

Information Theoretic Clustering for an Intelligent Multilingual Tutoring System

<http://dx.doi.org/10.3991/ijet.v8i6.3235>

C. Troussas and M. Virvou
University of Piraeus, Piraeus, Greece

Abstract—People working as groups, collaborating, rather than people working individually, has unquestionably helped them develop and make accomplishments beyond our imagination. It is quite common to believe that human beings have an inner need to act as social beings. In computer science and particularly in intelligent tutoring systems, the related scientific literature enhances the conviction that even in learning, humans as students may improve the way they learn when working in groups and being in clusters. In this paper, we present the information theoretic clustering built up in the context of student collaboration in a learning environment of English and French languages. Collaborative student groups are created with respect to the corresponding user models.

Index Terms—information theoretic learning, intelligent tutoring systems, multiple language learning, user clustering.

I. INTRODUCTION

During the last few years, the ever increasing role that computers can play in education has been emphasized and this evolution of including computer-assisted learning in addition to data clustering, has been realized [1]. The massive use of intelligent tutoring systems widens the avenues of learning. Furthermore, young people's everyday practice is consequently interconnected with their language learning activities [2]. Consequently, on the one hand, we have societal knowledge of young people's considerable interaction with computers using English and on the other hand, we have language education that still has to increase the understanding of changing conditions for languages to make more informed use of the language-learning potentials in this new arena. When incorporating computers in educational contexts, we have to recognize that young people's engagement in them during their everyday life belongs to their "self-directed practices" [3], which are different from school practices in many respects. This fact implies that students have different ways of representing and expressing themselves in the naturally occurring linguistic activities during their everyday life in computers compared to the more instructionally designed language teaching and learning practices in schools. Therefore, when computer-assisted approaches are used for educational purposes, the discrepancies in the views of learning, namely what counts as knowledge and the goals of the different practices, implicitly lead to tensions and practical challenges. Furthermore, there is not always an awareness of the divergence [4].

User clustering has significant educational implications in language learning and supports the learners' personal relationships and collaboration with their peers. Therefore, the support of classification in language learning promotes the educational process. When adaptive personalized e-learning systems could accelerate the learning process by revealing the strengths and weaknesses of each student, they could dynamically plan lessons and personalize the communication and didactic strategy. Machine learning techniques are used for acquiring models of individual users interacting with educational applications and group them into communities or stereotypes with common interests. User clustering 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.

The authors of this paper have used the k-means algorithm in order to cluster the students in a multiple language learning environment and ameliorate the learning process in [1], [5], and [6]. Selecting an appropriate distance measure (or metric) is fundamental to many learning algorithms such as k-means. However, choosing such a measure is highly problem-specific and ultimately dictates the success (or failure) of the learning algorithm. To this end, there have been several recent approaches that attempt to learn distance functions [7], [8]. These methods work by exploiting distance information that is intrinsically available in many learning settings. Concerning the problem of semi-supervised clustering, points are constrained to be either similar (namely, the distance between them should be relatively small) or dissimilar (the distance should be larger). In information retrieval settings, constraints between pairs of distances can be gathered from click-through feedback. In fully supervised settings, constraints can be inferred so that points in the same class have smaller distances to each other than to points in different classes [9]. Although existing algorithms for metric learning have been shown to perform well across various learning tasks, each fails to satisfy some basic requirement. First, a metric learning algorithm should be sufficiently flexible to support the variety of constraints realized across different learning paradigms and second, the algorithm must be able to learn a distance function that generalizes well to unseen test data [9]. Finally, the algorithm should be fast and scalable.

In view of the above, we propose a novel approach of information theoretic clustering, based on entropy. Our approach generalizes the standard Euclidean distance, used in k-means clustering algorithm, by admitting arbitrary linear scaling and rotations of the feature space

and models the problem in an information-theoretic setting. In this way, we offer the possibility of qualitative collaboration among students of the same cluster, so that they are capable of succeeding in multiple language learning, namely in the learning of the English and French language.

This paper is organized as follows. First, we present related scientific work. In section 2, we discuss the entropy-based information theoretic clustering and the general architecture system's architecture. Following, in section 3, we present an evaluation concerning the student clustering. Finally, in section 4, we come up with a discussion about the usability of the resulting system and we present our future plans.

II. RELATED WORK

In this section, we present the related scientific work, firstly related to machine learning in education and user clustering and secondly to Intelligent Tutoring Systems (ITSs).

