

Optimizing Dynamic Composition of Bayesian Networks for Context Sensing and Inference

Korbinian Frank^{*†} and Matthias Röckl^{*}

^{*}Institute of Communications and Navigation
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
{korbinian.frank | matthias.roeckl}@dlr.de

Tom Pfeifer[‡]

[‡]Telecommunications Software & Systems Group (TSSG)
Waterford Institute of Technology
Waterford, Ireland
t.pfeifer@computer.org

Abstract—Breaking Bayesian Networks for Context Inference from Sensor Networks into smaller Bayeslets is a proven approach for optimizing performance in adaptive resource-constraint ubiquitous computing and networking environments.

Automatic selection and composition of such Bayeslets faces the challenge that the related cost factors (inference time, memory consumption) grow exponentially with the number of components.

The paper discusses optimising approaches to evaluate the added value of using a particular Bayeslet vs. its cost to prune the dynamic composition graph.

Index Terms—Ad hoc & sensor networks; Ubiquitous networking; Bayesian Networks; Reasoning

I. INTRODUCTION

Networked Pervasive Computing Systems frequently collect sensor information from heterogeneous and redundant sources, which is fused and evaluated to form context information [1].

Bayesian algorithms are well-explored in the data fusion layers of such systems [2], moving from full instrumentation with plentiful sensors towards a more limited set of input parameters.

This requires that the algorithms inside mature, allowing to recognise situations from fewer sensing input and to cope with rising expectations of the quality of automated decisions, based on more complex contextual information.

With the complexity of context information rising, some elements of this contextual information must be *inferred* from a variety of simpler sources of input data.

Well-known inference algorithms, that work well on fully resourced computers, present particular challenges to resource-constrained mobile and ubiquitous computing environments.

The demand from the inference algorithms rises faster than the processing capabilities develop. Particular research is required to provide intelligent solutions to keep context inference feasible.

The objective of this paper is to present a novel idea for the dynamic assembly of context inference rules based on Bayeslets. This assembly or composition will happen during run time of the system (dynamically) and has to take into account all relevant but no unnecessary or redundant information. We propose two criteria for the decision upon assembly, *Mutual Information* and the *Value of Information*.

II. BAYESIAN NETWORKS FOR CONTEXT INFERENCE

Bayesian networks (BN) have evolved as a major tool in a wide area of scientific disciplines requiring sound statistical analysis, automated reasoning or exploitation of knowledge hidden in noisy data. Recently they have been presented for context reasoning [3], [4].

A BN is a probabilistic model consisting of a Triplet (V, E, P) , with a set of *Random Variables (RVs)* $V = \{A_1, A_2, \dots, A_n\}$, a set of dependencies $E = \{(A_i, A_j) | i \neq j, A_i, A_j \in V\}$ without directed cycles between these RVs and a *joint probability distribution (JPD)*

$$P(V) = P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^n P(A_i | pa(A_i)), \quad (1)$$

where $pa(A_j) = \{A_i | \forall A_i \in V \wedge (A_i, A_j) \in E\}$ are the *parents of A_j* . Using the interpretation of causal networks [5] in Context Modelling and Inference, every random variable represents a contextual entity (e.g. context attribute, sensor or service), its values the states in which the entity can be. The structure and the probabilities in the CPDs encode the existing knowledge about a certain context. This can come from human expertise or from statistical analysis of contextual knowledge bases. Context Inference is the process of generating knowledge for at least one of the context entities for which no knowledge is available, in terms of Fig. 1 for instance the level of a person’s availability depending on the knowledge about the current time and his current activity. With Bayesian Networks this process is performed by probabilistic inference, which assigns the RVs in question the marginalised and normalised probabilities for all its values.

Probabilistic Inference in BNs however is NP-hard in the number of RVs [6]. In order to keep inference tractable, the factors determining complexity (number of nodes, number of edges, number of values) have to be controlled by a variety of approaches, e.g. dynamic value ranges [7].

III. THE BAYESLET CONCEPT

The approach of *Bayeslets* [8] is supposed to partition full BNs into thematic RV groups, managed and evaluated separately, in order to limit the number of RVs involved in one inference process.

