Modular Bayesian Network for Uncertainty Handling on Mobile Device

Keum-Sung Hwang
Department of Computer Science, Yonsei University, Seoul, Korea
yellowg@ cs.yonsei.ac.kr

Sung-Bae Cho
Department of Computer Science, Yonsei University, Seoul, Korea
sbcho@ cs.yonsei.ac.kr

Abstract

Mobile devices can now handle a great deal of information thanks to the convergence of diverse functionalities. Mobile environments have already shown great potential in terms of providing customized services to users because they can record meaningful and private information continually for long periods of time. Most of this information has been generally ignored because of the limitations of mobile devices and the uncertainty of mobile environments in real world. In this paper, we propose an approach based on modular Bayesian networks to overcome these problems and to analyze various kinds of log data. The method adopts a probabilistic approach to manage the uncertainty and decomposes the probabilistic model automatically to decrease complexity and how to infer the model, which is called cooperative reasoning. In the experimental results, the proposed methods were evaluated with mobile log data collected in the real world.

Keywords: Modularized probabilistic reasoning, Landmark detection.

1 Introduction

Mobile environments have very different characteristics from desktop computer environments. First of all, mobile devices can collect and manage various kinds of user information, for example, by logging a user's calls, SMS (short message service), photography, music-playing and GPS (global positioning system) information. Also, mobile devices can be customized to fit any given user's preferences. Furthermore, mobile devices can collect everyday information effectively. Such features allow for the possibility of diverse and convenient services, and have attracted the attention of researchers and developers. Recent research conducted by Nokia is a good example [1]. Especially, the context-aware technique that has recently been widely researched is more applicable to mobile environments, so many intelligent services such as intelligent calling services [2], messaging services [3], analysis, collection and management of mobile logs [4,5] have been actively investigated.

However, mobile devices do present some limitations. They contain relatively insufficient memory capacity, lower CPU power (data-processing speed), smaller screen sizes, awkward input interfaces, and limited battery lives when compared to desktop PCs. In addition, they have to operate in the changeable real world, which means that they require more active and effective adaptation functions [6].

In this paper, we propose a novel way of analyzing mobile log data effectively and extracting semantic information and memory landmarks, which can be used as special ways of helping recall specific functions [7].

The proposed method adopts a Bayesian probabilistic model to efficiently manage various uncertainties that can occur when working with mobile environments, including real-world irregularities, like varying levels of attention and emotions, inaccuracy of sensors, and uncertain causal factors.

The proposed model uses a cooperative reasoning method with a modular Bayesian network (BN) in order to work competently in mobile environments.

We also discuss how to discover and update the Bayesian inference model by using the proposed BN learning method with training data. The proposed method was applied to a PC system, using real mobile log data collected with

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 402–408
Torremolinos (Málaga), June 22–27, 2008

a smart phone over a period of sixteen days in the real world.

1.1 Prior Work

In [8], the authors proposed a method for identifying landmarks on mobile devices. The preliminary investigation is the foundation of the present work.

1.2 Related Work

There have already been various attempts to analyze log data and to support expanded services by using the probabilistic approach. Li, et al. used a probabilistic model for active affective state detection of user [9]. They utilized dynamic Bayesian network and utility theory to reason the 'fatigue,' 'nervous,' and 'confused' states. They showed that the probabilistic approach was good at management of uncertain information like affection.

Y. Zhang, et al. proposed an active and dynamic information fusion method for multi-sensor systems with dynamic Bayesian networks [10]. This showed the usefulness of Bayesian approach for information fusion. These works showed the Bayesian probabilistic approach was good tool for handling, reasoning, and combining uncertain information.

Krause, et al. clustered the sensor and log data collected on mobile devices, discovered a context classifier that reflected a given user's preferences, and estimated the user's situation in order to provide smart services [11]. The context classifier was constructed using the BN model, which was based on a general learning method for a small domain of classification subjects.

