Generating Cartoon-Style Summary of Daily Life with Multimedia Mobile Devices*

Sung-Bae Cho, Kyung-Joong Kim, and Keum-Sung Hwang

Dept. of Computer Science, Yonsei University
134 Shinchon-dong, Sudaemoon-ku, Seoul 120-749, South Korea
{kjkim, sbcho}@cs.yonsei.ac.kr

Abstract. Mobile devices are treasure boxes of personal information containing user's context, personal schedule, diary, short messages, photos, and videos. Also, user's usage information on Smartphone can be recorded on the device and they can be used as useful sources of high-level inference. Furthermore, stored multimedia contents can be also regarded as relevant evidences for inferring user's daily life. Without user's consciousness, the device continuously collects information and it can be used as an extended memory of human users. However, the amount of information collected is extremely huge and it is difficult to extract useful information manually from the raw data. In this paper, AniDiary (Anywhere Diary) is proposed to summarize user's daily life in a form of cartoon-style diary. Because it is not efficient to show all events in a day, selected landmark events (memorable events) are automatically converted to the cartoon images. The identification of landmark events is done by modeling causal-effect relationships among various events with a number of Bayesian networks. Experimental results on synthetic data showed that the proposed system provides an efficient and user-friendly way to summarize user's daily life.

1 Introduction

Many people used to write a diary in order to memorize their daily life and it can help them recall what they did. When they write a diary, they attempt to remember daily events. Because it is not easy to remember all the events in the day, he needs to check other information sources such as pictures taken, call logs, SMS (short message service), scheduler and memos. Recent advances in mobile phone technologies allow the device not only to log user's daily activity (location, call log, SMS, and proximity information (Bluetooth)) but also to store personal multimedia information (photo, video and documents). The information is very private and contains relevant information to reveal user's daily events.

Our research aims to summarize user's daily life in a form of cartoons based on the information collected from mobile devices such as Smartphone. Cartoon-style representation has been used to summarize user's conference visit experience on PDA from user's explicit input without much exploitation of user's logs [1]. Although their

* This research was supported in part by MIC, Korea under ITRC IITA-2006-(C1090-0603-0046).

H.G. Okuno and M. Ali (Eds.): IEA/AIE 2007, LNAI 4570, pp. 135–144, 2007.

works showed that cartoon-style representation was useful compared to plain text representation, the scope of comic diary was limited to the conference visit scenario. In this work, the scope is extended to the daily life and all the available information sources in the Smartphone are used to derive cartoon-style diary automatically. The procedure of the diary generation is composed of information logging, preprocessing, landmark detection and selection, and cartoon generation. Landmark means very relevant or novel events that are useful to recall a sequence of events. Its organization is similar to the human memory structure and researchers try to adopt such concept in real-world systems [4].

In logging stage, GPS, call log, SMS, MP3 play lists, battery level and photo viewer usage information are recorded. Because extracting high-level information from the raw information directly is not efficient, statistical analysis on them is first done before the inference. Min, max and impact analysis are used to extract relevant trends or patterns of user's daily life. In landmark detection module, a number of Bayesian networks designed by experts are used to find memorable events. Bayesian networks are one of efficient tools to infer the situation given uncertain and partial information. Given memorable events selected, the most probable candidates are selected based on the relevance. Finally, they are converted to the cartoons by composing pre-stored cartoon image components. Figure 1 shows the procedure.

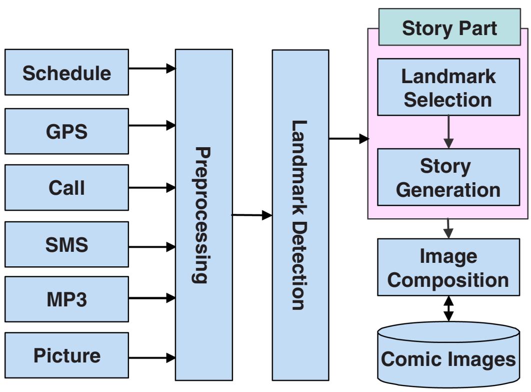


Fig. 1. The procedure of Anidiary generation

2 Related Works

Automatic diary generation from user's log (explicit or implicit) is one of hot research topics. Sumi *et al.* designed comic diary to summarize the user's conference tour in a cartoon-style form [1]. Their system is based on the explicit user input and user's schedule information. Eagle tried to develop a diary system based on the log information collected from cellular phone [2]. Because they show the raw information directly through GUI, it is difficult to understand the whole picture of the day in an intuitive way. A large amount of information is continuously stored from cellular phone and it is required to learn high-level context-information from the raw data but they do not focus much about the point. Nokia's Lifeblog provides a way to store and manage user's photo, multimedia and SMS in a chronological manner [3]. It does not utilize any abstraction and summarization methods.

