Some Approaches in Learning Style Diagnosis

Nikolaos Safouris, Sotirios Botsios and Dimitrios Georgiou

Department of Electrical & Computer Engineering, School of Engineering,
Democritus University of Thrace, GR 67100 Xanthi, Greece

Abstract

Adaptive Educational Hypermedia Systems among the emerging technologies serves Knowledge and Learning Systems. Aiming to improve pervasiveness and efficiency of asynchronous e-learning we give emphasis to the development of diagnostic tools capable to recognize certain learner's characteristics to the purpose of providing learning material tailored to the learner's specific needs. Learning Style diagnosis can be approached either by the use of probabilistic expert systems or by the use of fuzzy expert systems. In order to establish a probabilistic Learning Style's diagnostic expert system we propose a method based on Bayesian Networks. Our system analyzes information supplied by the system's descendant users, as well as providing a stochastic analysis of the problem. It is a system that utilizes what appear to be reasoning capabilities so as to reach Learning Style's estimations. In order to classify learners in a predefined set of classes, the proposed system takes under consideration learners' answers to a certain questionnaire, as well as a classification of learners who have been previously examined. As it was expected, our method takes advantage of previously accumulated knowledge and proves to be more accurate than Learning Style's direct estimation, i.e. an estimation based only on user responses. Further to this, we propose a Fuzzy Cognitive Maps technique, which modifies the Learning Style into appropriate quantitative characteristics. The importance of Fuzzy Cognitive Map's technique becomes even greater since this method takes into consideration the previously gained experience on learners' style diagnoses.

Keywords: Learning Style Estimation, Adaptive Educational Hypermedia Systems, Bayesian Networks, Fuzzy Cognitive Maps, Expert Systems, User Modelling, Probabilistic Models

1. Introduction

Learning Theories diverge with respect to the fact that students learn and acquire knowledge in many different ways, which have been classified as Learning Styles (LS). Learning behavior has been extensively examined in cognitive psychology. There exists a great variety of models and theories in the literature regarding learning behavior and cognitive characteristics i.e. LS or Cognitive Styles (CS) [Sternberg et. Al. (2001)]. Most of the authors categorize LS and/or CS into groups and propose certain inventories and methodologies capable of classifying learners accordingly [Smith (2001)].

In Web Based Education (WBE) researchers focus their attention on creating virtual tutors who behave like the best possible human tutors, without any of their potential limitations. Thanks to advanced technologies, it is now easier than ever to develop such virtual tutors, capable of teaching us how to learn and reason in the best possible way. Within the field of Adaptive Educational Hypermedia Systems (AEHS) we seek to combine the advantages of a computer with some of the tutors' intelligence characteristics for the purpose of making inferences, and taking decisions. Computers are fast, accurate, and have a reliable memory. The idea is to introduce human knowledge and experience in the computer in order to make it behave like the best possible tutor, who can adjusts his/her teaching to the learners' characteristics and abilities. This is expected to result in the optimum acquisition of knowledge. In AEHS learners are supported by virtual teachers who can be adapted to their LSs. LSs are considered relevant for the adaptation process in the user model, and have been used as a basis for adaptation in AEHS [Brown et. Al. (2005)]; [Karampiperis et. Al. (2005)]; [Georgiou et. Al. (2004)].

In order to develop AEHS capable of estimating LS, researchers face a wide range of relations which arise in complex dynamic systems. The existing relations (which might be unrecognizable) are necessarily poor approximations of complex dynamic systems, and some allowance must be made for uncertainty at this level of description. From a probabilistic point of view such hidden relations mean that there is a degree of uncertainty involved in the LS estimation.

In WEB, we look forward to improve the efficiency of the LS estimation, using expert systems. Development of such systems can be based on deterministic, probabilistic or even on fuzzy methodologies. In this paper we propose two expert systems, based on probabilistic and fuzzy analysis methods.

