

A Bayesian Network Model for User's Preference Estimation of Personalized TV Service

Han-kyu Lee*, Jihun Cha*, Munchurl Kim**

* Electronics and Telecommunications Research Institute, Korea

** Korea Advanced Institute of Science and Technology, Korea

hkl@etri.re.kr, jihun@etri.re.kr, mkim@ee.kaist.ac.kr

Abstract— In this paper, we propose a statistical method to inference user's preference on watching TV programs. The inference of preference is one of core modules for the personalized TV service, and we introduce a structure for the service. We designed a signal model for usage history and user preference data, and a statistical model as a Bayesian network, and developed an inference method based on message passing algorithm. With a set of real TV viewers' watching records at terrestrial TV receiver, we tested our inference model and present the inference results or genre or channel preference given day and time.

Keywords— User preference, usage history, inference, Bayesian network model, personalized TV

I. INTRODUCTION

With convergence of broadcasting and interactive telecommunication services, consumers want more smart functionalities to TV, such as video recording, video and data on-demand, interactive applications, program guide and recommendations[1]. Such functional extensions make consumer feel uncomfortable and complex to control or access contents at TV. The intelligent functions based on preference analysis and estimation becomes one of important tools to alleviate such complexity to accessing smart TV service.

The Bayesian networks become popular for applications of speech recognition, machine learning, and expert systems, etc. A Bayesian network is a set of graphical structural model and local probability model. The graphical model represents causal and effect relationships between a set of random variables, and the local probability model represents causal strength between the random variables[2,3].

In this paper, a general architecture for a personalized TV system is presented. And for the inference engine, a Bayesian network model for user preference and a statistical inference method using message passing algorithm for the model are proposed. Then, using a set of usage history data from real watching environment, we will show the experiment results to test our model and the inference method.

II. ARCHITECTURE

The personalized TV service provides a user-oriented multimedia experience at the TV terminal. For example, while a program guide provides TV viewers to browse the broadcasting program information, a personalized program guide provides the program information prioritized by user preference. Currently, the smart TV is one of important example requesting the personalized service to users.

Figure 1 shows a structure of terminal for the personalized TV service. The main feature of the structure is a flow of usage history and user preference data in addition to existing data flow of the TV terminal.

Usage history data is a record set to monitor user's behaviour such as watching time and channel. Note that the usage history data becomes diverge with the functional expansion such as interactive user events at data applications, playback control for digital video recorder or on-demand video service, and so on.

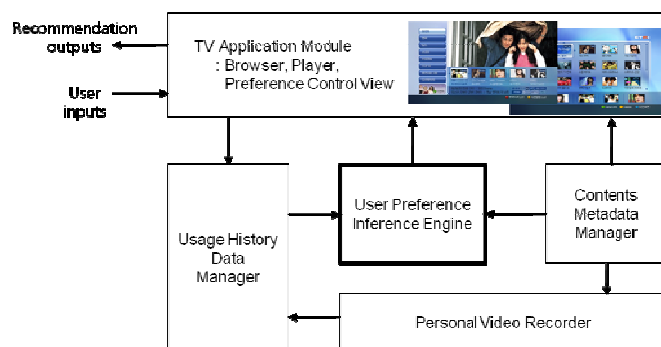


Figure 1. A structure of the personalized TV

The user preference inference module deduces intentions of what user wants at watching TV, and the user preference information describe that intentions. User preference information describes the degree of preference to broadcasting contents' attributes such as genre, channel, keyword, etc. The inference of the user preference is based on the user's usage history data from the usage history data manager and contents

information as contents description metadata from contents metadata manager.

At the structure of personalized TV, the physical location of modules depends on the service type. In the case of a uni-directional broadcasting service, the set-top box or television terminal include all the modules in the structure. In the bi-directional service, the modules can be located at the service provider's side.

For functional expansions of TV service, the inference engine has to easily adapt to the expansion of type of usage history, queries for inference, and services. The Bayesian network model can be one of good solutions for user preference modeling for such knowledge management and expansion [4,5].

III. BAYESIAN NETWORK MODEL

A. Data Modelling

For inference of user preference, we define a data model for usage history data and user preference data.

As input data, the usage history data record user's behavioral events of watching TV. The data structure and one example instance of the data are shown at Table 1. The contents metadata describes program's schedule information, title, genre data, so we can know what user watched on TV at each event of usage history.

