Dynamic Bayesian Network for Time-Dependent Classification Problems in Robotics

Cristiano Premebida, Francisco A. A. Souza and
Diego R. Faria

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.70059

Abstract

This chapter discusses the use of dynamic Bayesian networks (DBNs) for time-dependent classification problems in mobile robotics, where Bayesian inference is used to infer the class, or category of interest, given the observed data and prior knowledge. Formulating the DBN as a time-dependent classification problem, and by making some assumptions, a general expression for a DBN is given in terms of classifier priors and likelihoods through the time steps. Since multi-class problems are addressed, and because of the number of time slices in the model, additive smoothing is used to prevent the values of priors from being close to zero. To demonstrate the effectiveness of DBN in time-dependent classification problems, some experimental results are reported regarding semantic place recognition and daily-activity classification.

Keywords: dynamic Bayesian network, Bayesian inference, probabilistic classification, mobile robotics, social robotics

1. Introduction

Bayesian inference finds applications in many areas of engineering, and mobile robotics is not an exception. When time is a variable to be considered, the dynamic Bayesian network (DBN) [1–5] is a powerful approach to be considered. Due to its graphical representation and modelling versatility, DBN facilitates the problem-solving process in probabilistic time-dependent applications. Therefore, DBNs provide an effective way to model time-based (dynamic) probabilistic problems and also enable a very suitable and intuitive representation by means of a graph-based tree.

Depending on the structure of the DBN, the joint probabilistic distribution that governs a given system can be decomposed by a tractable product of probabilities, where the conditional terms only depend on their directly linked nodes. This chapter concentrates on inference problems using DBN where the variable to be inferred from a feature vector (data) represents a set of semantic classes $\mathcal{C} = \left\{ c_1, c_2, ..., c_{nc} \right\}$ or categories, in the context of intelligent perception systems for mobile robotics applications. Namely, we will address problems where $\mathcal{C}$ denotes semantic places in a given indoor environment [6, 7] e.g. $\mathcal{C} = \{$ 'corridor', 'office', ..., 'kitchen'$\}$ and also when the classes of interest are daily-live activities $\mathscr{C} = \{$'drinking', 'talking', ..., 'walking'$\}$ [8, 9].

The principle of Bayesian inference basically depends on two elements: the prior and the likelihood; in practical problems, the evidence probability acts 'only' as a normalization to guarantees that the posterior sums to one. In this chapter, we will deal with the problems of the classical Bayesian form $posterior \propto likelihood \cdot prior$, but the incorporation of (past) time will be explicitly modelled in a discrete-time basis, and the *past* information is assumed to be contained in the prior probabilities. Inference will be considered beyond the first-order Markov assumption, which means that a DBN with a finite number of time slices (T) will be addressed. Current time step t and previous/past time steps will be considered in the formulation of the DBN; thus, the time interval is $\{t, t-1, ..., t-T\}$.

The observed data enters the DBN in the form of a vector of features X calculated from sensory data; examples of sensors are laser scanners (or 2D Lidar) and RGB-D camera, as shown in **Figure 1**. Later, in the formulation of the DBN, we will consider that the feature vector at a given time step (X^t) is conditionally independent of previous time steps; therefore, $P(X^t \mid X^{t-1}) = P(X^t)$.

The use of Bayesian inference in mobile robotics for purpose of localization, simultaneous localization and mapping (SLAM), object detection, path planning and navigation, has been addressed in many scientific works; see Ref. [10] for a review. The majority of those applications involve stochastic filtering, such as Kalman filter (KF), particle filter (PF), Monte Carlo techniques and hidden Markov model (HMM) [11, 12]. However, when the parameter of interest has to be inferred from multidimensional feature vectors (e.g. feature vectors with hundreds of elements) and also when the distribution that the observed data were drawn is not known (in unseen/knew or testing scenarios) then, a DBN can be used to handle such complex problems. In robotics, semantic place classification [6, 7] and activity recognition [8, 9] are examples of such problems and belong to the research area of pattern recognition. For

Figure 1. Sensors commonly used in mobile robotics for perception systems.

these application cases, the class-conditional probabilities (or likelihoods) can be modelled using machine learning techniques, for example, naive Bayes classifier (NBC), support vector machines (SVMs) and artificial neural networks (ANNs) [13, 14].