ContextPhone is a context logging software for Nokia 60 series Smartphone [5] and its source is available to public. It collects information including photo, sound, battery level, location, SMS, MMS, call logs, Bluetooth, and active applications. However, it is not easy to use such software in general because Nokia 60 series Smartphone is not available in some places. Because of this, we have developed a new context logging software and applied to Samsung Smartphone M4300.

Eric Horvitz *et al.* attempt to re-organize personal information storage in desk top PC into an episodic style memory [4]. Previous research on human memory revealed that the organizational principle is episodic storage and retrieval. Related events are grouped as an episode and landmark event is used to recall them from the storage. By finding landmark, he can also recall the related items. He learns Bayesian networks to detect landmark event from the data stored in outlook scheduler. His work called Bayesphone uses client-server communication for Bayesian network whose inference is done on server-side and the results are transmitted to device through network [6]. We expand the scope of his works to the personal storage in mobile devices.

MIT reality mining group develops serendipity service using the ContextPhone software [7]. The group collaborates with MIT common sense reasoning group to generate diary automatically. Because the research is at early stage, there is no concrete result about that. Only visualization tool for collected log is available in their paper. However, their work shows a new way to generate more interpretable high-level diary using common sense. Basic details about the common sense knowledge can be found in [8]. Our work is based on the ontology and it can be expanded to the more general model using such a common sense corpus.

Although there are many initial-stage papers about personal information management on mobile device, visualization of daily life, and detection of memory landmark, they are not integrated into one system for useful applications. In this paper, we will design an integrated system of logging, preprocessing, landmark detection, and visualization.

3 Cartoon Generation for Summary of Daily Life

In logging stage, basic user's log from system, sensors and web is collected and stored to file system. In preprocessing stage, the raw information is analyzed using statistical methods such as average, min, max, and more sophisticated tools. Impact factor is proposed to reflect the importance of event's density over time. In landmark detection stage, modular structure is adopted to minimize computational cost for inference and the expert's effort to design the knowledge-base. A number of Bayesian networks are designed at this stage to infer the probability of landmark given evidences. At story generation stage, the most probable landmarks among them are selected and their order for visualization is determined based on pre-stored knowledge (causal relationships). Finally, cartoon generation module generates a diary by converting the landmarks to cartoons.

3.1 Logging , Preprocessing and Memory Landmark Detection

We have implemented a logging system on Samsung Smartphone M4300 with small GPS receiver attached to the device. Weather information is retrieved from http://www.kma.go.kr (Korea Meteorological Administration). GPS and battery level are sampled per 1 sec. Photo viewer and MP3 player are modified based on free source code to log usage information. User's current position is inferred from the GPS value that is a pair of (longitude, latitude). The raw information is transformed into semantic label using pre-stored information about the relationships between GPS value and semantic labeling information (building name and street name). The impact factor reflects the density of events. When the event occurs, the impact factor is increased and it decreases after pre-defined time. However, some events continuously occur after the event, the impact factor also continuously increases. Many events with less density (sparse distribution) will show low impact factor.

Various mobile log data are preprocessed in advance, and then the landmark reasoning module detects the landmarks. The BN (Bayesian Network) reasoning module performs complicated and probabilistic inference.

There are two general differences between the proposed Bayesian networks and the conventional Bayesian networks for landmark detection. Firstly, we modularize the Bayesian inference models according to their separated sub-domains (Figure 2) [10]. Secondly, in order to consider the co-causality of modularized BN, the proposed method operates 2-pass inference stages. The output of one module can be inputted to the other module for the 2nd inference.