However, these methods were not suitable for mobile devices that were limited in terms of capacity and power. For larger domains, the general BN and BN learning method require highly complex computation. This is a crucial problem when it comes to modeling everyday life situations with mobile devices. To overcome these problems, a more appropriate approach was necessary. The following researchers have studied methods of reducing the levels of complexity.

Marengoni, et al. [12] tried to reduce the complexity levels of the BN model by dividing it into several multi-level modules and using procedural reasoning of the connected BNs (just like chain inference). However, this method required procedural and classified properties of the target functions.

Tu, et al. [13] proposed a hybrid BN model that allowed hierarchical hybridization of BNs and HMMs. However, it supported only links from lower level HMMs to higher level BNs without consideration of links between same level BNs. They also remained the hybridization of low and high level BNs as future works.

2 Landmark Detection on Mobile Device

The overall process of landmark extraction from the mobile log data used in this paper is shown in Fig. 1. Various mobile log data is preprocessed in advance, and then the landmark-reasoning module detects the landmarks. The preprocessing module is operated by the techniques of statistical function and some rules. The BN reasoning module performs probabilistic inference.

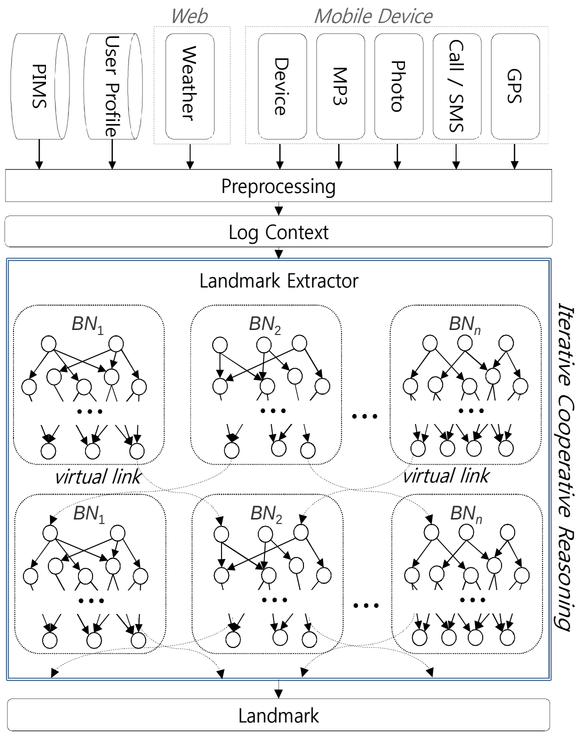

Fig. 1: The process of the landmark extraction from mobile log data. The iterative cooperative reasoning method is used to infer the modular Bayesian networks. The dotted arrow indicates virtual linking.

BNs refer to models that can express large probability distributions with relatively low complexity. They have the structure of a directed acyclic graph (DAG) that represents the link (arc) relations of the node, and has conditional probability tables (CPTs) that are constrained by the DAG structure [14].

2.1 Collection and Preprocessing

Table 1 shows log information collected on a mobile device and on the internet. The GPS log presents the places that the user visited, and the call and SMS logs show the user's calling patterns. The MP3 (a music file format) player log offers an idea of the user's emotions and the photograph log shows when the user wanted to memorize something.

Table 1: The log information that was collected on a mobile device

Log	Information
GPS	latitude, longitude, velocity, direction, date, time
Call	caller's phone number, call/receive/absence log, time span, start/end time
SMS	sender's phone number, call/receive/absence log, time span, start/end time
Photographing	photo file name, taking time
Weather	weather, visibility range (km), cloud degree (%), temperature (°C), discomfort index, effective temperature (°C), rainfall (mm), snowfall (cm), humidity (%), wind direction, wind velocity (m/s), barometer (hPa)
MP3 Player	title, time span, start/end time
Charging	charging status, time span, start/end time

Since logs have temporal properties, we considered their time spans, frequencies (per hour, daily, weekly), and start/end times. The coordinates from the GPS log are used to get place names. In this paper, we divided the domain area into a lattice and then labeled each region. The user profiles and PIMS (personal information management system) datasets were used to find the user's social position (student, worker), gender, the position of their home, and the names and phone numbers of their friends and relatives.