Aiming to establish a probabilistic expert system which is capable of estimating learner's LS, Botsios et. Al. propose an expert system based on Bayesian Networks (BN), the algorithm LSEvBN [Botsios et. Al. (2007)]. The system analyzes information from a questionnaire supplied by the system's descendant users, as well as providing a stochastic analysis of the problem. It is a system that utilizes what appear to be reasoning capabilities so as to reach LS estimations. In order to classify learners in a predefined set of classes, the proposed system takes under consideration learners' answers to a certain questionnaire, as well as a classification of learners who have been previously examined.

In terms of fuzzy analysis approach, Georgiou and Makry propose a model based on Fuzzy Cognitive Maps [Georgiou et. Al. (2004)]. A major advantage of Fuzzy Cognitive Maps (FCMs) is that they can handle even incomplete or conflicting information. This is considered to be of great importance because, quite often, pieces of information may be missing or be unreliable or difficult to integrate with information expressed differently. In this paper we shall present a FCM methodology for the recognition of learner’s cognitive characteristics. The importance of Fuzzy Cognitive Map’s technique becomes even greater since this method takes into consideration the previously gained experience on LS diagnoses.

2. LS Recognition

2.1 Direct Estimation

Cognitive scientists face a very challenging task in seeking to develop teaching strategies, in order to improve the way students learn. Therefore, they have created a number of inventories and questionnaires, resulted to an optimum estimation of learners’ LS. These inventories although having remarkable results, have limited efficiencies, due to:

- slippery answers and lucky guesses [Reye (2004)]
- cultural behavior [Dunn et. Al. (1990)]
- ‘grey’ areas in LSs discrimination
- time and effort consuming paperwork put on printed inventories (student, human supervisor)

2.2 Bayesian Networks

One of the primary roles of a Bayesian model is to allow the model creator to use commonsense and real-world knowledge to eliminate needless complexity in the model.

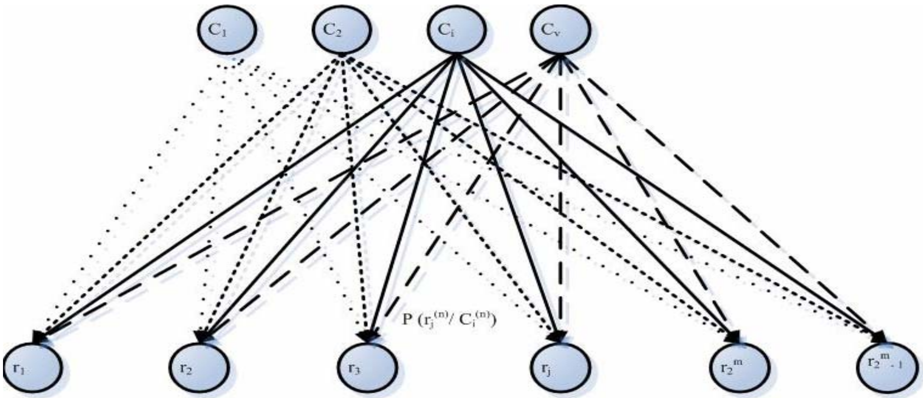

Figure 1. *The model of LSEvBN*

Let us consider the BN=(V, A, P) where $V=\pi_{\upsilon}\cup M$ and $\pi_{\upsilon}=LS=\{C_1, C_2, \ldots, C_v\}$ is the set of LSs. A learner is recognized as being of class C_i, $(i=1, 2, \ldots,v)$ according to his/her responses to a given set of m questions. Each question can be answered by yes or not. Let $Q=\{Q_1(k), Q_2(k), \ldots, Q_m(k)\}$ be the set of answers where k is a Boolean operator taking the values TRUE or FALSE whenever $Q_1(k)$ represents the answer YES or NOT respectively. There are 2^m different sets of such responses to the questionnaire. Let us consider the index j, where $j \in \{1, 2, \ldots, 2^m\}$. A learner's responses to the set of questions formulates an element

$$r_j = \bigcup_{l=1}^{m} Q_l^k \tag{1}$$

where $r_j \subset M$ the set of roots. Obviously, $r_i \neq r_j$ for any pair $(r_i, r_j) \subset M$, with $i \neq j$.