TABLE 1. STRUCTURE OF USAGE HISTORY DATA

Date	Day of week	v_start_time	v_end_time	Channel
02-Dec-02	MON.	21:35	21:59	KBS1

For inference of user preference, we define four random variables, $\mathbf{X} = \{D, T, G, C\}$. The random variable $D = \{d_1, \dots, d_7\}$ is a set of amount of watching time at each day of week, from *Monday* to *Sunday*, respectively. $T = \{t_1, \dots, t_5\}$ is a set of amount of watching time at time zone, corresponding to *early morning*, *morning*, *afternoon*, *evening*, *night*. $G = \{g_1, \dots, g_8\}$ is a set of amount of watching time for genres, *News*, *Information*, *Entertainment*, *Drama&Movie*, *Sports*, *Education*, *Children*, *Others*. $C = \{c_1, \dots, c_5\}$ is a set of amount of watching time for channels, five terrestrial broadcasting channels.

B. Bayesian Network Modelling

In probabilistic theory, statistical inference is described by a probability query. Let $\mathbf{X}$ is a set of random variables, $\mathbf{X} = \{X_1, \dots, X_n\}$, a probability query is

$$P(\mathbf{Y}|\mathbf{E} = \mathbf{e}) \quad (1)$$

, where a set of query variables, $\mathbf{Y} = \mathbf{X} - \mathbf{E}$, and $\mathbf{E}$ is a set of evidences instantiated to $\mathbf{e}$.

At previous section, we define the data model for user preference, $\mathbf{X}$. Then, we define a probabilistic model for statistical inference using Bayesian network(BN) model. The BN model represents the causal relationship between random variables in the model, and the relationship is quantified by conditional probability of each variable given its parental variables. Figure 2(a) shows the proposed BN structure with the data model.

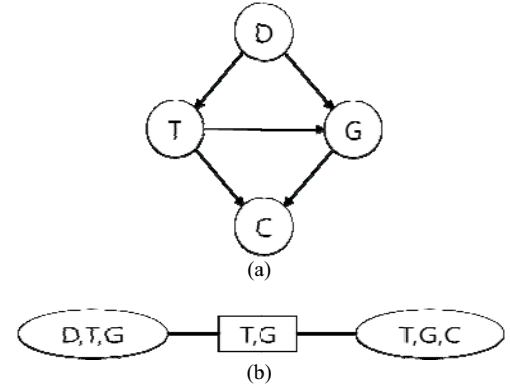


Figure 2. A Bayesian network structure for reasoning user preference: (a) the Bayesian network structure, (b) a junction tree for the BN.

With proposed BN model, we can define the joint probability $P(D, T, G, C)$ into more compact and factorized representations, a set of local probabilities.

$$P(D, T, G, C) = P(D)P(T|D)P(G|D, T)P(C|T, G) \quad (2)$$

From Eq. (2), to calculate the joint probability, we need to define the factored probability functions of left terms of the equation respectively.

Given $\mathbf{X}$, a set of input usage history data, the $P(D)$ is calculated as

$$P(D = d_i) = \frac{v(d_i)}{\sum_i v(d_i)} \quad (3)$$

, where $v(d_i)$ is a function calculates the total time amount of watching at i -th day of week from $\mathbf{X}$.

The conditional probability $P(T|D)$ is calculated as

$$P(T = t_i | D = d_j) = \frac{P(T, D)}{P(D)} \quad (4)$$

, where $P(T, D)$ is

$$P(T = t_i, D = d_j) = \frac{v(t_i, d_j)}{\sum_i \sum_j v(t_i, d_j)} \quad (5)$$

, where $v(t_i, d_j)$ is a function calculates the total time amount of jointly watching time at i -th time zone and j -th day of week from the input usage history data, $\mathbf{X}$.

Similar to $P(T|D)$, we can calculate the joint probabilities $P(G|D, T)$, $P(C|T, G)$, respectively.

After the definition of BN structure and conditional probabilities at each node of the BN, we can evaluate probability queries for user preference such as $P(G|D, T)$,

$P(C|T, G)$. Next, we propose an exact inference method based on the BN model.

C. Exact Inference by Message Passing Algorithm

To calculate the probability query, we apply HUGIN message passing algorithm, which requires converting of a BN model to a junction tree structure. The junction tree is a tree of cliques, clusters of random variables[3]. Figure 2 (b) shows a junction tree structure to our BN model.

At an initial state of the junction tree, each clique represents probabilistic information of the BN, $\mathbf{t}_V$, $\mathbf{t}_W$ as Eq. (6) and (7) respectively.

$$\mathbf{t}_V = P(D)P(T|D)P(G|D, T) \quad (6)$$

$$\mathbf{t}_W = P(C|T, G) \quad (7)$$

And, the separator, represented by a rectangle between two cliques, $\mathbf{t}_V$ and $\mathbf{t}_W$, holds $\mathbf{t}_S$, a table of ones with dimensions of $|Val(T)| \times |Val(G)|$.