The remainder of this chapter is organized as follows: a brief review of the DBN is given in Section 2. Section 3 addresses inference in DBN, formulated for purposes of pattern recognition in robotics, followed by the use of additive smoothing on the prior distributions. In Section 4, experimental results on semantic place classifications and activity recognition are presented. Finally, Section 5 presents our conclusions.

2. Preliminaries on DBN

Basically, a DBN is used to express the joint probability of events that characterizes a time-based (dynamic) system, where the relationships between events are expressed by conditional probabilities. Given evidence (observations) about events of the DBN, and prior probabilities, statistical inference is accomplished using the Bayes theorem. Inference in pattern recognition applications is the process of estimating the probability of the classes/categories given the observations, the class-conditional probabilities, and the priors [15, 16]. When time is involved, usually the system is assumed to evolve according to the first-order Markov assumption and, as consequence, a single time slice is considered.

In this chapter, we address DBN structures with more than one time slice. Moreover, the conditional probabilities of the DBN will be modelled by supervised machine learning techniques (also known as classifier or classification method). Two case studies will be particularly discussed: activity recognition for human-robot interaction and semantic place classification for mobile robotics navigation.

The observed data variable, denoted by $X = \left\{ X_1, ..., X_{nx} \right\}$, enters into the DBN in the form of conditional probabilities $P(X|C)$, where the values of X are feature vectors. To give an idea of the dimensionality of X, in semantic place classification [6], the number of features can be $nx = 50$, while in activity recognition we have 51 features [8]. Given such dimensionalities, which can be even higher, it becomes infeasible to estimate the probability distribution that characterizes $P(X|C)$ without the use of advanced algorithms. Although a simple Naïve Bayes classifier can be incorporated in a DBN to model $P(X|C)$, more powerful solutions, such as the ensemble of classifiers in the DBMM approach introduced in Ref. [8], tend to achieve higher classification performance.

In summary, DBN is a direct acyclic graph (DAG) that consists of a finite set of events (the nodes or vertices) connected through edges (or arcs) that model the dependencies among the events and also the time variable. Here, the nodes are given by the variables $\{X, C\}$, and the dynamic (time-based) behaviour of the BDN is considered to be governed by the current time t and by a finite set of previous time slices $\{t-1, t-2, ..., t-T\}$. So, future time slices will be not considered. **Figure 2** shows the structure of the DBN, with $T + 1$ time slices, that will be considered in the problem formulation presented in the sequel.
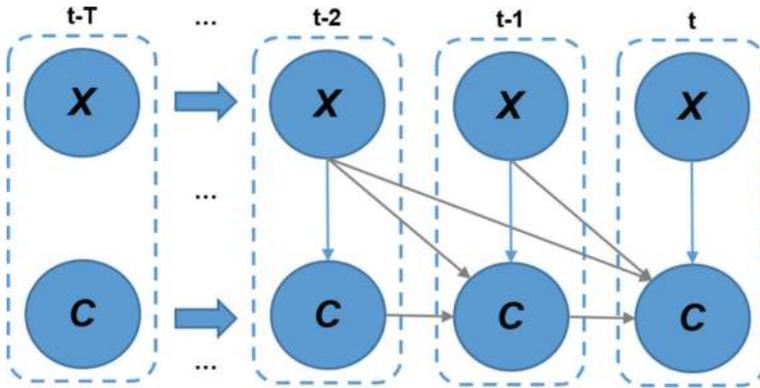
Figure 2. An example of a DBN with $T + 1$ time slices and two nodes $\{C, X\}$.

3. Inference with DBN

The problem is formulated by considering $P(X^t, X^{t-1}, \ldots, X^{t-T}, C^t, C^{t-1}, \ldots, C^{t-T})$ i.e., the joint distribution of the nodes over the time up to T. The goal is to infer the current-time value of the class C^t given the data $X^{t:t-T} = \{X^t, X^{t-1}, \ldots, X^{t-T}\}$ and the prior knowledge of the class, which is attained by the a-posteriori probability $P(C^t \mid C^{t-1:t-T}, X^{t:t-T})$. The superscript notation denotes the set of values over a time interval: $\{t : t - T\} = \{t, t - 1, t - 2, \ldots, t - T\}$.

