INFORMATION RETRIEVAL USING MARKOV MODEL MEDIATORS IN MULTIMEDIA DATABASE SYSTEMS

Mei-Ling Shyu, Shu-Ching Chen, and R. L. Kashyap

School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN 47907-1285
{shyu, shuching, kashyap}@ecn.purdue.edu

ABSTRACT

Recent progress in high-speed communication networks, large capacity storage devices, digitalized media, and data compression technologies have resulted in a variety of multimedia applications using the integration of text, images, audio, graphics, animation, and full-motion video. For traditional text-based database management systems, data access and manipulation have advanced considerably. However, for multimedia database systems, information retrieval is more difficult than that of the conventional data since it is necessary to incorporate diverse media with diverse characteristics. The need for information retrieval in multimedia database systems increases proportional to the continuous growth of diverse information sources and the proliferation of independent but related user applications. Therefore, the ability to query the databases and to locate specific information directly as needed is important for multimedia database systems. For this purpose, a mathematical sound framework, called *Markov model mediators (MMMs)* which employ the principle of Markov models and the concept of mediators, is introduced in this paper. The proposed MMM mechanism performs information retrieval via a stochastic process which generates a list of possible state sequences with respect to a given query and indicates which particular media objects to query.

1 Introduction

Recently, much attention has been focused on multimedia information technologies and applications. Information of all sorts (video, audio, pictures, text, and data) in varied formats are highly volatile. Multimedia information has been used in several applications including manufacturing, medicine, education, business, entertainment, etc. For example, the merchandise of a manufacturer can be advertised by providing audio descriptions, video demonstrations, and prices in textual format. With the increasing use of multimedia database systems and the fact that information retrieval in multimedia database systems is more difficult than the conventional database systems, there is the need for a *multimedia database management system (MDBMS)* which has the capabilities to provide a suitable environment for storing, retrieving, and managing the data in multimedia systems.

Recent papers related to multimedia database systems can be categorized in the following application domains: speech recognition, word recognition, signal processing, handwriting recognition, and document/passage retrieval [1] [6] [7] [8] [9]. However, the focus of the above researches is on the low-level feature recognition of multimedia data; while our approach addresses the need for a mechanism at the database management point of view. Toward this end, we have proposed a unified model that allows us to query different media types and manage the rich semantic multimedia data by using a mathematical structure, called a *Markov model mediator (MMM)*. Since the primitive constructed or manipulated entities in most multimedia systems are called *media objects* which could be a video clip, an image, a text file, or a complex entity of these simpler entities [2], a media object is represented as a node in an MMM and is associated with an ATN. An ATN is a model for multimedia presentations, multimedia database searching, and multimedia browsing [3] [4] [5].

Since the MMMs possess the stochastic property of Markov models, the processes of locating the required media objects for a query are based on complex statistical and probabilistic analyses which are best understood by examing the network-like structure in which those statistics are stored. Hence, the proposed MMM mechanism plays as an MDBMS by two stochastic processes. The first stochastic process discovers the summarized knowledge to construct a federation of data warehouses [10]. A second stochastic process generates a list of possible state sequences with respect to a given query and indicates which particular media objects to query over the constructed data warehouses. When the required media objects are predicted, the corresponding ATNs are traversed for information retrieval. Moreover, since there might be multiple data warehouses constructed in the first stochastic process, if one integrated MMM could not provide all the information for a query, then the second stochastic process is applied to other integrated MMMs until all the information for the query is found.

The rest of the paper is organized as follows. The components of an MMM (local or integrated) are introduced in Section 2. The proposed approach for information retrieval using integrated MMMs is presented in Section 3. In Section 4, a query example to illustrate how our information retrieval approach based on the proposed MMM mechanism works is given. The conclusions are drawn in Section 5.