2.2 Cooperative Reasoning of Modular Bayesian Networks

There are two general differences between the proposed BNs and conventional BNs. Firstly, we modularized the Bayesian inference models according to their separated sub-domains. The BN model essentially requires more computing power depending on the number of nodes and arcs. Especially, since the computational complexity of Bayesian inference is approximately proportional to $O(k^N)$, where k is the number of states and N is the number of causal nodes, the modularized BN is more efficient.

Secondly, to consider the co-causality of the modularized BN, the proposed method shows N-pass inference stages (in this paper, we take 2-pass inference) [8]. A virtual linking technique is utilized to reflect the co-causal evidence more correctly. The technique is performed to add the virtual nodes and regulate their conditional probability values (CPVs) to apply the probability of the evidence, which is called virtual evidence. The concept of the virtual node and evidence is shown in Figs. 2 and 3.

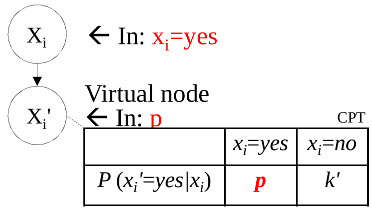

Fig. 2: A virtual node to use virtual evidence containing probabilistic value.

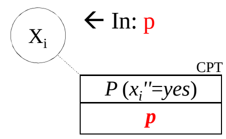

Fig. 3: A virtual root node that does not have additional node. The prior probability value is substituted for evidence.

2.3 Complexity Analysis

Equation (1) shows the time complexity of the exact inference of the BN using the Lauritzen Speigelhalter (LS) algorithm [15], which is the most popular exact inference algorithm and a junction tree-based algorithm, where n represents the number of nodes, k represents the maximum number of parents for each node, r denotes the number of values for each node, and w represents the maximum clique that each parameter used in the LS algorithm [16].

$$cmpx_{\text{inf}} = O(k^3 n^k + wn^2 + (wr^w + r^w)n) \tag{1}$$

We have replaced the maximum clique size w with k, since the clique size is proportional to the parents' size, and the number of values r is about 2 in this paper. The ideal modularization technique roughly divides the number of nodes by the number of modules d but it has to compute d times for d modules, so the exact inference complexity of the modular BNs is as shown in Equation (2) where the new number of nodes is (n/d).

$$cmpx_{\inf}' \cong O(\frac{k^3}{d^{k-1}}n^k + \frac{k}{d}n^2 + 2^k(k+1)n) \quad (2)$$

3 Modular Bayesian Networks Modeling

In this section we introduce a method of discovering the proposed modular BNs automatically from the training data set.

3.1 Discovering of the Bayesian Network

The BN model G can be defined as (B_s, Θ), which means a network structure B_s and a probability parameter set Θ. Θ ={B_Φ, B_p} is composed of the conditional probability table B_Φ and the prior probability distribution B_p. In this section, we build the structure B_s with the BN learning technique, and the parameter θ is calculated from the training data set D by using Equation (3).

$$\theta^* = \arg\max_\theta P(D \mid \theta)P(\theta) \quad (3)$$

$P(\theta)$ means prior probability. The general discovery process is shown by Equation (4) where Z_T={z_1, z_2, ..., z_T} represents a set of T status variables, and Y_T is a set of T observation result variables.

$$P(Z_T, Y_T, \theta) = P(Y_T \mid Z_T, B_\phi)P(Z_T \mid B_P) \quad (4)$$

The conditional probability table B_Φ is calculated from the relational frequency between the observation values and the status variables, meaning the probabilistic histogram of the training data.