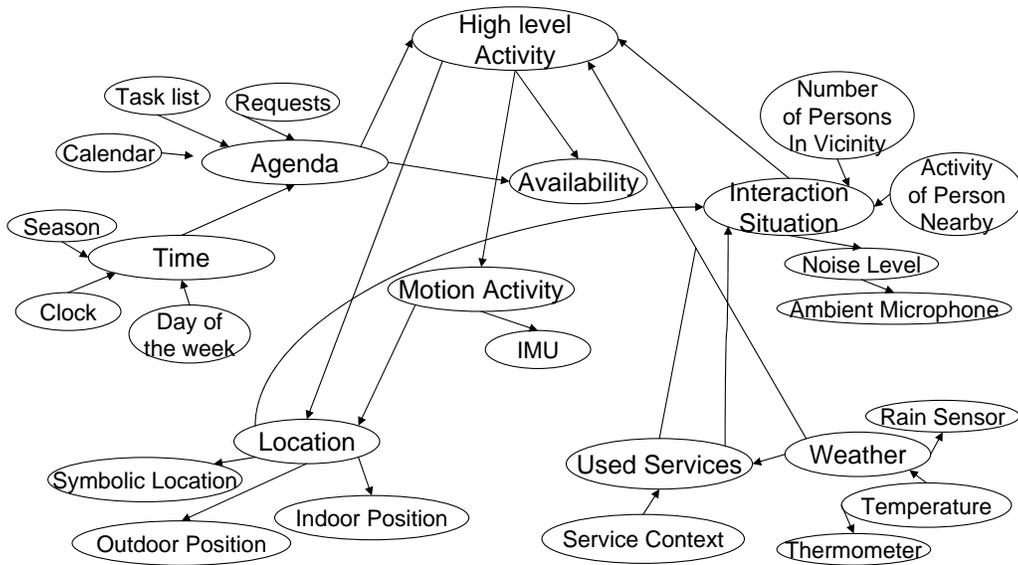


Fig. 1. An example Bayesian Network representing a view on important contextual aspects of a pervasive system user. His or her high level context is influenced by environmental conditions, location, agenda, the presence and interaction level of other persons, as well as his motions and behaviour. All these input factors depend themselves again on other information or sensors, again represented as random variables.

They represent enclosed domains of knowledge (*encapsulation*) that are usable alone, but also in conjunction with a “mother” BN or further Bayeslets as shown in Fig. 2 (*modularity*). They connect to further Bayeslets through pre-defined interfaces, specially tagged random variables that we call *input* or *output nodes*, respectively. In Fig. 2 we see three Bayeslets being connected via their output nodes to a general BN and via the latter also to each other.

Bayeslets that can be plugged to an input node do not need to have exactly the same interface, but exactly the output node demanded by this input node (a form of *polymorphism*). These interface nodes can and have to be addressed precisely such as by a name space with the name of their owner.

Context information pertaining to a user of a Pervasive Computing System may in general encompass a very large range of sensor readings and information such as calendars, as well as such data relating to other people – potentially all represented as random variables in a very large BN. Obviously it is impossible to include all sets of such RVs in a representation used in inference, as the resulting BN would be too large and too complex.

For practical relevance it will be necessary to impose boundaries on which RVs will be used to represent the information “around” her activity. In particular in the area of resource-constrained mobile devices, usage of Bayeslets for context inference is advantageous. They are small and therefore easy to store, they only need little processing power to evaluate, allow for distributed evaluation of inference rules and personalisation of context inference.

IV. DYNAMIC COMPOSITION OF BAYESLETS

As Bayeslets only represent knowledge domains, that may be influenced by each other, for context inference, the related Bayeslets have to be selected and composed to represent all the relevant knowledge. As this relevancy depends on the current situation and the requested output node (there may be more than one in one Bayeslet), composition has to be a dynamic process at run time of the context aware framework.