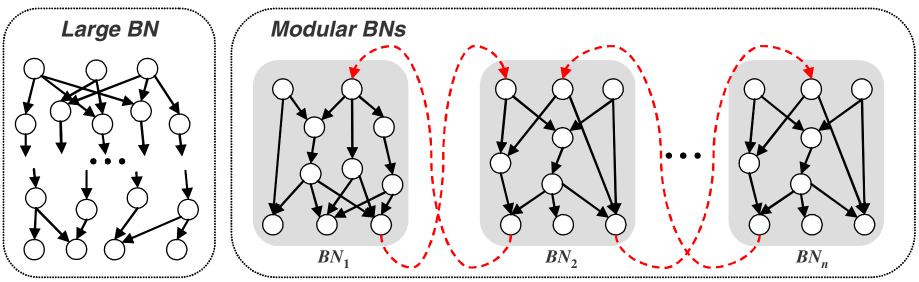


Fig. 2. The 2-pass inference process for the cooperation of modular Bayesian networks. The dotted line indicates the stream of the 2nd stage of inference processing.

There are four kinds of BNs, and totally 39 BNs are used. The BNs are as follows: Place-activity BNs={house, religion, shopping, photo, hospital, nature, meeting, workplace, sports, movement, food, call, music, school, traffic, watch, rest}, Emotion/condition BNs={joy, hungriness, hot, nonsense, surprise, busy, tiredness, drunken, anger, cold, fret, amusement, gloom, sick, bored}, Circumstance situation BNs={space, climate, time, device, group}, and event BNs={anniversary, event}.

3.2 Story and Cartoon Generation

Single cartoon cut is composed of 5 different types of images: text, sub character, main character, sub background, and main background (Figure 3) [9]. Designers prepare a set of images for each type. Cartoon generation module selects the character and background images that are the most appropriate given landmarks and user profiles. In the selection process, semantic similarity between images and landmark event is calculated based on the predefined annotations for the images. After selecting the cartoon cuts for each landmark, they are organized into a story stream to make a plausible cartoon story using consistency constraints.

The semantic similarity between the cartoon images and landmark events are calculated using the equation (1). User profile is a set of values that reflect the user's preference about the cartoon image. From the images whose similarities are higher than certain threshold, the candidate list of cartoon cuts is generated. The threshold for image selection is defined in consideration of cartoon diversity and computational cost. Lower threshold composes more diverse cartoon expression, but requires more computational cost for the consistency constraint violation check.

$$Similarity\ (event_m, image_{kmh}) = \frac{|(p(event_m) \cup p(UP)) \cap p(image_{kmh})|}{|p(event_m) \cup p(UP) \cup p(image_{kmh})|} \qquad (1)$$

$Similarity(event_m, image_{kmh})$ = similarities between $event_m$ and $image_{kmh}$
$event_m$ = the m^{th} landmark event
$image_{kmh}$ = cartoon image of type h in the k^{th} feasible cartoon cut composition for the m^{th} landmark event
\# of feasible cartoon cut composition
= # of Main BG×# of Sub BG×#of Main CH×# of Sub CH×# of Text
UP = User profile
$p(X)$ = A set of attributes that object X has

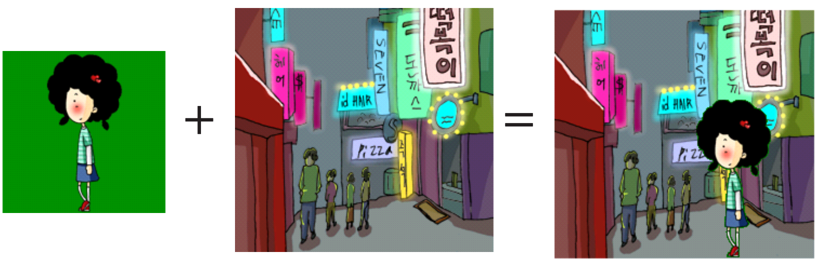


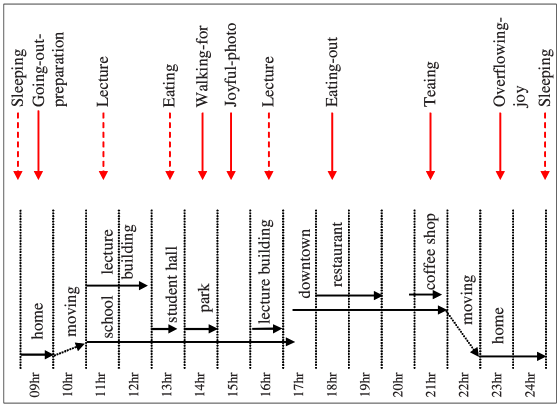
Fig. 3. Composition of image components for single cartoon cut