3.2 Discovering the Modular Bayesian Networks

The proposed discovering process of the modular BNs is shown in Fig. 4. The method includes structure learning, modularization, and parameter learning process of BN. In this paper we adopts K2 algorithm [17] to learn the network structure from the training data, which is denoted in the Fig. 5.

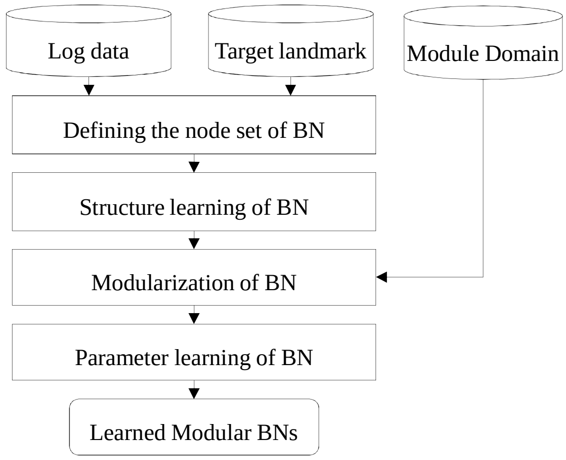

Fig. 4: The learning procedure of modular Bayesian network.

The Fig. 5 includes a parameter setting process (for node set X, the topological order set of nodes O, a level set of nodes V and the maximum size of parents that each node had p). The node set of BN (X) is composed of the union of the collected log context set (L) and the target landmark set (LM). The values of the landmark set are defined by users.

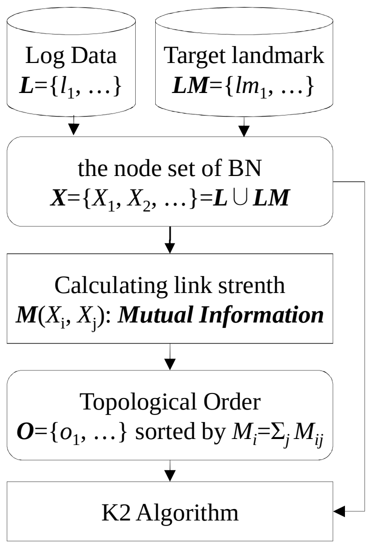

Fig. 5: The BN structure learning procedure

To determine the topological order (O), we ranks each variable based on its total influence on the other variables with mutual information. The K2 learning method and the mutual information computation are presented in the next section.

The network structures discovered are divided into several modules by a modularization process as shown in Fig. 6, where G is the network structure and d denotes the number of modules. The 1st process defines the nodes of the module-BNs based on the module domain and the log context set. The 2nd Process defines the arcs of the module BNs based on the network structure (*G*). The 3rd Process defines the virtual nodes based on the network structure (*G*).

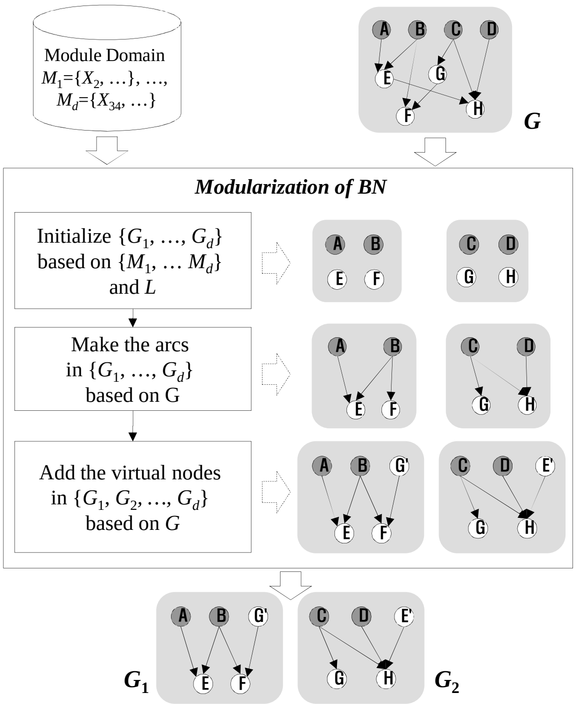

Fig. 6: The modularization process of BN. The right side shows an example, in which the gray nodes denote evidence nodes and the white nodes mean landmark nodes or their virtual nodes.