2 The Markov Model Mediators (MMMs)

There are two types of MMMs - local MMMs and integrated MMMs. Each multimedia database is modeled as a local MMM and each data warehouse is modeled as an integrated MMM. An MMM (local or integrated) is represented by a 6-tuple $\lambda = (\mathcal{S}, \mathcal{F}, \mathcal{A}, \mathcal{B}, \Pi, \Psi)$ where $\mathcal{S}$ is a set of media objects called states; $\mathcal{F}$ is a set of attributes/features; $\mathcal{A}$ is the state transition probability distribution; $\mathcal{B}$ is the observation symbol probability distribution; Π is the initial state probability distribution; and Ψ is a set of augmented transition networks (ATNs).

An MMM consists of a sequence of states which represent the media objects (in $\mathcal{S}$) in the multimedia databases. The states are connected by directed arcs (transitions) which contain probabilistic and other data used to determine which state should be selected next. All transitions $S_i \rightarrow S_j$ such that $Pr(S_j \mid S_i) > 0$ are said to be allowed, the rest are prohibited. $Pr(S_j \mid S_i)$ is greater than 0 when the media objects S_i and S_j have been accessed together by a set of historical queries or have structural equivalence relationship. Also, since different media objects may have different types of attributes or features, each media object has its own set of attributes/features (in $\mathcal{F}$). $\mathcal{A}, \mathcal{B}$, and Π are the probability distributions for an MMM and play as the major roles in the stochastic processes. The elements in $\mathcal{S}$ and $\mathcal{F}$ determine the dimensions of $\mathcal{A}$ and $\mathcal{B}$. The formulations of $\mathcal{A}, \mathcal{B}$, and Π for an MMM and the construction of the data warehouses are shown in [10]. Since those local MMMs which are accessed frequently are placed in the same data warehouse, the integrated MMMs are used in the second stochastic process to find the possible list of state sequence for a query.

The augmented transition network (ATN) is a semantic model to model multimedia presentations, multimedia database searching, and multimedia browsing. The arcs in an ATN represent the time flow from one state node to another. An arc represents an allowable transition from the node at its tail to the node at its head, and the labeled arc represents the transition function. An input string is accepted by an ATN if there is a path of transitions which corresponds to the sequence of symbols in the string and which leads from a specified initial state to one of a set of specified final states. In addition, subnetworks are developed to allow the users to choose the scenarios relative to the spatio-temporal relations of the video or image contents or to specify the criteria based on a keyword or a combination of keywords in the queries. Information in text databases can be accessed by keywords via the text subnetworks. The inputs for ATNs are modeled by multimedia input strings. Also, each subnetwork has its own multimedia input string. Database searching in ATNs is performed via substring matching between the multimedia input string(s) of the ATN (and its subnetworks) and the multimedia input string of a given query. For example, if a text subnetwork contains the keyword "Purdue University Library", then the Purdue University library database is linked via a query with this keyword. Therefore, each media object has an associated ATN and when the required media objects are predicted, the corresponding ATNs are traversed for information retrieval. For the details of ATNs, please see [3] [4] [5].

3 Information Retrieval Using the MMM Mechanism

Under multimedia database systems, the need for efficient information retrieval is strong because searching databases one by one is very time-consuming and expensive. The cost for query processing usually is very high and the complexity of a query depends heavily on the order in which the network is searched for a successful path. To speed up query processing, an efficient way to identify a successful path or to locate information for a query is very crucial. The MMM mechanism and the stochastic processes are proposed for this purpose. The integrated MMMs are the units for database searching and information retrieval. A lattice (or trellis) structure which yields a list of possible state sequences for a specific query with a given integrated MMM is first created. Then, we use *dynamic programming* on the lattice for the possible paths.