When our Bayesian model is actually used, the end user provides evidence about his or her LS through the responses to a given questionnaire. This information is applied to the model by "instantiating" or "clamping" a variable to a state that is consistent with the responses. Then the necessary mathematical mechanics are performed to update the probabilities of all the other LS and the array of responses variables that are connected to the variable representing the new evidence. Let n be the number of learners who made use of the system, and n_{r_i} be the number of those who responded to the questionnaire with r_i. The a priori probability that the $(n+1)^{th}$ user responded to the questionnaire with an element r_i is

$$P\left(r_i^{(n+1)}\right) = \frac{(n+1)_{r_i}}{n+1} \tag{2}$$

In this case, the LSEvBN in use is a weighted and oriented $K_{2^m}^{v}$ graph, i.e. a weighted and oriented complete bipartite graph on n and 2^m nodes. At each edge of the network's graph we adjust the conditional probability $P(r_i (n)/C_j (n))$. This probability expresses the ratio of users who responded to the questionnaire with the element r_i and were finally classified in C_j, in terms of the total number of r_i responses. Thus, the measure $P(C_j (n+1))$ is the probability that the LS of the $(n+1)^{th}$ learner belongs to C_j. This probability is given by the relation

$$P\left(C_j^{(n+1)}\right) = \sum_{i=1}^{2^m} P\left(C_j^{(n)} / r_i^{(n)}\right) P\left(r_i^{(n)}\right), \quad j = 1, 2, \ldots, v \tag{3}$$

where h an element r_i is

$$P\left(C_j^{(n)} / r_i^{(n)}\right) = \frac{P\left(r_i^{(n)} / C_j^{(n)}\right) P\left(C_j^{(n)}\right)}{\sum_{k=1}^{v} P\left(r_i^{(n)} / C_k^{(n)}\right) P\left(C_k^{(n)}\right)}, \quad \forall (i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, 2^m\} \tag{4}$$

where $P(Ci^{(n+1)}) = P(Cj^{(n+1)})$ for $i \neq j$, the learner can be classified either in class C_i or in class C_j. Due to this conflict with the procedure, and in order to avoid such a situation, the system redirects the programme flow to a subsystem where the whole procedure is repeated on a LSEvBN which has only includes the dominant classes C_i and C_j.

Finally, the learner's dominant LS is given through a combination of the LSEvBN and inventory response outcomes. If $C_j^{(n+1)}$ denotes the LS of the n+1 user, and '$score_j$' the final inventory's outcome score, then for

$$Value_j = P\left(C_j^{(n+1)}\right)(score_j) \quad \textbf{(5)}$$

The system's diagnosis therefore is the LS graded with the maximum value in (5). In the following section, the proposed model is applied using Kolb's LS Inventory [Kolb (1999)].

2.3 Fuzzy Cognitive Maps

Let us now consider a system of N nodes. Each node represents a concept C_i (I=1, 2, …, N) with a numeric value $V(C_i)$ which indicates the quantitative measure of the concept's presence in the systems model. Each two distinct nodes are joined by one, at the most, weighted arc. The arcs represent the causal relationships between adjoint concepts. The casuality degree from concept C_i to concept C_j is expressed by the value of the corresponding weight $w_{i,j}$. Experts describe this degree by using linguistic variables to express the weights. Weights vary from –1 to 1.

There are three types of causal relationships expressing the type of influence among the concepts, as represented by the weights $w_{i,j}$. Weights can be positive, negative or zero. Positive weight means the increasing influence a concept has to its adjacent concept of the graph. On the other hand, negative weight means that as concept C_i increases, concept C_j decreases on the $w_{i,j}$ ratio. In case of absence of relation between C_i and C_j, the weight $w_{i,j}$ equals zero.

FCMs converge either to a fixed point or limit cycle or a strange attractor [Dickerson et. Al. (1997)]. In case the FCM is called to support decision making process, as the recognition of LS, it is expected that FCM will result on a closed interval.