3.3 K2 Algorithm

The K2 method was proposed by Cooper and Herskovits [17]. It is shown in Fig. 7. It is the most popular BN algorithm and the basis of many advanced discovery algorithms. This algorithm adopts a score metric known as the K2 metric, which calculates scores based on the difference between the BN graph and the training data distribution. The search process of the K2 algorithm is greedy and heuristic.

The K2 algorithm uses a topological order to maintain the graph as the DAG by maintaining that the prior node cannot be the child of the posterior node without any other DAG checking rules. However, we have to optimize the topological order since a different topological order will have led to a different BN structure. In this paper, we compute the influence score of all the nodes by using mutual information [18], and sorted the topological order with the score. Equations (5) and (6) show the influence score and the mutual information calculation.

$$M_i = \sum_j M(X_i; X_j) \quad \textbf{(5)}$$

$$M(X;Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \quad \textbf{(6)}$$

Sort_nodes_by_topological_order(X, O)
for i=1 **to** n **do**:
    $\pi_i = \Phi$ // initialize the parent set
    $Score[i]=g(x_i, \pi_i)$ // k2 metric score of node x_i
    Continue=true
    repeat
        Z= $\arg\max_j g(x_i, \pi_i \cup \{x_j\})$ **and** $j<i$ **and** $x_j \notin \pi_i$
        $Score'[i]= g(x_i, \pi_i \cup \{x_Z\})$
        if $Score'[i]>Score[i]$ **then**
            $Score[i] = Score'[i]$, $\pi_i = \pi_i \cup \{x_Z\}$
        else *Continue=false*
    until $|\pi_i|<MaxParentNum$ **and** *Continue=true*

Fig. 7. The K2 learning algorithm. *MaxParentNum* is a limitation of the number of parents and π_i is a parent set of x_i and $g()$ is calculated by K2 metric function[17].

4 Experiments

The log data used in this paper include a GPS log, call log, SMS log, picture log, music playing log, device charging log, and weather log obtained from a website. We collected data from three college students (women) with real-world smart phones for 14 days. These users performed subtasks (such as writing activity diaries, shopping, walking and calling) to make the data more substantial.

The final experimental set was selected based on sixteen days because the other data was not collected well enough (especially, the GPS data was often missing). The data set was segmented into units of ten minutes and does not contain redundant data (779 datasets). We defined 48 landmarks shown in the table 2. To discover the modular BNs, we divided the landmarks into four modules based on four categories (Emotion & status, Everyday life, Events, School life). All data set were discretized and the state set of the most landmarks was defined as {yes, no}.

Table 2: Module domain definition of landmark nodes

Domain	Landmarks
Emotion & status	bored, busy, cold, concentration, fret, hungry, joy, overflowing joy, sad, sleepy, surprising, throb, tired, troublesome, with leisure, yearning
Everyday life	eat (Chinese), eat (western), eat (Korean), home activity, meet family, moving, ready to go out, ride a vehicle, run, sleeping, supper, using vehicle, walk
Event	date with my date, drinking(alcohol), eat (tee), eat out, hair-cut, meet friend, meet kin, take a walk, traffic jam, weight-training
School life	employment counsel, extracurricular lecture, go to school, late for school, lecture, school activity, school-club activity, study, test in school

4.1 Comparison of the Discovered BNs

In this section, we describes the test result of the landmark extraction model. We set the *MaxParentNum* parameters (p) as 4 and 8 in the experiment. Fig. 8 shows the modular BNs discovered with (parameter p=8). The average number of nodes, parents, and conditional probability values and the level of complexity are shown in Table 3. The complexities are calculated by Equation (2). We are able to observe the decrement of the complexity of the modular BNs.