The simplest case is for a single time slice where the posterior reduces to $P(C^t \mid X^t) \propto P(X^t \mid C^t)P(C^t)$. For two time slices, we have

$$P(C^t \mid C^{t-1}, X^{t:t-1}) \propto P(X^t \mid X^{t-1}, C^{t:t-1})P(X^{t-1} \mid C^{t:t-1})P(C^t \mid C^{t-1})P(C^{t-1}). \tag{1}$$

As the number of time slices increases, the problem of inferring the class becomes more complex; therefore, some assumptions can be made in order to find a tractable solution. As a first assumption, let the nodes be independent of later (subsequent in time) nodes. As a consequence, and taking as the example for $T = 1$, the probability $P(X^{t-1} \mid C^{t:t-1}) = P(X^{t-1} \mid C^{t-1})$ that is, the node X^{t-1} does not depend on the node C^t which is after a time-slice. The second assumption, more strong, is that the feature-vector node X is independent for all time slices hence, and following the previous example, $P(X^t \mid X^{t-1}, C^{t:t-1})$ becomes $P(X^t \mid C^{t:t-1})$. Given these two assumptions, we can state the general problem of calculating the posterior probability of a DBN with $T + 1$ time slices by the expression

$$P(C^t \mid C^{t-1:t-T}, X^{t:t-T}) = \frac{1}{\beta} \prod_{k=t}^{t-T} \left\{ P(X^k \mid C^k)P(C^k) \right\}. \tag{2}$$

where β is the scale (normalization) factor to guarantee that the values of the a-posteriori sum to one. The class-conditional probabilities $P(X^k \mid C^k)$ come from a supervised classifier or from an ensemble of classifiers as in Ref. [8], while $P(C^k)$ assumes the value of the previous posterior probability; thus, $P(C^t) \leftarrow posterior^{t-1}$.

This strategy for 'updating' the values of the prior by taking the values of previous posteriors is a very common and effective technique used in Bayesian sequential systems. The steps involved in the calculation of the posterior probability, as expressed in Eq. (2), are illustrated in **Figure 3**.

Selection of the class-conditional model to express $P(X|C)$ is an important part of the approach and can be achieved by well-known probabilistic machine learning methods. Although generative methods (e.g. Naïve Bayes, GMM and HMM) provide direct probabilistic interpretation and, therefore, constitute appropriate choices, discriminative methods (e.g. SVM, random forest and ANN) tend to have better classification performance. However, to be a suitable model, a given discriminative method has to be of a probabilistic form; this implies, at least, that the outcomes from the classifier sum to one. A more advanced method can be used to model $P(X|C)$ in a DBN, as the dynamic Bayesian mixture model (DBMM) [8], where a mixture of n classifiers is used to model the conditional probability which assumes the form $P(X|C) = \sum_{j=1}^{n} \omega_j P (X|C)_j$, $j = 1, ..., n$; where ω_j are the weighting parameters and $P(X|C)_j$ are the probabilities from the classifiers. Further details are provided in Ref. [6].

The product of likelihoods and priors, in the expression of the a-posteriori Eq. (2), has the consequence of penalizing the classes that are less likely to occur. In other words, the classes with low probability, i.e. close to zero, will have an even more low values of posterior; this effect is intensified as the number of time slices increases. Because the priors are recursively assigned by assuming the values of the previous posteriors, we suggest to use additive smoothing to avoid values of priors to be very close to zero.

Additive smoothing, also called Lidstone smoothing, adds a term (α) to the prior distribution and can be expressed as
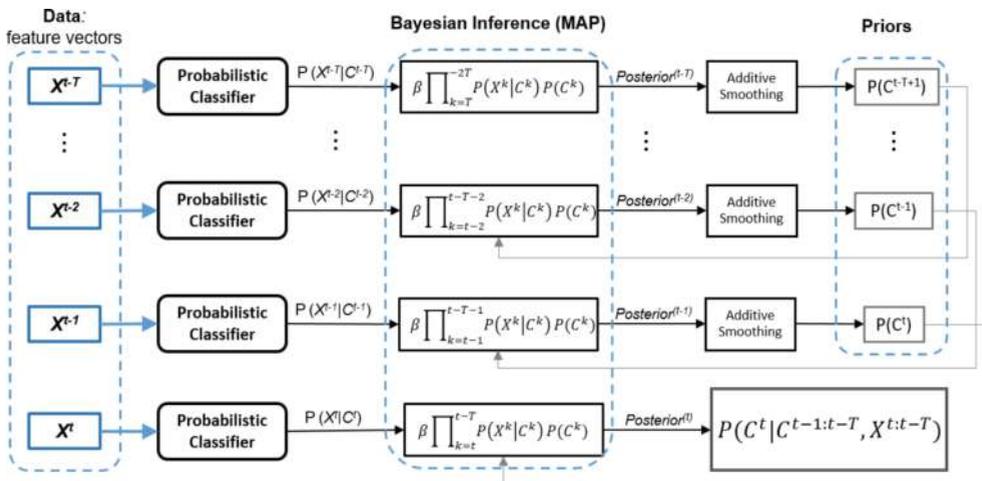