A. Machine learning in education and User Clustering

In [10], the authors proposed the exploitation of machine learning techniques to improve and adapt the set of user model stereotypes by making use of user log interactions with the system. To do this, a clustering technique is exploited to create a set of user models prototypes; then, an induction module is run on these aggregated classes in order to improve a set of rules aimed as classifying new and unseen users. Their approach exploited the knowledge extracted by the analysis of log interaction data without requiring an explicit feedback from the user.

In [11], the author presented a snapshot of what has been investigated in terms of the relationship between machine translation (MT) and foreign language (FL) teaching and learning. Moreover, the author outlined some of the implications of the use of MT and of free online MT for FL learning.

In [12], the authors investigated which human factors are responsible for the behavior and the stereotypes of digital libraries users so that these human factors can be justified to be considered for personalization. To achieve this aim, the authors have studied if there is a statistical significance between the stereotypes created by robust clustering and each human factor, including cognitive styles, levels of expertise and gender differences.

In [13], the authors focused on machine learning approaches for inducing student profiles, based on Inductive Logic Programming and on methods using numeric algorithms, to be exploited in this environment. Moreover, an experimental session has been carried out from the authors, comparing the effectiveness of these methods along with an evaluation of their efficiency in order to decide how to best exploit them in the induction of student profiles.

In [14], the authors studied the problem of unsupervised domain adaptation, which aims to adapt classifiers trained on a labeled source domain to an unlabeled target domain, since many existing approaches first learn domain-invariant features and then construct classifiers with them. They propose a novel approach that jointly learn the both.

In [9], the authors presented an information-theoretic approach to learning a Mahalanobis distance function and

formulate the problem as that of minimizing the differential relative entropy between two multivariate Gaussians under constraints on the distance function.

In [15], the authors presented an organized study of information theoretic measures for clustering comparison. They have shown that the normalized information distance (NID) and normalized variation of information (NVI) satisfy both the normalization and the metric properties. Between the two, the NID is preferable since the tighter upper bound of the MI used for normalization allows it to better use the $[0,1]$ range. They highlighted the importance of correcting these measures for chance agreement, especially when the number of data points is relatively small compared with the number of clusters.

In [16], the authors proposed to embed the clustering problem into a Bayesian framework to automatically detect the number of clusters. The entropy is considered to define a prior and enables them to overcome the problem of defining a priori the number of clusters and an initialization of their centers. A deterministic algorithm derived from the standard k-means algorithm was proposed and compared with simulated annealing algorithms.

B. Intelligent Tutoring Systems

In [6], the authors presented an electronic learning platform that addresses the problem of grouping students in order to provide sophisticated user models.

Another CALL system is SignMT, presented in [17], which tries to translate sentences/phrases from different sources in four steps, which are word transformation, word constraint, word addition and word ordering.

Another computer-based program on second language acquisition is Diglot Reader, presented in [18], which is used in a way that students may read a native language text with second language vocabulary and grammatical structures increasingly embedded within the text.

TAGARELA is an individualized instruction program, presented in [19], which analyzes student input for different activities and provides individual feedback.

Finally, in [20], the authors described a ubiquitous e-learning tutoring system for multiple language learning, called CAMELL (Computer Assisted Multilingual E-Language Learning). It is a post-desktop model of human-computer interaction in which students "naturally" interact with the system in order to get used to electronically support computer-based learning. Their system presents advances in user modeling, error proneness and user interface design.

In [21], the author discussed the effects of the inclusion of a Facebook activity as part of the language classroom. They examined students' feedback toward the activity and the use of language in their interaction. Based on their findings, the said activity proved to be advantageous because it kept the students connected with their friends while they take the language course.

In [22], the author investigated the benefits of SNS communities for learning Japanese as a second language (L2). They found out that SNSs has provided a "a portal for L2 learners to access other information and sources, and present a safe introduction to wider communication in the L2". Seeing the usefulness of the SNSs communities as a learning tool, they recommend that "researchers and

educators should direct their attention to such new tools, and the communities that form around them”.

In [23], the author has found out that simple activities in Facebook has helped the less language-proficient students to become more actively involved in the language learning process.

In [24], the author investigated the possibility of gathering learners’ experience and views pertaining to issues on English language learning problems in secondary school, college or university, gathering learners’ views on English language teaching and learning in secondary school, college and university and gathering learners’ suggestions on ways to enhance English language learning and teaching.

In [25], the authors focused on the utilization of Facebook to enhance students’ interaction using English in the English Speaking Zone. In undertaking this study, the authors had two research objectives: to understand the ways that social network can be used to provide the students with opportunities to use English as a medium of communication, and to reflect on and improve our own practice with an aim to create an environment that can continue to be used to enhance opportunities to use the language.