Consider one fixed state sequence $\mathcal{S}=\{S_1, S_2, \ldots, S_N\}$ for a given observation set $O = \{o_1, o_2, \ldots, o_T\} \subseteq \mathcal{F}$, where S_i denotes a state (media object), o_j represents an attribute/feature, N is the number of states, and T is the number of attributes/features required in a query. Define $W_t(i,j)$ to be the edge cost of the edge $S_i \rightarrow S_j$ at time t and $D_t(j)$ to be the cumulative node cost of the node S_j at time t, where $1 \leq i,j \leq N$, $1 \leq t \leq T-1$.

$$W_1(i,j) = \begin{cases} \pi_{S_i} b_{S_i}(o_1) & \text{i=j} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$D_1(j) = \max_i W_1(i,j) = W_1(j,j) \tag{2}$$

$$W_{t+1}(i,j) = D_t(i) a_{S_i,S_j} b_{S_j}(o_{t+1}). \tag{3}$$

$$D_{t+1}(j) = \max_i (D_t(i) + W_{t+1}(i,j)). \tag{4}$$

Here, $\mathcal{A}$, $\mathcal{B}$, and Π denote the state transition probability distribution, the observation symbol probability distribution, and the initial state probability distribution for an integrated MMM, respectively.

$\mathcal{A} = \{a_{S_i,S_j}\}$, where a_{S_i,S_j}=Pr(S_j at t+1 | S_i at t).
$\mathcal{B} = \{b_{S_j}(o_k)\}$, where $b_{S_j}(o_k)$=Pr(o_k at t | S_j at t).
$\Pi = \{\pi_{S_i}\}$, where π_{S_i}=Pr(S_i at t=1).

At time t=1, $W_1(i,j)$ is assigned the value of the joint probability of the state S_i with probability π_{S_i} and the attribute/feature o_1 with probability $b_{S_i}(o_1)$ when $i=j$; $W_1(i,j) = 0$ if $i \neq j$. Since

$D_t(j)$ is the cumulative node cost of the node S_j at time t, $D_1(j)$ is assigned to the value of $W_1(j,j)$ when $t=1$ given that all $W_1(i,j)$=0 if $i \neq j$. Then, a transition goes from state S_i to state S_j with probability a_{S_i,S_j} and the attribute/feature o_2 is generated with probability $b_{S_j}(o_2)$ as time goes from t=1 to t=2.

The lattice is generated in the following way. N nodes are created at the beginning (i.e. t=0) and for each time slot $t > 0$ with each node representing a state of the lattice. A node S_i at time $t-1$ connects to all the N nodes at time t with the edge cost $W_t(i,j)$ and the cumulative node cost $D_t(j)$, where $1 \leq i,j \leq N$ and $1 \leq t \leq T$. The process continues until all the attributes/features in the observation set are generated at time $t = T$. The list of possible state sequences are ranked based on the values of $W_t(i,j)$ and to suggest the paths to retrieve information for the query. Only those paths which have positive edge costs from $t=0$ to $t=T$ are considered and they are ranked in the following manner. First, the $D_T(j)$ values where $1 \leq j \leq N$ are sorted. Then, the top ranked path is the one which constitutes maximal $D_T(j)$ value, the second ranked path is the one with the second ranked $D_T(j)$ value, and so on. In other words, the top ranked path is a state sequence that provides the information for the query with maximal cumulative node cost. If the top ranked path cannot provide the information required for the query, then the second ranked path is considered. This is repeated until the information needed by the query can be obtained. From our experience, most of the probabilities are zeros at each time slot. Hence, the path ranking procedure becomes easy since there will not be many paths with all positive edge costs along the path in the lattice. The information retrieval method is in effect based on the lattice structure constructed for a specific query since there are only N states at each time slot $t > 0$ in the lattice. No matter how many attributes an observation set has, all the possible state sequences will be merged into these N nodes. Moreover, the information retrieval methods shows physical interpretation in trellis and is computationally cheaper.

4 An Example

We use the following query example to explain how our MMM mechanism is used to find a list of possible state sequences for retrieving information. Figure 1(a) is the integrated MMM used for this query. There are six states (media objects) in the integrated MMM and each state has an associated ATN with it. For simplicity, only the ATN for S_1 is shown (see Figure 1(b)). Since S_1 contains video frames and texts, two subnetworks are created for them (see Figure 1(c) and 1(d)).

