Hidden Markov Model Based Characterization of Content Access Patterns in an e-Learning Environment

Apple W P Fok, H S Wong and Y S Chen
AIM*tech* Centre and Department of Computer Science
City University of Hong Kong

Abstract: Personalized Education (PE) emphasizes the importance of individual differences in learning. To deliver personalized e-learning services and content, PE encompasses the abilities of identifying and understanding individual learner's needs and competence so as to deploy appropriate learning pedagogy and content to enhance learning. In this paper, we introduce a Hidden Markov Model Based Classification approach to enable a multimedia e-learning system to characterize different types of users through their navigation or content access patterns. Our experiments show that the proposed approach is capable of assigning student users to their corresponding categories with high accuracies. The results of such classifications would find applications in adaptive user interface design, user profiling and as supportive tools in personalized e-learning.

1. Introduction

Worldwide Education Reform raises the importance of Information and Communication Technology (ICT) in Education. To fully utilize the huge amount of useful and usable hypermedia that can potentially be accessed from the Internet, we extend the concept of Personalization in e-commerce to education, namely "Personalized Education" (PE). PE encompasses the identification and understanding of individual learner's needs and competence and then adopts and deploys appropriate learning pedagogy and multimedia content to bridge the learner's knowledge gap. In order to deploy the appropriate "Personalization Strategies" to serve the learning need of different learners in a different way, an agent-based framework for the development of a Personalized Education System (PES) has been proposed [1]. PES composes of a number of functional units that exploit personalization technologies extensively to support the various learning activities in a personalized manner. This includes intelligent user profiling and multimedia content searching and clustering as well as an intelligent man-machine interface that adapts dynamically to individual user's behavior and interactions with the system. Early in development, through positive and negative cultural and educational experiences, individuals develop a set of coherent interests, personality, and abilities [2]. This leads to the belief that PE is a high-impact environment for the next generation supportive tools for e-learning.

Empirical studies support the argument that web-based learning environments can significantly assist dissemination of knowledge within modern educational settings, however, little is known about the content access patterns or navigation patterns of students in a web-based multimedia learning environment.

The personalized user-interface, an elementary component of PES, not only gives a pleasant, enthusiastic and positive greeting when user starts to interact with PES, more importantly, PES incorporates the "learning while doing" strategy in the interface design; a collection of specific features for a particular user/user-group. Personalized interface offers guidance to individual user at each step of the user's interaction, supports user to use his/her natural intelligence to make choices. Hence, user modeling is essential in the process of personalized services delivery because it is the only possible way that a system would response more accurately to the identified user group with the appropriate learning activities/instructions.

To realize an intelligent user interface that is able to dynamically adapt itself to a user through his/her interactive behavior with the system. In this paper, we introduce a formal approach based on the Hidden Markov Model (HMM) [3] to learn and to characterize user behavior from his/her interactions with a web-based learning system. The resulting HMM trained from some scenarios of user interactions with the system are used to instantiate and identify user profile which help the system to predict and anticipate user needs and to dynamically adjust the user interface to suit the user's competence and learning needs.

2. Related work

Hidden Markov model (HMM) is a well-known approach to probabilistic sequence modeling and has been extensively applied to problems in speech recognition, motion analysis and shape classification [e.g. 3-4]. The Viterbi algorithm has been the most popular method for predicting optimal state sequence and its associated maximum posterior probability value. The Baum-Welch re-estimation, a variant of the well-known EM algorithm, has been the predominant model estimation technique given a HMM topology.

In HMM based learning, initialization is a crucial step due to the local behavior of the standard procedure used to estimate the HMM parameters. Training with

0-7803-9332-5/05/$20.00 ©2005 IEEE

inappropriate initializations may be trapped in a local minimum and should be avoided whenever possible. Another fundamental issue is the determination of the structure of the HMM. The choice of a good model structure is critical to the effectiveness of the learning process [4].