To answer an inference request, the Bayeslet with the target (output) node is selected and has to be composed with Bayeslets linkable into the input nodes. The available Bayeslets are retrieved from some knowledge base that manages stubs of Bayeslets stored with the Bayeslet owner himself, different system users or provided from third parties. The inference rule composed of Bayeslets has to be minimal, but not neglecting important links. Moreover assembly of

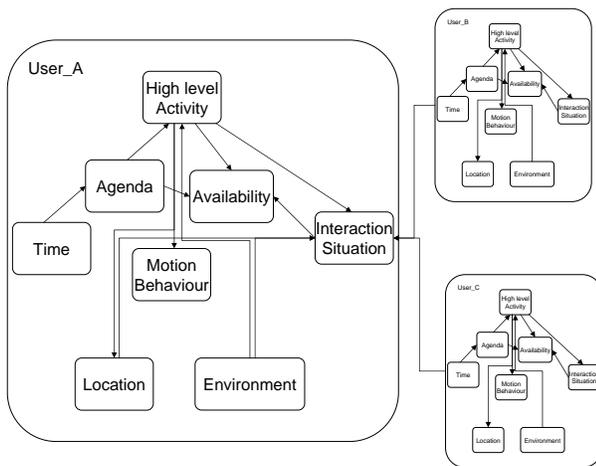


Fig. 2. This is an network of Bayeslets, grouping, fragmenting and extending the situation described in Figure 1. The main concepts are grouped into Bayeslets represented by rectangles with rounded corners, and similar Bayeslets of relevant other persons are linked in.

Bayeslets poses challenges, as it is recursive, because linked-in Bayeslets can allow/require input from other Bayeslets. This bears two main risks:

- **Loops:** If a directed graph composed of Bayeslets would result in having a loop, also the resulting graph of RVs contains a directed loop, which is forbidden in BNs. E.g. breadth-first search can be applied to exclude such cases.
- **Excessive Linking:** if all available interface nodes in Bayeslets are recursively joint with other Bayeslets, this would result in a very large network, possibly until all Bayeslets in the knowledge base are linked and the benefit of Bayeslets is lost, leading to unacceptable high delays.

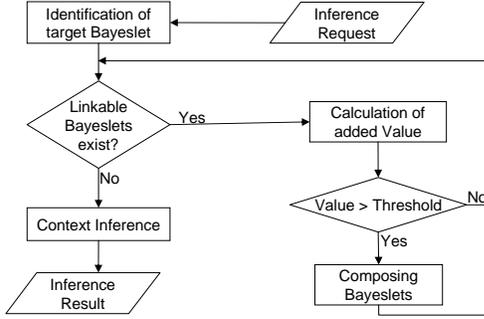


Fig. 3. The Process of Composing Bayeslets to a Complete Inference Rule

Therefore, we propose to join only those Bayeslets to the requested one that significantly add value in comparison to its costs, i.e. the evaluation and memory costs caused by the increased network. The selection of such Bayeslets has to be performed in a new step, preceding the inference itself as shown in Fig. 3.

V. COMPOSITION CRITERIA

This section presents two criteria to determine the benefit of an added Bayeslet. The outcome of such a criterion has to be comparable across different domains and with a threshold value. Furthermore, it has to take into account both, the costs induced by joining that Bayeslet (here in particular inference time) and value added, i.e. the reduced uncertainty in the inference outcome.

A. Mutual Information

A common measure for uncertainty of a random variable is the entropy as it has been identified by Shannon [9]. $H(X)$ for a discrete RV X with a probability distribution $P(X)$ and the expectation E_X is defined by:

$$\begin{aligned} H(X) &= - \sum_{x \in X} P(x) \log P(x) = \\ &= -E_X \log P(X) = E_X \frac{1}{\log P(X)}. \end{aligned} \quad (2)$$

A uniform distribution has the maximum entropy because all states are equiprobable and does not add knowledge. On the other extreme the minimum entropy is zero which reflects a random variable with a certain outcome. Therefore

$0 \leq H(X) \leq \log n$ always holds [10].

Additionally to the entropy definition for prior probability distributions, the entropy can also be regarded as conditional entropy “of a random variable given another as the expected value of the entropies of the conditional distributions, averaged over the conditioning random variable” [11]:

$$H(X|Y) = \sum_{y \in Y} P(y) H(X|y) = -E_{X,Y} \log P(X|Y) \quad (3)$$

Thus, the difference in the entropy of X with and without the knowledge of Y can be calculated by:

$$I(X : Y) = H(X) - H(X|Y) = E \log \frac{P(X|Y)}{P(X)} \quad (4)$$

$I(X : Y)$ is called *mutual information*. It is the reduction in the uncertainty of one random variable due to the knowledge of another one [9], [11]. A general rule is that the mutual information is non-negative and zero, iff X and Y are unconditionally independent.