The evidence input of a random variable is represented by a finding, table of zeros but one at the position of the evidence inputs. For example, let D has evidence of Monday, then the finding of D is represented as $\mathbf{e}_D = \{1, 0, 0, 0, 0, 0\}$. The evidence inputs are applied to the junction tree by product of findings to all cliques including random variables of evidences.

The message passing operations propagate the changes in probabilistic information by the evidential inputs through the edges of the junction tree. The message passing algorithm has two operations, *DistributeEvidence* and *CollectEvidence*, which is applied until the cliques are consistent.

For the above example, apply production of the finding vector $\mathbf{e}_D$ to D when the D has hard evidence, the *DistributeEvidence* operation has following steps.

$$\begin{aligned} \mathbf{t}_S^* &= \sum_{V \setminus S} \mathbf{t}_V \\ &= \sum_D P^*(D)P(T|D)P(G|D, T) \\ &= P^*(T, G) \end{aligned} \quad (8)$$

$$\mathbf{t}_S = \mathbf{t}_S^* \quad (9)$$

$$\begin{aligned} \mathbf{t}_W^* &= \mathbf{t}_W \frac{\mathbf{t}_S^*}{\mathbf{t}_S} \\ &= P(C, T, G) \frac{P^*(T, G)}{P(T, G)} \\ &= P^*(C, T, G) \end{aligned} \quad (10)$$

, where the $P^*(D)$ is updated probability of D by product of $\mathbf{e}_D$ hard evidence. So, by Eq. (8~10), the effect of hard evidence is propagated from $\mathbf{t}_V$ to $\mathbf{t}_W$ through the separator $\mathbf{t}_S$.

The *CollectEvidence* operations are similar to the *DistributeEvidence*.

$$\begin{aligned} \mathbf{t}_S^{**} &= \sum_{W \setminus S} \mathbf{t}_W^* \\ &= \sum_C P^*(C, T, G) \\ &= P^{**}(T, G) \end{aligned} \quad (11)$$

$$\mathbf{t}_S^* = \mathbf{t}_S^{**} \quad (12)$$

$$\begin{aligned} \mathbf{t}_V^{**} &= \mathbf{t}_V^* \frac{\mathbf{t}_S^{**}}{\mathbf{t}_S^*} \\ &= P^*(D, T, G) \frac{P^{**}(T, G)}{P^*(T, G)} \\ &= P^{**}(D, T, G) \end{aligned} \quad (13)$$

After the two evidence input's propagation is applied to junction tree, we can calculate probability queries by marginalization to the clique. For example, $P(G|d_i, t_j)$ is calculate as

$$P(G|d_i, t_j) = \frac{\mathbf{t}_V^*}{\sum_G \mathbf{t}_V^*}.$$

Note that the evidence inputs of D, T are applied by product of findings of D, T to the cliques before above message passing procedure.

IV. EXPERIMENTS AND THEIR RESULTS

To evaluate our inference model, we use a set of TV viewers' usage history data of 6 months, which have the structure of Table 1. To evaluate our inference model, we use a set of TV viewers' usage history data of 6 months. We applied the usage history of the first 5 months to learn the BN's prior probabilities, and then we calculated the posterior probabilities as the user's preference estimation, which are the genre and channel preferences given Day and Time. The accuracy of the preference estimation is measured by L_1 distance between the posterior conditional probabilities and last 1month's corresponding probabilities. Figure 3 shows a genre preference given day and time zone estimated from a user's usage history data.

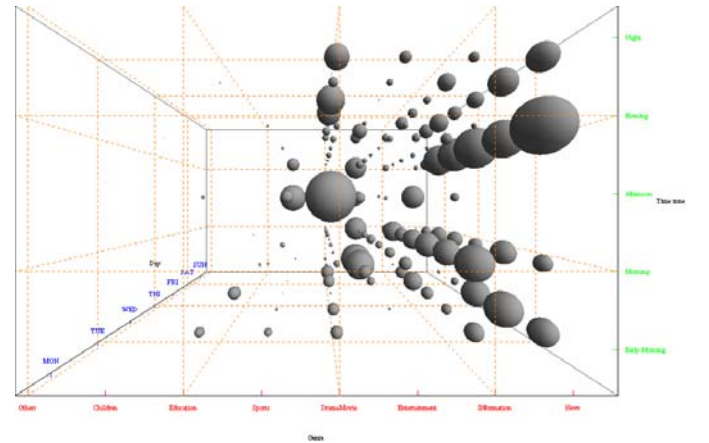


Figure 3. A genre preference estimation results given day and timezone

For a certain day of week, d_i and time zone, t_j , we can estimate the preference by the value of $P(G|d_i, t_j)$. Figure 4 shows 3 cases of the preference estimation results, Monday morning, Wednesday late night, and Sunday evening,