The problem of characterizing a user's behavior on a web-site has gained popularity due to the rapid growth of the need to personalize and influence users' browsing experience. Markov models have been found well suited for addressing such problems. In general, the input for these problems is a sequence of some web pages accessed by a user and the objective is to build Markov models that can characterize the user's behavior [5-6]. However, in most relevant work the web users' behavior is modeled by a simple Markov chain where the individual states are directly associated with particular navigation links on web pages. In other words, there are no hidden states for representing different internal cognitive behaviors of users. Here, we propose to remedy the above drawback by taking into account the hidden cognitive contexts behind users' observable actions.

This paper aims at exploiting the capabilities of hidden Markov models for smart user characterization in web based learning systems. Our objective is to investigate whether the HMM approach can successfully distinguish between two specific student groups based solely on their navigation or content access patterns in a web-based learning system. Specifically, we address the issues of model selection and initialization, which are critical to the success of relevant applications but have been frequently ignored in related literature.

3. HMM modelling of content access patterns in e-learning

3.1 A brief introduction to hidden Markov models

Basically, A discrete HMM λ is formulated by a 3-tuple $\lambda = \{A, B, \pi\}$, where A is the transition matrix, which defines the transition probabilities between N hidden states, B is the emission matrix, which defines the probabilities of M discrete observations under each state; and π is a vector of initial state probabilities.

A typical HMM-based classifier adopts the maximum-likelihood criterion, where an unknown sequence O is assigned to the class showing the highest likelihood, i.e.

$$Class(O) = \arg\max_{i} P(O \mid \lambda_i) \tag{1}$$

where λ_i is the HMM corresponding to the i-th class. This requires training C HMMs for a C-class problem. The probability term $P(O \mid \lambda_i)$ in equation (1) can be computed with the well-known forward-backward procedure.

3.2 Overall framework

In this work, we focus on automatically learning user models, characterizing and classifying user groups in a personalized education system. First, we want to know whether there exist some learning patterns for students belonging to different groups that can be reflected in their educational content access activities and characterized by a carefully designed model. Second, we attempt to effectively generate the model from limited training data and then employ it on future unseen data to classify or predict the behavior of new students.

Our basic assumption is that different classes of users will have different preferences of the various learning activities in a web-based learning system, which translates into a characteristic navigation or content access sequence when members of a particular user class interact with the system. Given this assumption, our goal is to design a user model for each class that can correctly characterize the navigation sequence for the corresponding class members. The example navigation sequences are used to estimate the optimal parameters for a particular user class model.

HMM is a natural choice to the above questions since it is ideally suited for modeling the stochastic aspect of discrete temporal sequences like web content access actions. The non-stationary aspect of the navigation behavior can be better modeled with multiple hidden states in contrast to a single probability distribution. Moreover, individual user classes can be characterized by their corresponding HMM, and new user categories can easily be included by adding new HMM to the original ensemble.

3.3 Training and testing

Given a set of example navigation sequences U^n from the *n*-th user class, the Baum-Welch algorithm is used to iteratively adjust the parameters λ^n such that the likelihood of the sequences is maximized. In other words, the set of optimal parameters λ^n should satisfy:

$$\lambda^{n*} = \arg\max_{\lambda^n} P(U^n \mid \lambda^n) \tag{2}$$

where $P(U^n \mid \lambda^n)$ is the likelihood function. This procedure is repeated for each of the user classes to determine their respective sets of parameters. In this manner, different categories of user behaviors can be summarized in terms of a set of transition probabilities between different internal user states, and the probability of choosing a particular navigation link in each of these states. To perform user classification, a new user is first requested to interact with the web-based interface briefly such that a navigation sequence U can be obtained. The forward-backward procedure can then be applied to evaluate the likelihood of this sequence with respect to each of the HMMs and then assign the user to the class associated with the HMM with the maximum likelihood.

3.4 Model selection

A fundamental issue in this approach is to determine the topology and the number of hidden states of the HMM. Individual navigation links or buttons on the web-based interface naturally act as the observation sequences after necessary preprocessing. We additionally suppose that there are several abstract psychological states hidden behind these observations and dominate the generation of the observation sequences with a statistical model. Since there is no strong evidence showing temporal order between different states, we adopt a generic ergodic hidden Markov model to exploit full relationships among all hidden states.