The procedure starts assigning a value to each concept. The values of the concepts change in the sequence as they are influenced by the adjacent concepts and their corresponding weights, according to formula (6).

$$V^{n+1}(C_i) = f(k_1 \sum_{\substack{j=1 \\ i \neq j}}^{k} w_{ji} V^n(C_j) + k_2 V^n(C_i)) \quad \textbf{(6)}$$

where $V^{n+1}(C_i)$ is the value of the concept C_i at the discrete time n+1, $0 \le k_1 \le 1, and\ 0 \le k_2 \le 1$ and f is a predefined threshold function. From now on the unipolar sigmoid will be used as *f*.

$$f(x) = \frac{1}{1+e^{-\lambda x}} \quad \textbf{(7)}$$

where λ determines the steepness of the sigmoid with λ>0.

The coefficient k_1 defines the concept's dependence on its inter-connected concepts, while the coefficient k_2 represents the proportion of contribution of the previous concept value in the computation of the next value. We selected $k_1=k_2=0.5$ as this results in smoother variation of the values of the concepts after each recalculation and more discrete final values.

Aiming to create an FCM based computer applicable algorithm Georgiou and Makry [Georgiou et. Al. (2004)] propose the following:

- Set the number k of learners.
- Set initial values n=0, $V_0(C_i)$ for i = 1, 2, …, N from the LS database. Data have been stored, while the learner responded to certain tests. Data have been stored as linguistic values Ai and have been turned into fuzzy degrees $V_0(C_i)$ for all concepts except those in LS. Concepts in LS are set equal to 0 for n = 0.
- Set the initial values for $w_{j,k}$ according to given information.
- For n=n+1, apply the relation (6) and set values $V_{n+1}(C_i)$. Update LS database. Following the defuzzification the weights at the edges of the graph are presented as elements of the adjacent matrix W_n.
- Set $\mathbf{V}^{n+1}=W_n\mathbf{V}^n$, where
- If a C_i does not be influenced by any C_j, j≠i then $w_{j,i}=0$ at present n.
- If a $V_n(C_i)$=m $[V_n(C_j)]$-1, for a given measure of competence 0<m<1, then set $w_{i,j}$= -m.
- Use the unipolar sigmoid function to transform the coordinates of V^{n+1} into the interval [0, 1].
- If $\underset{0 \prec I \prec k}{MAX}\left|V^{n+1}(C_i) - V^n(C_i)\right| \prec \varepsilon, (\varepsilon \succ 0)$ then stop and store as result the LS of the highest value $V^{n+1}(C_i)$.

2.4 Comparison of Methods

Due to the above numerous problems the direct estimation, has the worst efficiency among the methods. The probabilistic expert system based on Bayesian Networks and the fuzzy expert system, seem to have similar efficiency. Both systems can take into consideration the previously gained experience on LS diagnoses. Therefore, we can overcome any deficiencies caused by the dependence on direct use of the inventories. Initially, the Bayesian network provided suitable information for the FCM fine tuning, and at the sequel the systems collaborated to the purpose of errors correction. A joint use of the two systems in an educational platform seems to have rather surprising results, as it appears in a study case made on a test group of 885 university students. This study case is strong evidence to consider that using probabilistic and fuzzy methods in cooperation as a comparison of such LS estimations will add value to online LS estimation methods.

3. Applications

Kolb's learning theory sets out four distinct LS (or preferences), which are based on a four-stage learning cycle, which might also be interpreted as a ‘training cycle’.

According to David Kolb, diagnosis of LS can be based on the learner’s response to the inventory proposed by him [Kolb (1999)]. The learner, who responds to the inventory, marks the scores he or she has received on a separate scoring sheet. These scores are finally represented on a two-dimensional Cartesian plane giving a dominated vector located on a quadrant. In this paragraph based on Kolb's learning theory, we present an application for each one of the two dominant LS estimation methods.