4 Experimental Results

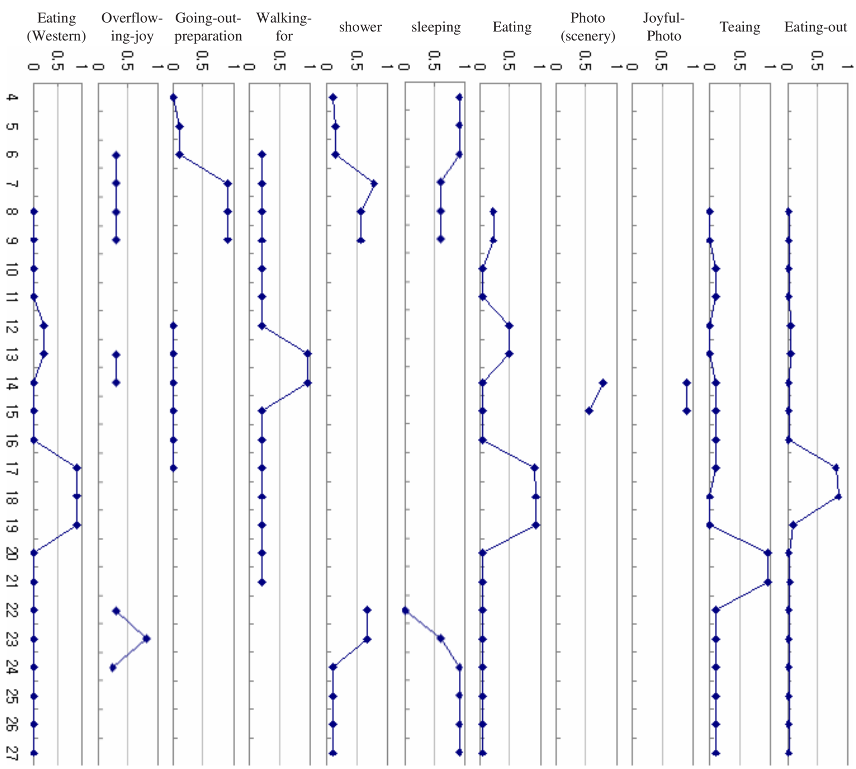
At this stage, the performance of the system is tested using synthetic data because collecting real data requires much time. It takes more than one month to get useful statistics.

4.1 A Case Study and Performance Evaluation on Long-Term Data

We have tested the proposed landmark reasoning model with a scenario in order to confirm the performance. The left side of Figure 4 shows a scenario used. The BN set strongly related to the scenario is {food, photo, movement, nature, joy, home}. The probabilities are calculated when the related evidences are given.



(a)



(b)

Fig. 4. (a) A scenario of an everyday life with mobile device of an undergraduate student for experiments. (b) The observation of the probability values of 11 target landmarks. The denoted time is from 4 o'clock to 27 o'clock (equal to 3 o'clock next day).

After generating the log contexts with quantity of a day, we have tested them. The right side of Figure 4 shows the inference results. We can see the increment of the probabilities of the related landmarks at the corresponding time. For example, there are ‘going-out-preparation’ and ‘shower’ landmarks at 7~9 o'clock, ‘eating’ at 12~13 and 17~19 o'clock, ‘walking-for’ at 13~14 and 20~21 o'clock, ‘joyful-photo’ at 14~15 o'clock, and ‘eating-out’ and ‘eating (western style)’ landmarks at 17~19 o'clock.

To evaluate the landmark reasoning performance, we grouped situations of everyday life on the two aspects: usual/unusual and idle/busy. We generated artificial high-level contexts for evidence because it was difficult to make raw data directly. For example, we made a context ‘a lot of phone calls’ instead of its phone call log data.

The target landmarks are categorized to three classes as shown in Table 1. There is no idle class because it is regarded as idle if it is not busy. The data set for a class was collected for 30 days and a day of data set contains two landmarks, which are selected randomly and one is before noon and the other is afternoon.