In [26], the authors presented the multivariate clustering for group learning in Facebook. By the incorporation of the multiple parameter based clustering, the problem of grouping students is further treated. Multivariate clustering is conducted by the k-means algorithm which takes as input multiple students’ characteristics.

In [27], the authors presented the misconception diagnosis in multiple language learning over social networks, assisted by user clustering. In particular, they introduced a prototype Facebook application that can diagnose the errors of users, who are confused while they learn multiple languages.

In [28], the authors 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.

However, after a thorough investigation in the related scientific literature, we came up with the result that there are not educational systems that incorporate the information theoretic learning for clustering. In this way, students can be classified in order to better collaborate. Hence, this is another step for the amelioration of the educational process.

III. ESTIMATING ENTROPY

Entropy plays an important role in forming many information theoretical quantities. In this paper, we are interested in how it is related to $I(X; Y)$, the mutual information between a random variable X and its cluster membership Y and the techniques for entropy, as precisely described in [15], have been also used in our research. Concretely [15],

$$I(X; Y) = H(X) - H(X|Y), \quad (1)$$

where the right-hand-side is the difference between the entropy and the conditional entropy of X .

For high-dimensional X , estimating entropies from samples is a challenging problem. Standard approaches include quantizing/binning X or constructing density estimators of X and they often do not work well due to the “curse of dimensionality”, where the number of data points grows exponentially in order to obtain an accurate estimation [15]. There has been a growing interest in using non-parametric statistics to estimate entropies. Specifically, it was shown that, given N samples $D = \{x_1, \dots, x_N\}$ where $x_i \in \mathbb{R}^D$, the entropy of X can be estimated [15] by

$$\hat{H}_k(X) = \frac{D}{N} \sum_i \log \|x_i - x_i^{(k)}\|_2^2 + \text{const}, \quad (2)$$

where $x_i^{(k)}$ is the k -th nearest neighbor of x_i in D [29], [30]. The estimator approaches $H(X)$ with a convergence rate of $O(1/\sqrt{N})$. Averaging $\hat{H}_k(X)$ over all possible k from 1 to $(N-1)$ leads to a simplified estimator [15],

$$\begin{aligned} \hat{H}(X) &= \frac{1}{(N-1)} \sum_{k=1}^{N-1} \hat{H}_k(X) \\ &= \frac{D}{N(N-1)} \sum_{i \neq j} \log \|x_i - x_j\|_2^2 + \text{const}. \end{aligned} \quad (3)$$

This estimator was first investigated in [31] and can be understood intuitively as follows. To estimate the entropy, one would need to obtain an unbiased estimator of $-\log p(x_i)$ such that [15]

$$H(X) \approx \frac{1}{N} \sum_i -\log p(x_i). \quad (4)$$

For one-dimensional X , we can approximate $p(x_i)$ as a uniform distribution between x_i and x_j , which gives rise to [15]

$$-\log p(x_i) \approx -\log \frac{1}{|x_i - x_j|}. \quad (5)$$

Averaging this estimator over all possible $x_j \neq x_i$, an estimator is obtained in the form of (3). A detailed derivation of this result is given in [8]. $\hat{H}_k(X)$ is more computationally convenient than $\hat{H}(X)$ as it does not need to identify nearest neighbors. Thus, the focus is on $\hat{H}_k(X)$ in the rest of the paper. For the conditional entropy $H(X|Y)$, we estimate it with the following [15]

$$\hat{H}(X|Y) = \sum_y \hat{p}(Y = y) \hat{H}(X, Y = y), \quad (6)$$

where $\hat{p}(Y = y)$ is the (empirical) prior distribution and $\hat{H}(X, Y = y)$ is the entropy of data samples whose corresponding Y is y . Specifically, in the context of clustering, Y stands for cluster memberships. We assume that there are K clusters, each with N_k data points. The conditional entropy is thus given by (up to a constant) [15]

$$\begin{aligned} \hat{H}(X|Y) &= \sum_{k=1}^K \frac{N_k}{N} \frac{D}{N_k(N_k-1)} \sum_{Y_i=Y_j=k} \log \|x_i - x_j\|_2^2 \\ &= \sum_{k=1}^K \frac{D}{N(N_k-1)} \sum_{Y_i=Y_j=k} \log \|x_i - x_j\|_2^2 \end{aligned} \quad (7)$$

where the inner summation is over data points x_i and x_j which are both assigned to the cluster k .