Figure 3. This figure illustrates the DBN, with $T + 1$ time slices, as formulated according to the assumptions presented in Section 3. The product of likelihoods and priors, over the time interval $[t − T, t]$, becomes the posterior probability as expressed in Eq. (2).

$$\hat{P}(C_i) = \frac{P(C_i) + \alpha}{1 + \alpha \cdot (nc)}, \ i = 1, \ldots, nc \tag{3}$$

where α is the additive smoothing factor and nc is the number of classes. The influence of α on the smoothed prior $\hat{P}(C_i)$ has to be such that the values of $\hat{P}(C_i)$ are greater than zero ($\hat{P}(C_i) > 0, \forall i$) and, moreover, the prior distribution $\hat{P}(C_i)$ should be consistent (the values of $\hat{P}(C_i)$ must of course sum to one). A practical range is $0 < \alpha < 0.1$.

Figure 4 provides an example of the impact of α on a given prior, with values of α equal to {0, 0.01, 0.05 and 0.1}. As the value of α increases, the prior distribution tends to lose its initial definiteness due to the uniform 'bias' introduced by α. In the example shown in **Figure 4**, we have considered a five-class case ($nc = 5$).
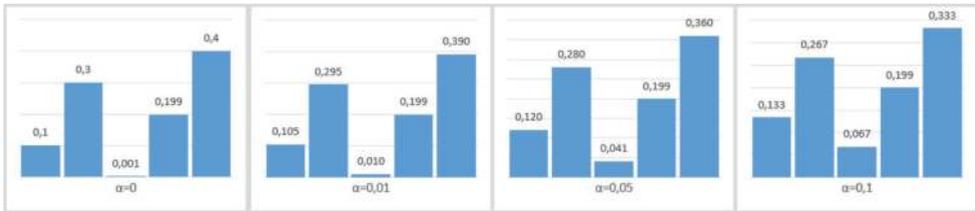

Figure 4. An example of the influence of the additive factor (α) in a given $P(C_i)$, $i = 1, \ldots, 5$.

4. Experiments on classification: mobile robotics case studies

In order to demonstrate the use of the DBN as formulated above, we will consider two classification problems that find applications in mobile robotics: semantic place recognition [6] and activity classification [8].

4.1. Semantic place recognition

Figure 5 illustrates a probabilistic system for semantic place recognition where data comes from a laser scanner sensor. In a practical application, the sensor is mounted on-board a mobile robot [6, 7]. Based on **Figure 5**, we can make a direct correspondence with the DBN discussed above by verifying that the feature vector is X, the probabilistic classifier outputs the class-conditional probability $P(X|C)$ and the priors transmit the time-based information through the network.