Query: Find the video clips of manufacturer A's advertisement beginning with a salesman named M who is alone and ending with M holding an InletNeedle product with diameter equals 0.25.

To retrieve information for this query, several steps need to be executed. First, the query is translated into a multimedia input string. Second, since the attributes/features specified in this query are the employee name (M), the manufacturer (or company) name (A) and the diameter (0.25) for the InletNeedle, the observation set is specified as $O = \{emp\_name, mname, diameter\}$. Third, construct the lattice and compute the $W_t(i,j)$ and $D_t(j)$ values for the lattice, where $1 \leq t \leq 3$ and $1 \leq i,j \leq 6$. Figure 2 shows the lattice for the query. In Figure 2, the positive edge costs are shown in bold lines and others have zero edge costs. Next, the paths with positive edge costs are ranked. It can be clearly seen that only one possible path exists in Figure 2 and so the top ranked path can be determined as $S_1 \rightarrow S_6 \rightarrow S_5$. Once the required media objects are identified, information retrieval becomes handy. Simply traverse the ATNs of those media objects. If the media object contains image, video frames, or texts, then the corresponding subnetworks are traversed, too. In this example, since it asks for the advertisement video clips of the manufacturer (or company) A with respect to an employee M, some video and text elements are involved in the searching. Hence, V_1 and T_1 in the ATN of S_1 should be used. The multimedia input string for the query is translated to be (M)(M&P) and the multimedia input string of the subnetwork for V_1 is (M)(M&B)(M&P). Substring matching is conducted between these two multimedia input strings. Information in text databases are accessed by keywords via the text subnetworks. For example, the keyword "emp_name" in T_1 is used to search for the information for the employee named M. The details of the multimedia input strings and the substring matching procedures are discussed in [5].

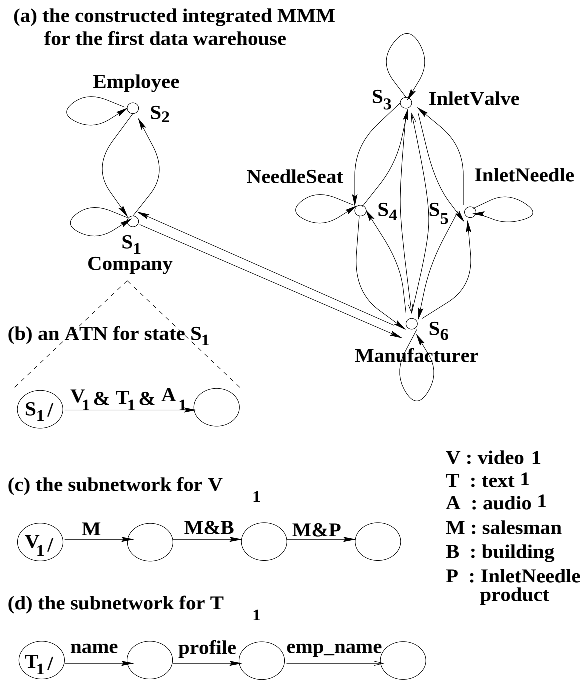

Figure 1: An example of an integrated MMM – each state in turn is represented by an ATN. (a) is the example integrated MMM. (b) is the ATN for state S_1. (c) and (d) are the subnetworks for V_1 and T_1, respectively.