Since the entropy $H(X)$ does not depend on how data points to different clusters are assigned, maximizing the mutual information – a clustering criterion to be described in detail in the next section – is equivalent to minimizing the conditional entropy. Further insight is gained by contrasting the conditional entropy to the criterion minimized in the K-means [15]:

$$J(X, Y) = \sum_{k=1}^K \sum_{Y_i=k} \|x_i - \mu_k\|_2^2 = \sum_{k=1}^K \sum_{Y_i=Y_j=k} \frac{1}{N_k} \|x_i - x_j\|_2^2 \quad (8)$$

where μ_k is the centroid of the cluster k . Both $H(Y|X)$ and $J(X, Y)$ measure how tight the clusters are. However, the conditional entropy uses the logarithm of the distances, which grows slower than the linear function used by the K-means and showed in past research of the authors [1], [5] and [6]. Thus, arguably the conditional entropy tends to be less sensitive and more robust to outliers. In the following, we describe how optimal clustering in the information theoretical sense can be obtained.

IV. INFORMATION THEORETIC CLUSTERING AND GENERAL ARCHITECTURE

The conditional entropy of (7) depends on the cluster memberships of the data points. The minimization of this quantity over all possible assignments is referred as information theoretical clustering (ITC) in [31]. Experimental results reported there have shown this is an effective and useful clustering criterion. Despite its similarity to the K-means objective function in (8), minimizing the conditional entropy does not admit the two-step alternate minimization procedure often used in K-means. Specifically, it is not obvious how to define a single centroid (as in K-means) for each cluster and iteratively update the locations of these centroids.

Instead, in [31], the authors proposed a local search procedure that greedily assigns data points to clusters. The procedure starts with a random assignment. Then a data point is cyclically but randomly chosen from D and evaluated. If changing its cluster membership k to a different k' would result in reduction in the conditional entropy, the data point's cluster membership will be updated to k' . It is easy to see that the procedure converges to a local optimum. Furthermore, determining whether to change assignments can be performed efficiently, involving at most $(N-1)$ calculations of distances. However, there is no rigorous analysis on how many such evaluations are needed in order to converge or how good the converged solution is in terms of approximation guarantee.

In what follows, we show how we can apply information theoretic clustering for a multilingual learning system and we incorporate the techniques described in [32].

In data clustering, a label $l_k \in \{1, \dots, C\}$, $k=1, \dots, N$, is assigned to each d -dimensional data point x_k , $k=1, \dots, N$, comprising a data set and the data points with the same label form a group in the data set [32]. Each group should

correspond to a "natural" cluster of data points, such that members of the same cluster are more similar, in some sense, than members of different clusters and in this way, collaboration between members of the same cluster can enhance the tutoring. In the current information theoretic clustering approach, similarity is quantified by a divergence measure, $D(p_1, \dots, p_C)$, between the probability density functions (pdfs), $p_1(x), \dots, p_C(x)$, of the clusters and the divergence measure is based on the quadratic Renyi entropy [33], which is an information theoretic measure of the uncertainty of a random variable [32]. The goal is to assign labels such that an estimate of the divergence is maximized, because this corresponds to clusters which are maximally distant in terms of the cluster pdfs [32].

The quadratic Renyi entropy is given by [32]

$$H_2(p) = -\log \int p^2(x) dx \quad (9)$$

where $p(x)$ is the pdf. It is a special case of the family of Renyi entropies, $H_\alpha(p) = -\log \int p^\alpha(x) dx$, where α is the Renyi entropy parameter (Renyi, 1976). In the limit $\alpha \rightarrow 1$ the Shannon entropy is obtained.

Consider two pdfs, $p_1(x)$ and $p_2(x)$. By the Cauchy-Schwarz (CS) inequality, the following holds [32]

$$\int p_1(x)p_2(x) dx \leq \sqrt{\int p_1^2(x) dx} \sqrt{\int p_2^2(x) dx} \quad (10)$$

Using (10), a CS divergence measure between these two pdfs may therefore be defined as [34]

$$D_{CS}(p_1, p_2) = -\log \frac{\int p_1(x)p_2(x) dx}{\sqrt{\int p_1^2(x) dx} \sqrt{\int p_2^2(x) dx}} \quad (11)$$

This measure is principled in the sense that it is additive under statistical independence (joint densities are products of marginal densities), a fundamental requirement for an information measure, it is symmetric and obeys $D_{CS}(p_1, p_2) \in [0, \infty)$. It is zero only if $p_1(x) = p_2(x)$. It does not satisfy the triangle inequality and is for that reason not a distance metric.

The CS divergence is based on the quadratic Renyi entropy since $D_{CS}(p_1, p_2) = -\log [p_1(x)p_2(x) dx - 1/2H_2(p_1) - 1/2H_2(p_2)]$, where $-\log [p_1(x)p_2(x) dx]$ can be considered a "cross" Renyi entropy between $p_1(x)$ and $p_2(x)$. As shown in the next section, the relation to the Renyi entropy provides a convenient estimator for this quantity.