As an example of the DBN application in semantic place classification, let us report some results from Ref. [6], where a DBN was applied on the image database for robot localization (IDOL) dataset: available at http://www.cas.kth.se/IDOL/. In this context, the problem of semantic place classification can be stated as follows: 'given a set of features, calculated on
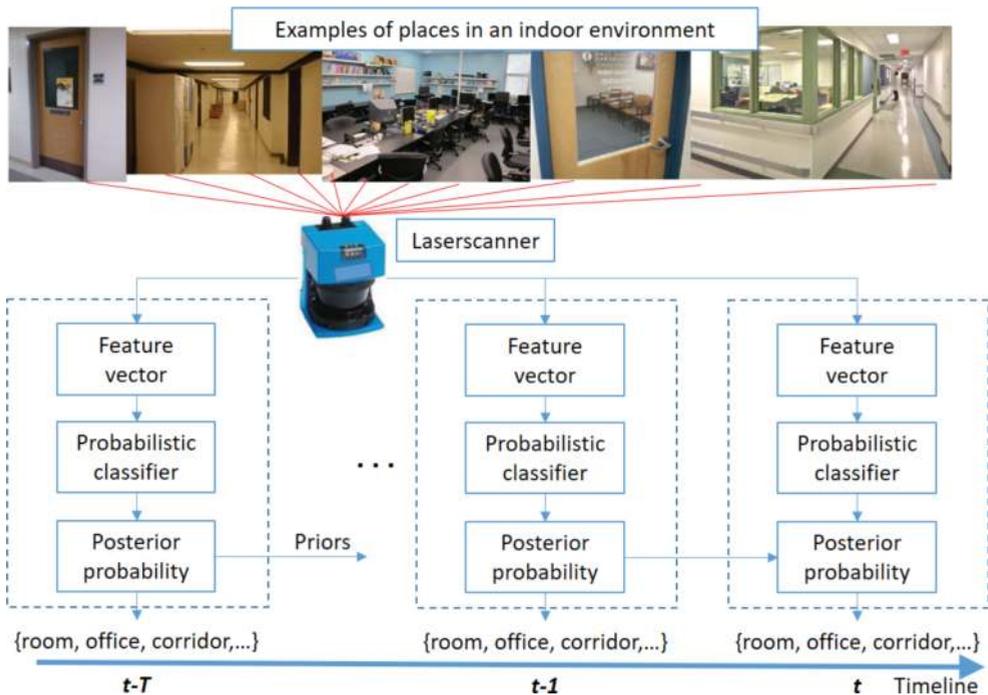
Figure 5. Illustration for a time-dependent probabilistic system applied in semantic place recognition. In this system, data obtained from a laser scanner.

data from laser scanner sensors (installed on-board a mobile robot), determine the semantic robot location ('corridor', 'room', 'office', etc) by using a classification method'. The experiments in Ref. [6] use a mixture of classifiers to model the class-conditional probability in the DBN; such approach is called DBMM [8].

Figure 6 shows recognition results in a sequence of nine frames from the IDOL dataset, where the first row depicts images of indoor places as captured by a camera mounted on-board a mobile robot. The second row provides classification results without time slices (i.e. time-base prior probabilities are not incorporated into the DBN), and the subsequent rows show classification probabilities for a DBN with time-slices up to three. In the figure, the vertical line (in red) indicates the transition between classes: from the class 'kitchen' (KT) to the class 'corridor' (CR).

4.2. Activity classification

In the case of the activity classification problem described here, the objective is to classify the human's daily activity based on spatiotemporal skeleton-based features. In such a case, mobile robots mounted with appropriated cameras can make use of such classification models to
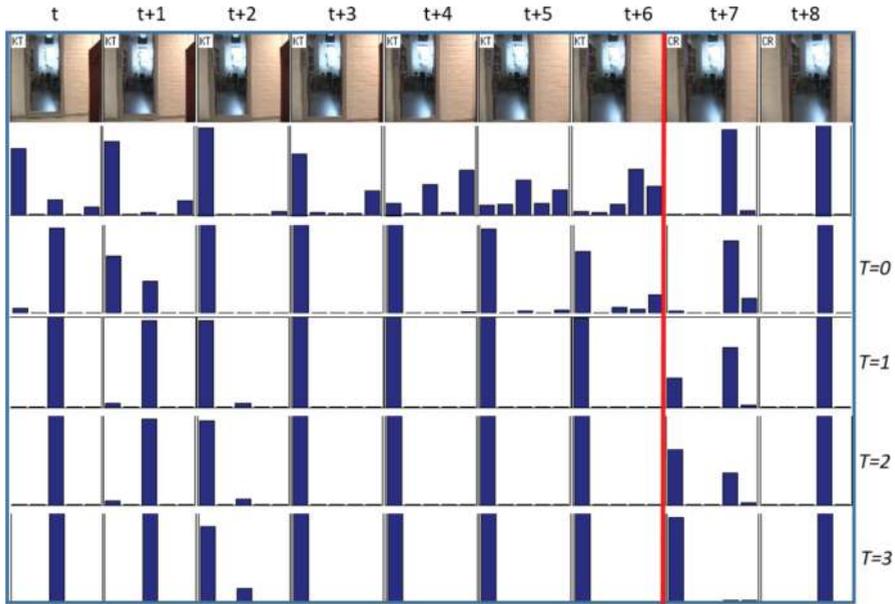
Figure 6. Classification results on a five-class semantic place recognition problem, extracted from reference [6], using a DBN with mixture models of three classifiers (DBMM [6, 8]).

improve the quality of life of, for example, old-age people, by assisting them in their daily life or detecting anomalous situations. Similar to semantic place recognition problem, the activity classification problem can also be seen as a time-dependent probabilistic system, where the feature vector X is the skeleton-based features. From Ref. [8], we report some results on the activity classification.