Table 1. The category of landmarks based on usual, unusual, and busy

Class (#)	**Related Landmarks**
busy (6)	busy, run with umbrella, joyful phone call, joyful SMS, annoying SMS, tired phone call
usual (82)	boarding airplane, …, boarding train, studying, meeting, …, walking for, showing, washing dish, caring for hair, washing face, shopping, sleeping, eating, painful, feel hot, overflowing-joy, …, sad, heavy traffic, joyful photo, …, self-photo, playing golf, …, playing basketball, mountain climbing, water skiing, swim, health-training
unusual (25)	Watch horse racing, …, watch performance, watching, anniversary, watching basket ball, very strange place, unusual place, watching baseball, trip, receiving celebration-SMS, …, make a celebration-phone-call, marriage, election, visiting the family graves, …, funeral ceremony, ancestor-memorial service of a festive day

Table 2 shows the statistics of experimental results and Table 3 shows some individual results. We have excluded the landmarks related to the default place 'home' and the low-weight landmarks from main landmarks set. The false-positive error of 'usual/idle' class is high and the precision is low, because 'usual' class includes many places and landmarks that have evidence duplication. For example, since the landmark 'boarding ship' of A-60302 in Table 3 is caused by the evidence 'sea' or 'river', a landmark 'swim (outdoor)' can be extracted. The false-positive error of 'usual/busy' class is low because the class includes relatively many landmarks that have distinct evidence. In the experiment, the overall recall rate was low as 75%. It results from the lack of tuning or the landmarks hard to detect.

Table 2. The experimental results with synthetic data. Two target objects (with few redundancies) are selected in each data set. 'unusual/busy' class data are composed of one 'unusual' landmark and one 'busy' landmark.

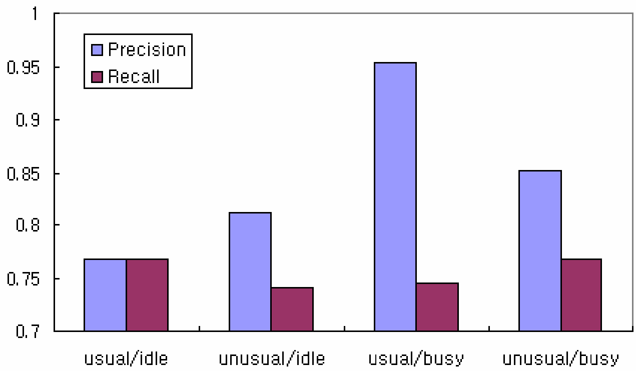


Table 3. A part of the obtained landmarks. We have excluded the landmarks related to the default place 'home' and the low-weight landmarks from main landmarks set. The target landmarks found are underlined. Abbreviations: OL (Number of Obtained Landmark), TP (True Positive), FP (False Positive). Data type: A (usual/idle), B (unusual/idle), C (usual/busy), D (unusual/busy).

Data	Target landmark	OL	TP	FP	Extracted main landmarks
A-60301	athletic meeting, boarding train	45	2	2	Athletic meeting, boarding subway train, boarding expressway bus, boarding train
A-60302	boarding ship, boarding train	47	1	3	Swimming (outdoor), boarding subway train, boarding expressway bus, boarding train
…	…	…	…	…	…
A-60330	boarding airplane, overflowing-joy	35	1	0	Swimming (outdoor), boarding airplane
Total	60 landmarks	866	46	14	78 landmarks

4.2 Image Generation Test

Because the fun of the composed cartoon story can be evaluated only in subjective measure, we evaluate the fun of the cartoon by evaluating the diversity and the consistency of the cartoon. To evaluate the generated cartoon story, we create sample scenario landmark events described in Table 4.

Figure 5 shows one of the composed cartoon story images. To evaluate the method, four composed cartoon story images were shown to participants. By composed cartoon stories, user evaluation tests for cartoon cuts and cartoon story were performed with five participants. For each question, 5-point scale measure is used to grade an item. At first, cartoon cuts generated by a landmark example are evaluated in the criteria of the diversity and the descriptiveness. Figure 6(a) shows the evaluation results. After cartoon cut evaluation, diversity and consistency of four cartoon stories were asked to the participants. The diversity of cartoon stories results in 4.0 of average with 1.0 of standard deviation, and the consistency of cartoon stories is depicted in Figure 6(b). These data are not enough for the statistical significance, but the average score for each criterion shows affirmative tendencies.