The CS divergence may be extended to the C-cluster case as follows [32]

$$D_{CS}(p_1, \dots, p_C) = -\log \frac{1}{\kappa} \sum_{i=1}^{C-1} \sum_{j>i} \frac{\int p_i(x)p_j(x) dx}{\sqrt{\int p_i^2(x) dx} \sqrt{\int p_j^2(x) dx}}, \quad (13)$$

where

$$\kappa = \sum_{c=1}^{C-1} c. \quad (14)$$

To be useful as a clustering cost function, the sample-based estimator of the CS divergence must be expressed in terms of the labels of the data points. The goal of the current clustering approach is precisely to adjust the labels, such that the cost function used obtains its

maximum value. This optimization procedure may be implemented in several different ways. In [34], the authors first devised an exhaustive search optimization. This means that every clustering possibility had to be examined by computing the cost function in each case, for then to pick the best result in the end. Since the computation of the cost function scales as $O(N^2)$, where N is the number of data points, this method became prohibitive for anything but very small and manageable data sets. The results obtained were however very promising, and instigated research focused on more efficient optimization rules.

Such an information theoretic divergence measure captures directly the statistical information contained in the data as expressed by the probability density function and can thus produce non-convex cluster boundaries. Student clustering (Fig. 1) offers the possibility of collaboration among students of the same cluster and thus the multiple learning of English and French can be ameliorated.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The approval of the educational software by Software human instructors and students consists one of the most significant aspects that play crucial role in the success of the process. Educational systems have a large number of users, by whom they should be approved. Hence, evaluation of this kind of software is an important phase that has to follow development at all times and, in particular, formative evaluation is one of the most critical steps in the development of learning materials, because it helps the designer improve the cost-effectiveness of the software and this increases the likelihood that the final product will achieve its stated goals [35]. The modeling of instructors and students constitutes the cornerstone of the educational process. Hence, in the scientific literature, there have been presented evaluation methods which are completely adjusted to educational software.

One such evaluation framework outlines three dimensions to evaluate: (i) context; (ii) interactions; and (iii) attitudes and outcomes [36]. The context determines the reason why the educational software is adopted in the first place, namely the underlying rationale for its development and use, users' interactions with the software reveal information about the users' learning processes and the "outcome" stage examines information from a variety

of sources, such as pre and post achievement tests, interviews and questionnaires with students and tutors [35].

The aforementioned framework, as precisely described in [35], has been used for the evaluation of our multiple language learning application. The underlying rationale of our system involves providing educational features, since ITSs provide opportunities within professional education, curriculum education, and learning. Furthermore, the context of the evaluation required an emphasis on the information theoretic clustering framework of the application. Moreover, users' interactions with the application were evaluated with respect to the users' learning processes. Finally, the "outcomes" stage involved pre and post-achievement tests before and after the use of the application. In addition, it involved questions to students and instructors, which focused mainly on evaluating the use of the aforementioned framework in the application.

In view of the above, the evaluation of our system involved both instructors and students and was conducted in two different phases. At the first phase, only the instructors took place at the evaluation. The second phase concerned the evaluation of the resulting educational application and involved both instructors and students. The instructors of the second phase were exactly the same as in the first phase, so that they could have a complete experience with our language learning application. At the first and second phase, 4 instructors participated in the evaluation. All of them are private instructors in the English language and were asked to use our system, namely to check in detail the tutoring of the multiple languages. All of the instructors who participated in the experiment were familiar with the use of computers.

When asked, all of the instructors confirmed that our language learning application had a user-friendly interface and that the clustering procedure was satisfactory. More specifically, three of them stated that they found the application very useful, while one of them stated that the application is useful. Concerning the question "How do you see the possibility of the system to create clusters and are you satisfied with the cluster, in which the system classified you?", the instructors stated that they were all very satisfied from the possibility of the system to create clusters and from the cluster in which they were classified.