Figure 7 exhibits an activity classification framework, based on Ref. [8], which uses a DBN with mixture models (the DBMM approach as previously described in the semantic place classification problem), where the data is acquired by using an RGB-D sensor, followed by the skeleton detection step and the feature extraction process, where the latter is based on geometrical features. From the training stage, global weights are computed using an uncertainty measure (e.g. entropy) as a confidence level for each base classifier based on their performance on the training set. During the test, given the input data (i.e. skeleton features for the current activity), base classifiers are used and merged as mixture models with time slices (using previous time instant classification) to reinforce the current classification.

The well-known dataset for activity recognition Cornell Activity Dataset (CAD60) [9, 17] was used to evaluate the proposed framework in Refs. [8, 18]. The CAD-60 dataset comprises video sequences and skeleton data of human daily activities acquired from a RGB-D sensor. There are 12 human' daily activities performed by four different subjects (two male and two female, one of them being left-handed) grouped in five different environments: office, kitchen,
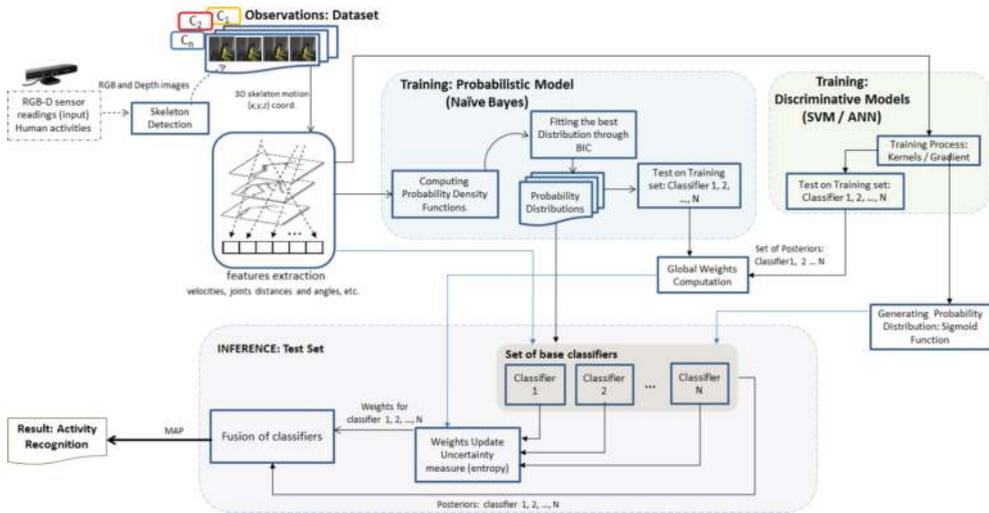
Figure 7. Illustration for a time-dependent probabilistic system applied to activity classification. In this system, data obtained from a RGB-D camera, which provides the spatiotemporal skeleton-based features.

bedroom, bathroom and living room. Additionally, the CAD-60 dataset has two more activities (random movements and still), which are used for classification assessment on test sets, in order to evaluate precision and generalization capacity of the approaches since these activities encompass similar movements to some other activities. We have adopted the same strategy described in Ref. [17], so that we present the classification results in terms of precision (Prec) and recall (Rec) for each scenario. The evaluation criterion was carried out using leave-one-out cross-validation. The idea is to verify the generalization capacity of the classifier by using the strategy of 'new person', i.e. learning from different persons and testing with an unseen person. The classification is made frame-by-frame to account for the accuracy of the frames correctly classified.