Table 4. Landmark context examples

ID	Behavioral landmarks	Environmental landmarks
1	Movement, MP3, Stand	Bus, Indoor
2	Study, Take a class, Sit	University, Classroom, Indoor
3	Eat, Korean Food, Sit	University, Dinning Room, Indoor
4	Talk, Phone, Stand	University, Outdoor
5	Cheer, Watch, Stand	Stage, Outdoor
6	Drink, Beer, Sit, Cheer	Drinking House, Bar, Indoor

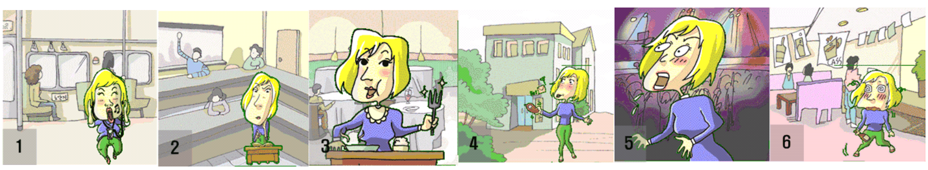
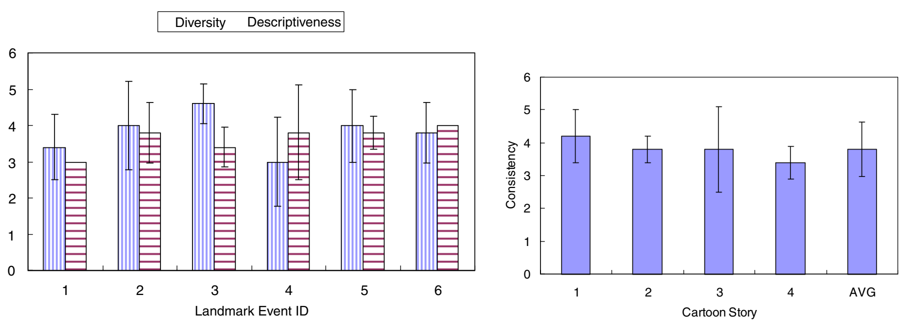


Fig. 5. A composed cartoon story from landmark examples in Table 4 to evaluate the diversity and the consistency



(a) (b)

Fig. 6. (a) Diversity and descriptiveness of cartoon cuts for each landmark event (n=5). (b) Consistency Scores of four generated cartoon stories (n=5).

5 Conclusions and Future Works

In this paper, we have proposed a cartoon diary generation system based on user's personal logs stored in mobile devices as a form of episodic memory. Modular Bayesian networks are used to infer the probability of landmark for each event. Some selected landmarks are arranged using consistency ontology and converted to cartoon images. Experimental results on synthetic data show the possibility of the proposed methods for detecting landmarks and generating meaningful images. User's enormous logs are summarized into user friendly cartoon images and this gives a new way for managing his personal memories.

As a long-term goal, the system needs to be evaluated from real logs from subjects' real life. Also, learning algorithm for landmark detection is required for personalized detection models.

References

1. Sumi, Y., Sakamoto, R., Nakao, K., Mase, K.: ComicDiary: Representing individual experiences in a comic style. Ubicomp 2002, pp. 16–32 (2002)
2. Eagle, N.: Machine Perception and Learning of Complex Social Systems. Ph. D. Thesis. In: Program in Media Arts and Sciences, Massachusetts Institute of Technology (2005)
3. Nokia Lifeblog, http://www.nokia.com/lifeblog
4. Horvitz, E., Dumais, S., Koch, P.: Learning predictive models of memory landmarks. In: 26th Annual Meeting of the Cognitive Science Society (2004)
5. Raento, M., Oulasvirta, A., Petit, R., Toivonen, H.: ContextPhone-A prototype platform for context-aware mobile applications. IEEE Pervasive Computing 4(2), 51–59 (2005)
6. Horvitz, E., Koch, P., Sarin, R., Apacible, J., Subramani, M.: Bayesphone: Precomputation of context-sensitive policies for inquiry and action in mobile devices. User Modeling 2005, 251–260 (2005)
7. Kurlander, D., Skelly, T., Salesin, D.: Comic chat. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pp. 225–236 (1996)
8. Singh, P., Barry, B., Liu, H.: Teaching machines about everyday life. BT Technology Journal 22(4), 227–240 (2004)
9. Song, I.-J., Jung, M.-C., Cho, S.-B.: Automatic generation of funny cartoons diary for everyday mobile life. In: Sattar, A., Kang, B.-H. (eds.) AI 2006. LNCS (LNAI), vol. 4304, pp. 443–452. Springer, Heidelberg (2006)
10. Hwang, K.-S., Cho, S.-B.: Modular Bayesian networks for inferring landmarks on mobile daily life. In: Sattar, A., Kang, B.-H. (eds.) AI 2006. LNCS (LNAI), vol. 4304, pp. 929–933. Springer, Heidelberg (2006)