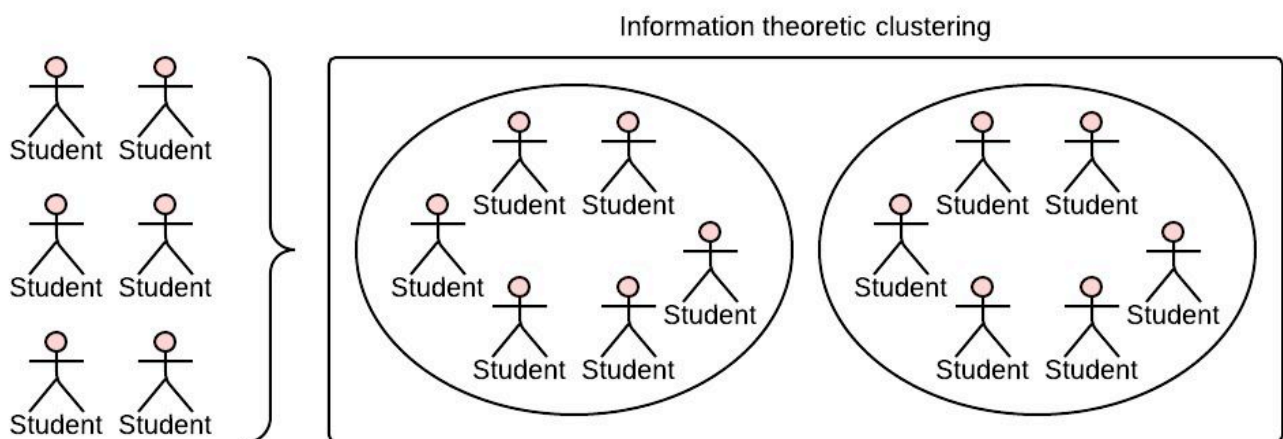


Figure 1. Information theoretic clustering for students

The second phase involved in total 60 users, from which 30 users are undergraduate students and 30 users are postgraduate students. The underlying rationale of the application lies on the hypothesis that these applications are more convenient and flexible to use (because of the fact that students can collaborate with their peers) while they retain the educational quality. At a first glance, the validity of this hypothesis might look obvious. However, there may be users who are not familiar with educational software in general and thus might not like the particular applications. On the other hand, there may be students, who are very familiar with computers and are eager to use them for educational purposes. Hence, one important aspect of the evaluation was to find out whether users were indeed helped by the whole environment.

After the completion of the use of the application, the users were asked about their experience. 77% found satisfactory the possibility of the system to create clusters and were satisfied by the cluster in which they were classified. Moreover, the students were very pleased and enthusiastic by the opportunity that our application gave them the possibility to collaborate with their peers, namely the members of the same cluster. Fig. 2 illustrates a pie-chart, representing this result (users' answers).

VI. CONCLUSIONS AND FUTURE WORK

ITSs have already become very popular among students and thus a new era has been imposed. However, many issues should be taken seriously into account so that the resulting applications can be educationally beneficial to users and the tutoring process is promoted. Moreover, instructors should be included in the educational design of such applications, as an integral part of the educational process.

Information theoretic clustering is the process of grouping, or clustering, the items comprising a data set, according to a divergence measure between probability density functions based on Renyi's quadratic entropy. Such an information theoretic divergence measure captures directly the statistical information contained in the data as expressed by the probability density function and can thus produce non-convex cluster boundaries. As is well known K-means clustering produce only convex clusters. The information divergence measure is estimated directly from the data with pairwise sample interactions using concepts from kernel density estimation.

In this paper, we have described how information theoretic learning can be used in the clustering process of students, so that the emerging clusters can offer the possibility of collaboration among their members and thus the educational procedure may be assisted. Our application for multiple language learning is designed to provide the relatively facilities to both instructors and users, while retaining high quality of the educational application with respect to interactivity, adaptivity, collaboration and user classification. The research, which was conducted, has resulted in the development of the aforementioned application that has been evaluated among instructors and users. The evaluation results were very encouraging, both for our language learning application. The results showed that the contribution of the entropy-based information theoretic clustering can be appreciated by instructors and simple users and can ameliorate the teaching of languages.

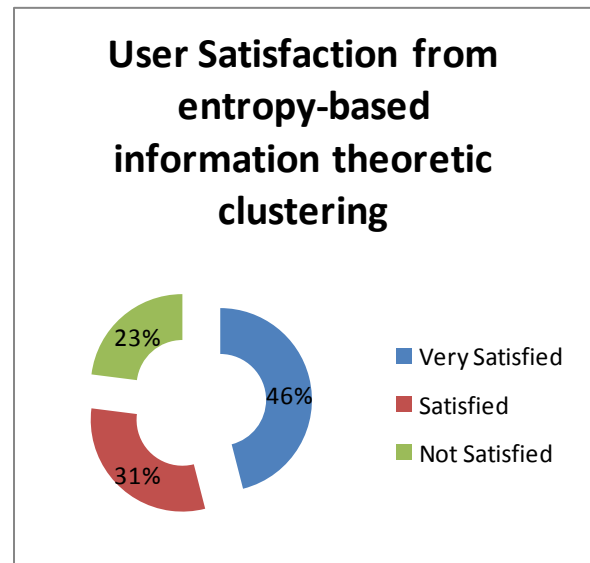


Figure 2. Rate of satisfaction from user clustering.