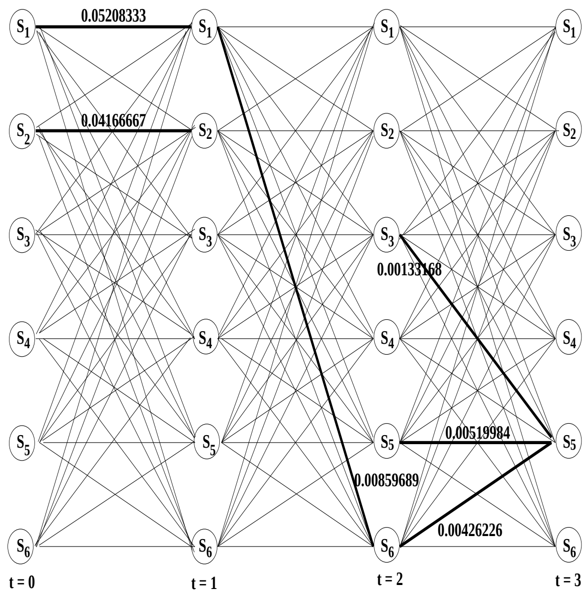

Figure 2: The lattice structure created for the query. The bold lines list all the positive edge costs $W_t(i,j)$. All the other edge costs are zeros.

5 Conclusions

The rapid growth of multimedia applications has increased the need for the development of MDBMSs as a tool for efficiently storing, retrieving, and managing the information in multimedia database systems. A new approach to retrieving information from multimedia database systems based on MMM mechanism to facilitate MDBMSs is proposed in this paper. The analysis as well as the example presented in Sections 3 and 4 show our information retrieval method is feasible since information retrieval can be performed in terms of probabilistic retrieval. Moreover, under our proposed information retrieval method, the media objects for a specific query can be identified efficiently. Therefore, the time for query processing can be reduced in the multimedia database systems.

Acknowledgements

This work has been partially supported by a grant from Allied Signal Corporation and National Science Foundation under contract IRI 9619812.

References

[1] E. J. Bellegrada et al., "Discrete Parameter HMM Approach to On-Line Handwriting Recognition," Proceedings of the 1995 20th International Conference on Acoustics, Speech, and Signal Processing. Part 5 (of 5), Detroit, MI, USA, May 1995, pp. 2631-2634.

[2] K.S. Candan, P.V. Rangan, and V.S. Subrahmanian, "Collaborative multimedia systems: synthesis of media objects," IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No. 3, pp. 433-457, May/June 1998.

[3] Shu-Ching Chen and R.L. Kashyap, "Temporal and spatial semantic models for multimedia presentations," 1997 International Symposium on Multimedia Information Processing, pp. 441-446, Dec. 11-13, 1997.

[4] Shu-Ching Chen and R.L. Kashyap, "Empirical studies of multimedia semantic models for multimedia presentations," 13th International Conference on Computer and Their Applications, Honolulu, Hawaii USA, pp. 226-229, March 25-27, 1998.

[5] Shu-Ching Chen and R.L. Kashyap, "A spatio-temporal semantic model for multimedia presentations and multimedia database systems," Submitted to IEEE Transactions on Knowledge and Data Engineering.

[6] H.C. Lin, L.L. Wang, and S.N. Yang, "Color image retrieval based on hidden Markov models," IEEE International Conference on Image Processing, Vol. 1, Washington DC., Oct. 23-26, 1995.

[7] E. Mittendorf and P. Schäuble, "Document and passage retrieval based on Hidden Markov Models," In ACM-SIGIR Conference on Research and Development in Information Retrieval, pp. 318-327, 1994.

[8] YK. Park and CK. Un, "Word Recognition by a Parallel-Branch subunit model based on misrecognized Data in Semi-Continuous HMM," IEEE Signal Processing Letters, v3 n3 1995 Mar., pp. 66-68.

[9] M. G. Rahim et al., "Signal Conditioning Techniques for Robust Speech Recognition," IEEE Signal Processing Letters v 3 n 4, pp. 107-109, April 1996.

[10] Mei-Ling Shyu, Shu-Ching Chen and R. L. Kashyap, "Database clustering and data warehousing," to appear in *1998 ICS Workshop on Software Engineering and Database Systems*, Dec. 17-19, 1998, Taiwan.