Since the hidden state is an abstract concept in our modeling context no intuitive clues are easily found to guide the determination of the number of states. A typical principle in treating this problem is to choose a suitable tradeoff between model flexibility and fitting performance. That is, we choose a score function that encapsulates the compromise to evaluate each possible model structure. The Bayesian Information Criterion (BIC) can be applied to assist us in making decisions based on the aforementioned principle. We adopt the following score function in model selection [7].

$$S_{BIC}(\lambda_k) = 2S_L(\lambda_k) + d_k \log n \qquad (3)$$

where S_L is a negative log-likelihood function measuring the goodness of fit of the model described by λ_k, d_k is the number of free parameters in λ_k, and n is the sample size taken into account. Broadly speaking, in this formulation the error function is penalized by an additional term to ensure a reasonable balance between goodness of fit and model complexity. Based on this score function, we try to achieve optimization over both model structures and parameter estimations by a nested searching algorithm. This strategy can guarantee to find reliable solutions with moderate computational cost in the presence of moderate training data.

3.5 Initialization

Another important but often overlooked issue is the initialization of the training procedure. Note that our major target is to optimize equation (2). In fact, initialization is crucial to such optimization procedures. Inappropriate starting point often leads to local minimum and stops at a suboptimal solution. Therefore, the solution is heavily dependent on the starting point of the searching algorithm. To reduce the risk of being trapped in local minimum various methods have been developed, mostly at the expense of higher computational complexity. Fortunately For many practical learning applications there is an attractive scheme. That is, the optimization is repeated several times with multiple randomly chosen starting points and the best result produced is adopted as the final solution. Although this stochastic search technique involves extra computation and overhead, it is quite suitable for our learning process due to the relatively simple model structures adopted. This approach ensures the reliability and robustness of the result.

Specifically, in our learning algorithm we have two matrices to initialize, i.e. the transition matrix A and the emission matrix B. Initializing both matrices randomly will significantly increase the computational burden and is not recommended. Experiments show that for general training algorithms the transition matrix is more sensitive to initialization than the emission matrix. That is, if the transition matrix is inappropriately initialized, the algorithm bears a high risk of getting stuck in local minimum. On the contrary, if the transition matrix is appropriately initialized but the emission matrix is not initialized to be close to the ground truth, the training algorithm can in general adaptively adjust the emission matrix gradually to the true values during training. This observation leads to the following initialization strategy: For each new searching step, we only randomly initialize the transition matrix A and set each row of the emission matrix B to a default unique distribution. This initialization makes the training procedure somewhat simpler and more effective in our application.

4. Experimental results

The experimental data are collected through a Web-based English Learning System (WELS) for primary school students [8]. The underlying assumption is that different groups of users will have different preferences for the various learning activities and content available in WELS which form the discrete HMM observation sequences in our experiment.

With the support of our collaborating school, 187 participant records are collected during a period of two weeks. The students are ranked according to their English performance in the past year, and based on this ranking, a top ranking group (Group A) and a low ranking group (Group B) of students are chosen respectively. In each group, a number of students are selected at random and their navigation sequences are used as the training data set while the others form the test data set.

For each group of students under study, a corresponding HMM is constructed to model their navigation behavior, which we denote as HMM A and HMM B respectively. Random transition matrix initialization is employed to reach global optimal solution. The number of hidden states is selected by BIC using equation (3). Table 1 gives the normalized score function values under several different state numbers, which shows that the 4-state topology leads to minimal score function and is thus the best choice for the model. This leads to an optimal topology with a 4x4 transition matrix and a 4x7 emission matrix.

Table 1. BIC score functions for different state selections

No. of states	2	3	4	5
$S_{BIC}(A)$	1.58	1.34	**1.23**	1.28
$S_{BIC}(B)$	1.41	1.27	**1.25**	1.28

Once the parameters of the two HMMs have been determined based on the training sequences, their classification performances are evaluated by applying these two models to navigation and content access sequences which are not included in the training set. For each group of test data, we test sequences with different lengths. The test results are shown in Figure 1, with longer sequences corresponding to larger sample index.