It is in our future plans to further evaluate our multilingual language learning system in order to examine the degree of usefulness of the individualized learning offered by our system. Furthermore, we are planning to further evaluate the system in terms of its usefulness of the multilingual support along with the efficiency of such applications within the education.

REFERENCES

- [1] Troussas, C., Virvou, M., & Alepis, E., "Multiple parameter cluster analysis in a multiple language learning system", in *The Fourth International Conference on Information, Intelligence, Systems and Applications*, 2013, pp. 281, under publication.
- [2] Lantz-Andersson, A., Vigmo S., & Bowen R., "Crossing boundaries in Facebook: Students' framing of language learning activities as extended spaces", in *Computer-Supported Collaborative Learning* (2013) 8:293-312.
- [3] Drotner, K. (2008), "Leisure is hard work: Digital practices and future competencies", in *D. Buckingham (Ed.), The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning* (pp. 167-184). Cambridge, MA: The MIT Press.
- [4] Bonderup Dohn, N. (2009). "Web 2.0: Inherent tensions and evident challenges for education", in *Computer-Supported Collaborative Learning*, 4(3), 343-363.
- [5] Troussas, C., Virvou, M., & Alepis, E. (2013). Comulang: towards a collaborative e-learning system that supports student group modeling. *SpringerPlus*, vol. 2, issue 387. <http://dx.doi.org/10.1186/2193-1801-2-387>
- [6] Virvou M., Troussas C., Alepis E., "Machine learning for user modeling in a multilingual learning system", in *International Conference on Information Society, i-Society 2012*, Article number 6284978, pp. 292-297, 2012.
- [7] Weinberger, K.Q. and Saul, L.K., "Distance metric learning for large margin nearest neighbor classification", *JMLR*, 10:207{244, 2009.
- [8] Goldberger, J., Roweis, S., Hinton, G., and Salakhutdinov, R. Neighbourhood components analysis. NIPS, 2004.
- [9] Davis, J., Kulis, B., Jain, P., Sra S., & Dhillon, I. Information Theoretic Metric-Learning, in *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, OR, 2007.
- [10] Basile T., Esposito F., and Ferilli S., "Improving User Stereotypes through Machine Learning", in *Communications in Computer and Information Science*, vol. 249, pp. 38-48, 2011.