Thus, additional information never increases the entropy. Although the logarithm of the relevance quotient can be negative, its expectation is always nonnegative.

Analogously to Röckl’s information dissemination in [10], the mutual information between input Bayeslets has to be made comparable among different input Bayeslets by normalisation with $H_{max}(X)$ and fair by consideration of the costs, which results in the *net normalized mutual information (NetNI)*:

$$NetNI(X : Y) = \frac{I(X : Y)}{\log |X|} - C(Y) \quad (5)$$

In our case X represents a Bayeslet and $NetNI(X : Y)$ will be calculated to decide whether to plug in Bayeslet Y . The respective output nodes will be used to represent the Bayeslets, where the probabilities of Y are marginalised over its Bayeslets without evidence. This information is static as well as its plugging costs. The measure for the costs has to be normalized to the interval $[0; 1]$, in order to stay comparable to the information measure. The following is an example for a cost measure for CPU costs, i.e. inference time:

$$C(X) = 1 - \frac{1}{n_X^a + b}, \quad (6)$$

with $a \geq 1$, $b \geq 0$ constants and n_X the size of the biggest clique in the junction tree [12] of the Bayeslet containing X . Other measures can include memory consumption and CPU costs, depending on the needs of the implementing system.

B. Value of Information

A drawback of the entropy based approach is, that it only depends on the distribution of probabilities of RVs and therefore not on the currently available evidence. To allow for that, we can make use of *Decision Networks* [13]. A Decision Network is a kind of BN that introduces two new types of nodes, *decision nodes* and *utility or value functions*, where the value of each decision variable [...] is imposed from the outside to meet some optimization objective [14]. Within a decision network, decision variables are depicted

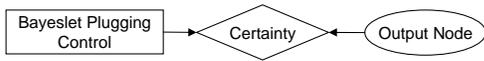


Fig. 4. A Decision Network with one decision node, one utility node and one standard node representing a random variable.

as rectangular boxes and value functions as diamond shaped boxes as shown in Fig. 4.

With Decision Networks, the *Maximum Expected Utility (MEU)* can be calculated given evidence ϵ .

$$MEU(A|\epsilon) = \max_{a \in A} EU(a|\epsilon) = \max_{a \in A} \sum_{x \in X} U(x)P(x|a, \epsilon), \quad (7)$$

where A represents the possible actions in the decision node and $U : X \rightarrow \mathbb{R}$ the value function.

The information gain by the acquisition of evidence ϵ_e from a linkable, *external* Bayeslet is the difference between the *MEU* without additional evidence $MEU(A|\epsilon)$ and $MEU(A|\epsilon, \epsilon_e)$ with additional evidence. It is expressed by the *Value of information (VoI)* [15]. Since for this calculation the state of ϵ_e is still unknown, an expectation over the states of ϵ_e with its posterior probability $P(\epsilon_e|\epsilon)$ given the already observed evidence ϵ has to be included, the *Expected Maximum Expected Utility (EMEU)*:

$$\begin{aligned} VoI(\epsilon_e|\epsilon) &= EMEU(A|\epsilon; \epsilon_e) - MEU(A|\epsilon) = \\ &= \sum_{\epsilon_e, i \in \epsilon_e} P(\epsilon_e, i|\epsilon) MEU(A|\epsilon; \epsilon_e) - MEU(A|\epsilon) \end{aligned} \quad (8)$$

The *VoI* is non-negative as additional information will never reduce the *MEU* [10]. Its value is zero, i.e. $EMEU = MEU$, iff the additional evidence does not provide any new information. This is the case if the evidence is (conditionally) independent of the situation represented by Bayeslet X .

Evidence ϵ includes evidence from the target Bayeslet and already joint Bayeslets. ϵ_e represents potential information from an external Bayeslet.

