate in the Department of Informatics at the University of Piraeus, Greece (e-mail: ctrouss@unipi.gr).

M. Virvou is Professor and Head of the Department of Informatics at the University of Piraeus, Greece (e-mail: mvirvou@unipi.gr).

J. Caro is Professor of Computer Science and Assistant Vice President for Development at University of the Philippines Diliman (e-mail: jdlcaro@dcs.upd.edu.ph).

K. J. Espinosa is Assistant Professor at the Department of Computer Science, University of the Philippines Cebu (e-mail: kspinoso@up.edu.ph).

Received 15 April 2013. Published as submitted by the authors 02 June 2013.