respectively. At Figure 4, we can see preference at specific genre at the give day of week and time zone of the TV viewer. This preference information is used for TV program recommendation at the corresponding given condition.

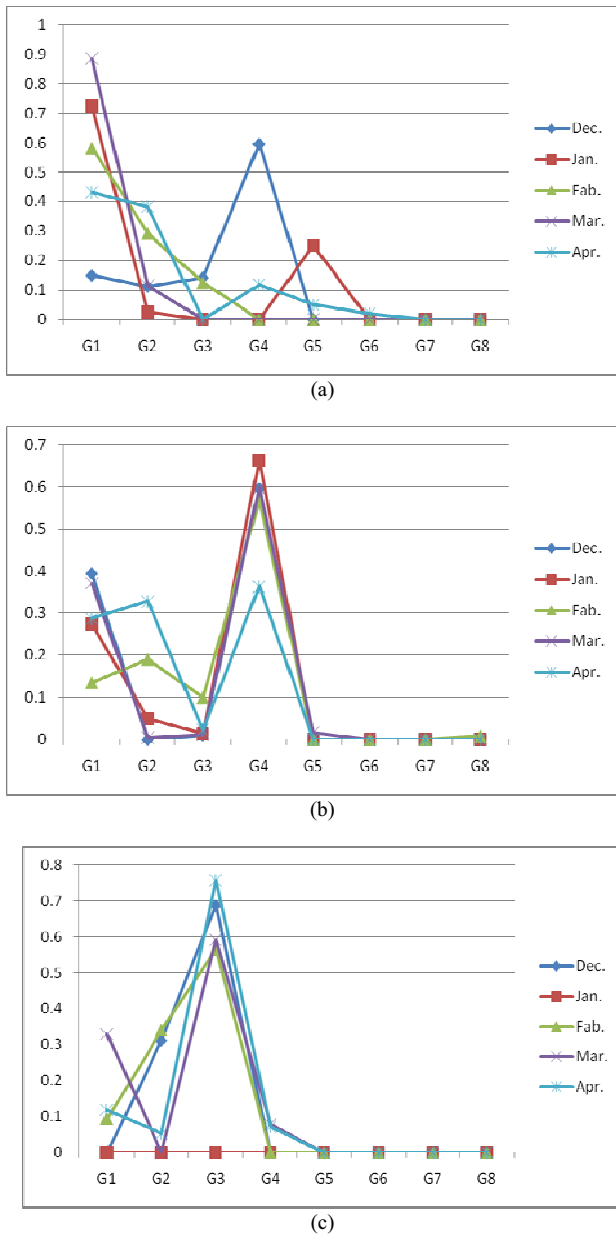


Figure 4. A genre preference estimation results (a) Monday's morning, (b) Wednesday's late night, and (c) Sunday's evening.

Similarly, we can estimate preference of channel given genre and time zone directly from the probabilistic information from the proposed BN network model.

V. CONCLUSIONS

The modeling and estimation of user preferences become more important for higher intelligent services on TV environment.

In this paper, we proposed a BN structure and its probabilities' updating method for statistical estimation of user preference. The BN model can be expanded to add variables as evidences and queries as extensions of TV functions. As future works, we will expand the model into a dynamic BN structure and will include the variables for interactive TV environment.

REFERENCES

- [1] ETSI, "TV-Anytime Part 1: Benchmark Features," TS 102 822-1
- [2] F. V. Jensen, *An Introduction to Bayesian Networks*, Springer, 1996.
- [3] D. Koller and N. Friedman, *Probabilistic Graphical Models Principles and Techniques*, MIT Press, 2009.
- [4] K. Kang, et al., "Metadata broadcasting for Personalized Service: A Practical Solution," *ETRI Journal*, Vol. 26, Number 5, p.452-466, 2004
- [5] S. Kang, J. Lim, and M. Kim, "Modeling the user preference of broadcasting content using Bayesian networks," *J. of Electronic Imaging*, vol. 14, 023022(2), 2005.