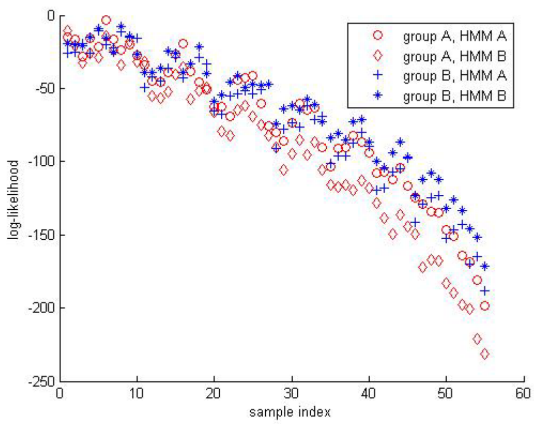

Figure 1. 2-category HMM classification result

Figure 1 shows intuitively that for most samples the "*" marks are associated with higher log-likelihood values than the corresponding "+" marker. Similar conclusion can be drawn with respect to the circle marks and the diamond marks. This means that the HMM can effectively distinguish between students of different categories. The global classification accuracy is 92.73% for group A and 87.27% for group B test data. Another interesting observation is that the results of the two HMMs for a single test sample are more separable for longer sequences and the accuracy increases with the increase of the length of the test sequence. When the length of the sequence exceeds 60 the accuracy reaches 100%. This shows that the model has some error resilient function capable of correcting early mistakes. The result is expected because the Viterbi algorithm employed here is itself a powerful error correction algorithm in the area of channel coding.

After the HMMs are built it is possible to do more quantitative evaluations about the models. For instance, the equilibrium state probabilities associated with both HMMs can be computed, as shown in Equation (4). The equilibrium probabilities reveal that group A bears a higher entropy compared with group B. This observation may be used to guide user activity prediction or interface adaptation, which can lead to very interesting future work.

$$\begin{aligned} \Pi_A &= [0.291 \quad 0.192 \quad 0.228 \quad 0.289] \\ \Pi_B &= [0.243 \quad 0.283 \quad 0.056 \quad 0.418] \end{aligned} \tag{4}$$

5. Conclusion

This paper exploits the usage of HMM classification in web-based education systems. BIC is employed to select proper model topology and random initialization is used to guarantee global optimization. Our experiments show that the proposed approach is capable of assigning new student users to their corresponding categories with high accuracies. To our knowledge, this is the first work that applies HMM to characterize student groups for the purpose of personalized education based solely on the sequence of navigation or content access actions within an e-learning environment.

Acknowledgement

This work is jointly supported by HKSAR Quality Education Fund Project No. 2003/0350, Sun Centre of Excellence on Pervasive Computing and CityU SRG Project No. 7001698.

References

[1] Apple W. P. Fok and Horace H. S. Ip, Personalized Education (PE) – Technology Integration for Individual Learning, Proc. of 3rd International Conference on Web-Based Education, pp.48-53, Austria, 2004.

[2] Phillip L. Ackerman, Traits and Knowledge as Determinants of Learning and Individual Differences: Putting It All Together, Learning and Individual Differences, Process, Trait, And Content Determinants, the American Psychological Association, P.437-460, 1999.

[3] L. Rabiner, A tutorial on Hidden Markov Models and selected applications in speech recognition, Proc. IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[4] M. Bicego, V. Murino, Investigating Hidden Markov Model's capabilities in 2D shape classification, IEEE Trans. PAMI, vol. 26, no. 2, pp. 281-286, 2004.

[5] R. Sarukkai, Link prediction and path analysis using Markov chains, Computer Networks 33: 377-386, 2000.

[6] M. Deshpande and G. Karypis, Selective Markov models for predicting web page accesses, ACM Trans. on Internet Technology 4(2): 163-184, 2004

[7] D. Hand et al., Principles of data mining, Cambridge, MIT Press, 2001.

[8] Apple W P Fok and Horace H S Ip, WELS -- A Supportive Teaching and Learning Web-based Environment for Primary Schools, Proc. of GCCCE 2004, pp 326-329, 2004.