- [11] Niño A., "Machine translation in foreign language learning: Language learners and tutors perceptions of its advantages and disadvantages", in *ReCALL*, vol. 21, pp. 241-258, 2009.
- [12] Friaz-Martinez E., Chen S. Y., Macredie R. D., and Liu X., "The role of human factors in stereotyping behavior and perception of digital library users: a robust clustering approach", in *User Modelling and User-Adapted Interaction*, vol. 13, pp. 305-337, 2007.
- [13] Licchelli O., Basile T. M. A., Di Mauro N., Esposito F., Semeraro G., and Ferilli S., "Machine learning Approaches for inducing student models", *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, vol. 3029, pp. 935-944, 2004.
- [14] Shi, Y., & Sha, F., "Information-Theoretical Learning of Discriminative Clusters for Unsupervised Domain Adaptation", in *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.
- [15] Vinh, N. X., Epps, J., & Bailey, J., "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance", in *Journal of Machine Learning Research 11* (2010) 2837-2854.
- [16] Palubinskas G., Descombes X., and Kruggel F., "An unsupervised clustering method using the entropy minimization", in *ICPR '98*, Brisbane, Australia, 1998.
- [17] Ditcharoen N., Naruedomkul K., and Cercone N., "SignMT: An alternative language learning tool", in *Computers and Education*, vol. 55, pp. 118-130, 2010. <http://dx.doi.org/10.1016/j.compedu.2009.12.009>
- [18] Christensen E., Merrill P., and Yanchar S., "Second language vocabulary acquisition using a diglot reader or a computer-based drill and practice program", in *Computer-Assisted Language Learning*, vol. 20, pp. 67-77, 2007.
- [19] Amaral L., and Meurers D., "Conceptualizing student models for ICALL", in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4511 LNCS, pp. 340-344, 2007.
- [20] Virvou M., and Troussas C., "CAMELL: Towards a ubiquitous multilingual e-learning system", in *CSEDU 2011 - Proceedings of the 3rd International Conference on Computer Supported Education*, vol. 2, pp. 509-513, 2011.
- [21] Piriyaasilpa Y., "See you in Facebook: The Effects of Incorporating Online Social Networking in the Language Classroom", *Journal of Global Management Research*, pp. 67-80, 2011.
- [22] Ota F., "A Study of Social Networking Sites for Learners of Japanese", *Sydney: Japan Foundation Sydney*, 2011.
- [23] Promnitz-Hayashi L., "A Learning Success Story Using Facebook", *Studies in Self-Access Learning Journal 2*, no. 4, pp. 309-316, 2011.
- [24] Hiew W., *English Language Teaching and Learning Issues in Malaysia*, Learners' Perceptions via Facebook Dialogue Journal." *Journal of Arts, Science & Commerce (ResearchWorld)*, 2011: 11-19.
- [25] Ho-Abdullah I., Ruzy S. H., Azhar J., and Rosnani I., "Enhancing Opportunities for Language Use through Web-based Social Networking", *2011 International Conference on Social Science and Humanity*. Singapore: IACSIT Press, pp. 136-139, 2011.
- [26] Troussas, C., Virvou, M., Caro, J., & Espinosa, K. J., "Multivariate clustering for group language learning for Facebook", in *Intelligent Interactive Multimedia Systems and Services - Proceedings of the 6th International Conference on Intelligent Interactive Multimedia Systems and Services (IIMSS)*, vol. 254, 2013, pp. 80-88.
- [27] Troussas, C., Virvou, M., Vougiouklidou, A., & Espinosa, K. J., "Automatic misconception diagnosis in multiple language learning over social networks", in *The Fourth International Conference on Information, Intelligence, Systems and Applications*, 2013, pp. 293-298, under publication.
- [28] Troussas, C., Virvou, M., Caro, J., & Espinosa, K. J. (2013). Language learning assisted by group profiling in social networks. *International Journal of Emerging Technologies in Learning*, vol. 8, issue 3, pp. 35-38.
- [29] Wang Q., Kulkarni S.R., and Verdú S., "Divergence estimation for multidimensional densities via k-nearestneighbor distances", *IEEE Transactions on Information Theory*, 55(5):2392-2405, 2009. ISSN 0018-9448.
- [30] Wang, M., & Sha, F., Information Theoretical Clustering via Semidefinite Programming, in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 2011*, Fort Lauderdale, FL, USA. Volume 15 of JMLR: W&CP 15.
- [31] Faivishevsky L., and Goldberger J. "A nonparametric information theoretic clustering algorithm", in *Proceedings of the 27th international conference on Machine learning*, pages 351-358. ACM, Omnipress, 2010.
- [32] Jenssen R., & Príncipe J. C., "Information theoretic clustering", *Scholarpedia*, vol. 5, 2010.
- [33] Renyi A. (1976). "On measures of entropy and information", *Selected Papers of Alfred Renyi, Akademiai Kiado*, Budapest, 2:565-580.
- [34] Gokcay E. and Principe J. C. (2002), "Information Theoretic Clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):158-170. <http://dx.doi.org/10.1109/34.982897>
- [35] Virvou M., Alepis E., "Mobile educational features in authoring tools for personalized tutoring", *Computers and Education*, vol. 44, pp. 53-68, 2005. <http://dx.doi.org/10.1016/j.compedu.2003.12.020>
- [36] Jones A., Scanlon E., Tosunoglu C., Morris E., Ross S., Butcher P., and Greenberg J., "Contexts for evaluating educational software", *Interacting with Computers*, 11(5), pp. 499-516, 1999. [http://dx.doi.org/10.1016/S0953-5438\(98\)00064-2](http://dx.doi.org/10.1016/S0953-5438(98)00064-2)

AUTHORS

C. Troussas is currently a Ph.D. Candidate in the Department of Informatics at the University of Piraeus in Greece. He received a M.Sc. degree in "Advanced Computing and Informatics Systems" (2010) and a B.Sc. degree in Informatics (2008) from the same Department. He has published over 25 articles in international conferences, books and journals. His current research interests are in the areas of user modeling, social networking services, artificial intelligence in education and adaptive systems.

M. Virvou is a Full Professor at the Department of Informatics, University of Piraeus in Greece. She received a Ph.D. Degree in Computer Science and Artificial Intelligence from the University of Sussex, UK (1993), a M.Sc. degree in Computer Science from the University College London, UK (1987) and a B.Sc. degree in Mathematics from the University of Athens, Greece (1986). She is Head of the Department of Informatics, University of Piraeus (Greece), Director of the Software Engineering Laboratory and also Director of postgraduate studies in the same Department. She has published over 300 articles in international conferences, books and journals. Her research interests are in the areas of user modeling, human computer interaction, knowledge-based software engineering, artificial intelligence in education, adaptive systems and affective computing.

Submitted 07 October 2013. Published as re-submitted by the authors 08 November 2013.