3.1 Bayesian Networks

Let us consider that the set π_υ=LS= {CE, RO, AC, AE} has elements belonging to Kolb’s distinct LSs. The set M consists of all possible responses to the proposed inventory. We recall that each response is a 96-digits binary word. That is because there are 8 items with 6 pairs of statements each. When referring to any pair of statements, the user is forced to choose a single one. Therefore, card(M)=2^{48} indicates all the possible elements, i.e. the total number of different responses to Kolb’s originally proposed inventory. The online inventory we use, accepts a considerably smaller number of responses (card (M) $\leq 2^{32}$) due to the exclusion of incorrect answers. For the purpose of applying the proposed LSEvBN it is necessary to mark a priori conditional probabilities. To this end the LSEvBN is trained by a direct classification via Kolb’s reformed online inventory. It follows that the LSEvBN is a weighted and oriented $K^4_{2^{32}}$ graph having as weights at its edges the conditional probabilities $P(r_i(n)/C_j(n))$.

During the implementation of LSEvBN, the user database was still in process of refinement. At the starting point no data had been stored and so there was direct use of the revised online inventory. Both the users’ responses and the LS’s classes appointed to them were stored. If the system recognized a response which was still unknown, it borrowed marks from the direct application of the revised Kolb’s inventory. If there was a previously recorded response of the same kind, the LSEvBN used current values of a priori probabilities in order to classify the system’s user. At the same time, some of the LSEvBN’s weights were updated according to the evidence they gathered. Special attention was paid to avoiding an initial uniform joint distribution that would result in the system's inability to detect the user's LS. To this end, further direct classification via Kolb's inventory was made, and use of LSEvBN was omitted. The produced data enriched the system's database and modulated the conditional probabilities. Practically, such implications are not expected to occur after the initial system's training. In fact, the test group gave no such indications.

3.2 Fuzzy Cognitive Maps

There are two types of concepts: Learner Characteristics (LC) and Learning Activity Factors (LAF). The four central concepts LS=*{CE, RO, AC, AE}* are of the Learner Characteristic type. They are considered to be the outcome of the LS, Recognition Model. The concepts of the second type, that surround the four central concepts, are the LAFs. These are subjects measured by the system. Concepts of this type are, for example, risk taking ability, collaborative learning preference etc. Such factors influence directly the learner's characteristics. The oriented connections between concepts – vertices of the graph are represented by arrows. The connections may show the positive or negative influence that LAFs can have to LCs. Thus, weights, assigned to these connections, indicate the degree of influence. The influence degree can be negative or positive. Negative weight shows that while the value of the concept origin increases the value of the concept target decreases. The outputs of the FCM are the LCs. This is owed to the fact that it is desirable to maintain "competence" between output concepts. Such competence allows two at the most of the LCs be assigned by significally higher than the rest of the values. According to this point of view CE and AC are complementary concepts. This is also the case for the characteristics AE and RO. The desirable "competence" of the output concepts is not valid for the LSs as, for example, one can have both Diverger's and Assimilator's characteristics. So, we organize the central nodes in pairs: CE-AC and AE-RO, and we consider of interconnections, among the members of the same pair, at the lowest negative weight, which is –1. Connections between nodes of the two different pairs do not exist.

In order to formulate the algorithm we introduce the following notations:

- Θ the set of concepts, $C_i \in \Theta$, where $\Theta=\{LAF\} \cup \{LC\}$;
- **A**, a linguistic term of a linguistic variable (e.g. almost absolute cause);
- X is a measurable numerical assignment compact interval and $X \in (-\infty,\infty)$;
- **V**$\in$X, a linguistic variable assigned to $C_i \in \Theta$;
- $\mu_A(C_i)$, the membership value representing the degree of membership of $\theta_\mathbf{i}$ to the set of elements determined by the linguistic term **A**.

Since we do not expect that all LAFs have the same degree of effectiveness and causality on their adjacent LCs, weights must determined in order to express the degree of effectiveness and causality in the case. As cognitive psychology experts describe mostly qualitative behavior by using linguistic variables, it is necessary to introduce a transforming algorithm to map the values of such linguistic variables into membership functions. Watanabe's [Watanabe (1979)] membership functions direct estimation methods take an approach by asking experts to grade an event on a scale. Using such grading, we make use of the transformation which appears in Georopoulos et. Al. [Georopoulos et. Al. (2003)]. According to the suggested scheme, each fuzzy set corresponds to a membership function, where fuzzy sets describe the degree of causality corresponding to membership functions $\mu_A(C_i)$, A={w, o, s, vs}.The proposed fuzzy sets and their corresponding membership functions are:

- M_w(weak cause) the fuzzy set for causality about 17.5 % with membership function μ_{wc}.
- M_o(ordinary cause) the fuzzy set for causality about 42.5 % with membership function μ_{oc}.
- M_s(strong cause) the fuzzy set for causality about 57.5 % with membership function μ_{strc}.
- M_{vs}(very strong cause) the fuzzy set for causality about 82.5 % with membership function μ_{esc}.

Domains of membership functions are not of the same size since it is desirable to have a finer distinction between grades in the edges of the influence scale.

FCM have the ability to describe systems where there are feedback relationships (represented by the term $k_2 V^n(C_i)$ in formula 6) and relationships between concepts (represented by the sum $k_1 \sum_{i \neq j}^{N} V(C_j)$ in formula 6). The interrelation of the concepts and the feedback relation is described by the matrix W. Weights in feedback relation are the values of the diagonal matrix elements, while the rest of the matrix elements are the weights of concepts interrelations.

4. Conclusions and Future Work

Before concluding, it should be mentioned that it was in the authors' intentions to develop an applicable on line LS estimation method based on prevalent learning theories. Also one of our targets is to test our algorithms in various different inventories.

Let us recall, that the effectiveness of the model depends on the cognitive characteristics of the test group under examination. In this study, we collected evidence that test groups taken from populations of various origins formulate LSs distributions of different types, and we intend to further investigate such differences.

It is our intention to create an online platform where tutors and students can interact with in order to receive comparative estimations for their personal learning style.

Following this, efforts will be made to extend the use of our platform via other mathetical tools, i.e. binary trees etc.

References

Botsios, S., Georgiou, D., Safouris, N. (2007), *Learning Style Estimation Using Bayesian Networks*. 3rd International Conference on Web Information Systems and Technologies, Barcelona, Spain.

Brown, E., Cristea, A., Stewart, C., Brailsford, T. (2005), *Patterns in Authoring of Adaptive Educational Hypermedia: A Taxonomy of Learning Styles*. Educational Technology & Society 8(3), 77-90.

Dickerson, J.A., Kosko, B. (1997), *Virtual Worlds in Fuzzy Cognitive Maps*, Fuzzy Engineering, editor B. Kosko, Prentice-Hall, Upper Saddle River, New Jersey.

Dunn, R., Gemake, J., Jalali, F., Zenhausern, R. (1990), *Cross-cultural differences in learning styles of elementary-age students from four ethnic backgrounds*. Journal of multicultural counselling and development 18, 68-93.

Georgiou, D., Makry, D. (2004), *A Learner's Style and Profile Recognition via Fuzzy Cognitive Map*. The 4th IEEE International Conference on Advanced Learning Technologies, Joensuu, pp. 36-40.

Georgopoulos V., Malandraki G., Stylios, Ch. (2003), *A Fuzzy Cognitive Map Approach to Differential Diagnosis of Specific Language Impairment*. Artificial Intelligence in Medicine, 29, 261-278.

Karampiperis, P., Sampson D. (2005), *Adaptive Learning Resources Sequencing in Educational Hypermedia Systems*. Educational Technology & Society 8(4), 128-147.

Kolb, D. (1999), *Learning Style Inventory – version 3: Technical Specifications*. TRG Hay/McBer, Training Resources Group.

Reye, J. (2004), *Student Modelling based on Belief Networks*. International Journal of Artificial Intelligence in Education 14, 63-96.

Smith, S.E. (2001), *The relationship between learning style and cognitive style*. Personality and Individual Differences 30(8), 609-616.

Sternberg, R.J., Zhang L.F. (2001), *Perspectives on Thinking, Learning, and Cognitive Style*. International Review of Education, 48 Netherlands: Springer.

Watanabe, N. (1979), *Statistical Methods for Estimating Membership Functions*. Japanese Journal of Fuzzy Theory and Systems, 5(4).