Table 3: The comparison of the complexity of learned BNs.

BN	N #	N $\#_{avg}$	P #	P $\#_{avg}$	C
1BN	115	115	298	2.59	O (2.6×10^{6})
mBN	189	47.25	182	1.14	O (9.58×10^{2})

1BN-monolithic BN, mBN-modular BN, N-node, P-parent, C-complexity, # - the number of, $\#_{avg}$ – the average number of.

Table 4 shows the results of the landmark reasoning evaluation. Because the number of training data was small, we used the leave-one-out validation method. We compared the monolithic BN and modular BNs with the parameters p=4 and p=8. The computation of the precision rate is (TP/(TP+FP)), the recall rate is (TP/(TP+FN)), and the hit rate is ((TP+TN)/(TP+TN+FP+FN)).

As shown by the results, the performance of the modular BNs is similar to that of the monolithic BN. This means the proposed method is valuable since it is used to reduce the BN model and increase efficiency.

Table 4: The correctness comparison of extracted landmarks.

BN	*P*	TP	TN	FP	FN	PR	RC	HR
1BN	8	135	35,845	64	1,348	0.678	0.091	0.962
mBN	8	133	35,853	56	1,350	0.704	0.090	0.962
1BN	4	420	35,807	102	1,063	0.805	0.283	0.969
mBN	4	420	35,807	102	1,063	0.805	0.283	0.969

P-parent size parameter, TP - true positive, TN - true negative, FP - false positive, FN - false negative, PR-precision, RC-recall, HR-Hit rate.

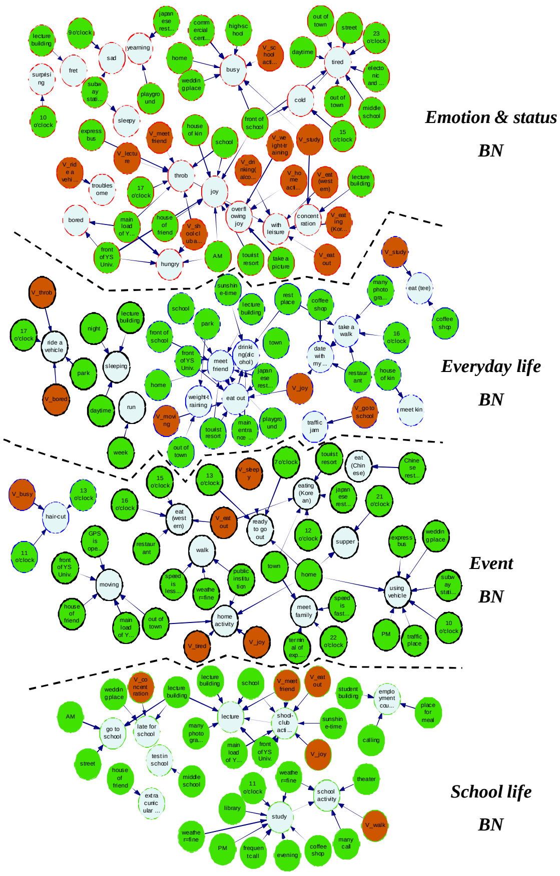

Fig. 8: The BN modules trained with *p*=8. The dark nodes: virtual nodes, the lighter nodes: landmark nodes, the others: evidence nodes.

5 Concluding Remarks

In this paper, we proposed a modular landmark inference model, which was efficient and suitable to mobile environments. We introduced the modularized BN model for efficient operations in mobile environments, and proposed the cooperative inference method by applying the virtual node concept. We discovered the modular BNs automatically from the given training data. In experimental results with mobile life log data, we observed that the proposed method was able to reduce the level of complexity.