Analogously to the *NetNI*, the *Net Value of Information (NetVoI)* is the *VoI* which can be achieved by acquiring additional evidence minus the costs for the acquisition:

$$NetVoI(\epsilon_e|\epsilon) = VoI(\epsilon_e|\epsilon) - C(\epsilon_e) \quad (9)$$

$C(\epsilon_e)$ has the same function as $C(Y)$ for $NetNI(X : Y)$. It represents optional evidence in the *VoI* approach. C is here not bound to $[0; 1]$, as U is neither. Both summands have to be harmonized however to reach the desired result.

C. Theoretical Evaluation

NetVoI fits the requirements better than *NetNI*, as it takes into account all available information at request time. The design of the utility function to represent the aim of reducing uncertainty in the target random variable of the Bayeslet is challenging. Adapting automatically such a utility function for machine learnt Bayeslets will be even more complicated. The final decision for one criterion has to be based on a realistic evaluation, taking into account the time for calculating *NetVoI* and *NetNI* for all linkable Bayeslets.

VI. CONCLUSION AND FUTURE WORK

This paper has presented a process for the dynamic composition of Bayesian Network based context inference rules taking into account the current requirements of an inference request. In order to keep the inference rule minimal, Bayeslets are only joint, if a criterion measuring its added value exceeds a certain threshold. Candidates for such criteria are the Net Normalized Mutual Information *NetNI* and the Net Value of Information *NetVoI*.

The context and location test bed at DLR and the TSSG laboratories in Waterford for the Persist project will allow for information collection in a large scale pervasive personal smart space environment. Its evaluation will help us to further improve Bayesian context inference in terms of efficiency and outcome quality.

ACKNOWLEDGMENTS

This work has received funding from the EC's Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 215098 of the *Persist* Collaborative Project and from the Irish HEA through the PRTL cycle 4 project "Serving Society: Management of Future Communication Networks and Services".

REFERENCES

- [1] T. Pfeifer, "Redundant positioning architecture," *Computer Communications*, Elsevier Science Publishers B.V. North-Holland, vol. 28, no. 13, pp. 1575–1585, August 2005.
- [2] D. Wagner and R. Wattenhofer, Eds., *Algorithms for Sensor and Ad Hoc Networks*, *Advanced Lectures*, ser. Lecture Notes in Computer Science, vol. 4621. Springer, 2007.
- [3] T. Gu, H. K. Pung, and D. Q. Zhang, "A bayesian approach for dealing with uncertain contexts," in *Proc. of the 2nd International Conference on Pervasive Computing (Pervasive '04)*, April 2004.
- [4] B. Beamon, "Evaluation of first order bayesian networks for context modeling & reasoning," in *Proceedings of Third Annual Google Ph.D. Forum on Pervasive Computing and Communications*, 2010.
- [5] J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.
- [6] G. F. Cooper, "Probabilistic inference using belief networks is NP-hard," Medical Computer Science Group, Knowledge Systems Laboratory, Stanford University, Stanford, CA, Tech. Rep. KSL-87-27, 1987.
- [7] K. Frank, S. Fortes Rodriguez, P. Robertson, and R. Barco Moreno, "Faster bayesian context inference by using dynamic value ranges," in *Proceedings of Workshop on Information Quality and Quality of Service for Pervasive Computing*, 2010.
- [8] K. Frank, M. Röckl, and P. Robertson, "The Bayeslet concept for modular context inference," in *Proceedings of Second International Conference on UBICOMM08*, J. Lloret Mauri et al., Ed. IEEE CPS, May 2008, pp. 96–101. [Online]. Available: <http://elib.dlr.de/54515/>
- [9] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, 1948.
- [10] M. Röckl, "Cooperative situation awareness in transportation," PhD thesis, Faculty of Mathematics, Computer Science and Physics of the University of Innsbruck, 2009.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley, 1991.
- [12] C. Huang and A. Darwiche, "Inference in belief networks: A procedural guide," *International Journal of Approximate Reasoning*, vol. 15, no. 3, pp. 225–263, 1996.
- [13] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1995.
- [14] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann, 1988.
- [15] R. A. Howard, "Information value theory," *IEEE Transactions on System Science and Cybernetics*, vol. 2, no. 1, pp. 22–26, 1966.