However, the recall performance of the BN was not so good because the training data was not enough and much biased to negative sample. In the future, we need to continue research with sufficient real world data. Since we used only K2 algorithm and information theory based ordering method, we have to study other algorithms with our modularization approach.

Acknowledgements

This research was supported by MIC, Korea under ITRC IITA-2008-(C1090-0801-0046).

References

[1] Nokia LifeBlog, http://www.nokia.com/lifeblog.

[2] A. Schmidt, A. Takaluoma, and J. Mntyjrvi, "Context-aware telephony over WAP," *Personal Technologies*, vol. 4, no. 4, pp. 225-229, 2000.

[3] B.P.L. Lo, S. Thiemjarus, and G.-Z. Yang, "Adaptive Bayesian networks for video processing," *Int. Conf. on Image Processing*, vol. 1, no. 1, pp. 889-892, 2003.

[4] M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen, "ContextPhone: A prototyping platform for context-aware mobile applications" *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 51-59, 2005.

[5] A. Krause, A. Smailagic, and D. P. Siewiorek, "Context-aware mobile computing: Learning context-dependent personal preferences from a wearable sensor array," *IEEE Trans. on Mobile Computing*, vol. 5, no. 2, pp. 113-127, 2006.

[6] P. Dourish, "What we talk about when we talk about context," *Personal and Ubiquitous Computing*, vol. 8, no. 1, pp. 19-30, 2004.

[7] E. Horvitz, S. Dumais, and P. Koch. "Learning predictive models of memory landmarks," *CogSci 2004: 26th Annual Meeting of the Cognitive Science Society*, pp. 1-6, 2004.

[8] K.-S. Hwang and S.-B. Cho, "Modular Bayesian Networks for Inferring Landmarks on Mobile Daily Life," *The 19th Australian Joint Conf. on Artificial Intelligence*, pp. 929-933, 2006.

[9] X. Li and Q. Ji, “Active affective state detection and user assistance with dynamic Bayesian networks,” *IEEE Trans. on Systems, Man, and Cybernetics, Part A*, vol. 35, no. 6, pp. 93-105, Jan. 2005.

[10] Y. Zhang and Q. Ji, “Active and dynamic information fusion for multisensor systems with dynamic Bayesian networks,” *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, vol. 36, no. 2, pp. 467-472, Apr. 2006.

[11] Krause, A. Smailagic, and D.P. Siewiorek, "Context-aware mobile computing: Learning context-dependent personal preferences from a wearable sensor array," *IEEE Trans. on Mobile Computing*, vol. 5, no. 2, pp. 113-127, 2006.

[12] M. Marengoni, A. Hanson, S. Zilberstein and E. Riseman, "Decision making and uncertainty management in a 3D reconstruction system," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no.7, pp.852-858, July 2003.

[13] H. Tu, J. Allanach, S. Singh, K. R. Pattipati and P. Willett, "Information integration via hierarchical and hybrid Bayesian networks,“ *IEEE Trans. On Systems, Man, and Cybernetics.Part A: Systems and Humans*, vol. 36, no. 1, pp. 19-33, Jan. 2006.

[14] K.B. Korb and A.E. Nicholson, *Bayesian Artificial Intelligence*, Chapman & Hall/CRC, 2003.

[15] S.L. Lauritzen and D. J. Spiegelhalter, “Local computations with probabilities on graphical structures and their application to expert systems,” *Journal of the Royal Statistical Society*, pp. 157-224, 1988.

[16] V. K. Namasivayam and V. K. Prasanna, “Salable parallel implementation of exact inference in Bayesian networks,” *Int. Conf. on Parallel and Distributed Systems*, pp. 143-150, July 2006.

[17] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, pp. 309-347, 1992.

[18] J. Su, H. Zhang, "Full Bayesian network classifiers," *Proc. of international conference on Machine learning*, pp. 897-904, 2006.