Results show the DBMM approach obtained better classification performance compared to other state-of-the-art methods presented in the ranked table in Ref. [17]. The overall results were precision: 94.83%; recall: 94.74% and accuracy: 94.74%. **Figure 8** presents the classification performance (i.e. precision and recall) for the 'new person' tested in each scenario. For comparison purposes, **Table 1** summarizes the results in terms of accuracy of state-of-the-art single classifiers and a simple averaged ensemble compared with the proposed DBMM for the bedroom (scenario with more misclassification), showing that our approach outperforms other classifiers. The classification performance in terms of overall accuracy, precision and recall has shown that our proposed framework outperforms state-of-the-art methods that use the same datasets [17].

In this section, we have shown the DBMM [8, 18] performance using an offline dataset. Additionally, further tests using a mobile platform with an RGB-D sensor on-board running on-the-fly in an assisted living context was also successfully validated with accuracy above

Figure 8. Performance on the CAD-60 ('new person'). Results are reported in terms of precision (Prec) and recall (Rec) and an average (AV) per scenario. Overall AV: precision 94.83%; recall: 94.74%. Activities in (a): Act1—rinsing water; Act2—brushing teeth; Act3—wearing lens; Act4—random + still; activities in (b): Act1—talking on phone; Act2—drinking water; Act3—opening container; Act4—random + still; activities in (c): Act1—talking on phone; Act2—drinking water; Act3—talking on coach; Act4—relaxing on coach; Act5—random + still; activities in (d) Act1—drinking water; Act2—cooking chopping; Act3—cooking stirring; Act4—opening container; Act5—random + still; activities in (e): Act1—talking on phone; Act2—writing on whiteboard; Act3—drinking water; Act4—working on computer; Act5—random + still.

Location	Activity	Bayes	ANN	SVM	AV	DBMM
Bedroom	1	79.90%	74.70%	74.90%	76.50%	84.10%
	2	72.70%	76.60%	81.40%	76.90%	86.40%
	3	79.60%	91.10%	93.10%	87.90%	98.30%
	4	65.70%	93.50%	92.60%	83.90%	97.40%
	Average	**74.48%**	**83.98%**	**85.50%**	**81.30%**	**91.55%**

Activity: 1—talk.on phone, 2—drink.water, 3—open.container, 4—random + still.

Table 1. Results in terms of accuracy on the bedroom scenario of the CAD-60 dataset ('new person') using single classifiers, a simple averaged ensemble (AV) and the DBMM.

90%, as reported in Ref. [18]. More details about the DBMM using a mobile robot for activity recognition and a video showing the classification performance can be found in Ref. [18].

5. Conclusion

In this chapter, the authors have presented a DBN formulation for classification of time-dependent problems together with experimental results on applications of two mobile robots. The first one regarding the semantic place classification and the second one based on activity classification. In both formulations, the DBN was used as basis to compose the DBMM [6, 8, 18], a more complex structure used to handle more complex scenarios. In both applications, the DBMM has shown to be a powerful choice in modelling of time-dependent scenarios.

When it comes to semantic place classification, the model could detect classes' transitions during the robot navigation, thanks to the different time slices (i.e. higher than 2) and the additive smoothing used in the model. In the case of activity recognition, since the activities in the dataset do not have classes' transitions, i.e. only one activity is performed during a task, in this case, a simple version of the DBMM using only one time slice is enough to correct classify all activities. For real-time applications using a mobile robot and in accordance with experimental results reported in Ref. [6], it is suggested to use more than two time slices in the mode.

Author details

Cristiano Premebida[1]*, Francisco A. A. Souza[1] and Diego R. Faria[2]

*Address all correspondence to: cpremebida@isr.uc.pt

1 Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal

2 School of Engineering & Applied Science, Aston University, Birmingham, UK

References

[1] Friedman N, Murphy K, Russell S. Learning the structure of dynamic probabilistic networks. In: Proceeding of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI'98); 24-26 July 1998; Madison, Wisconsin. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1998

[2] Korb KB, Nicholson AE. Bayesian Artificial Intelligence. 2nd ed. Boca Raton, FL: CRC Press, Inc. 2010

[3] Koller D, Friedman N. Probability graphical models: Principles and techniques. In: Adaptive Computation and Machine Learning. The MIT Press; Cambridge, MA, USA; 2009

[4]   Murphy KP. Dynamic Bayesian networks: Representation, inference and learning. Ph.D. Dissertation. University of California, Berkeley; 2002

[5]   Mihajlovic V, Petkovic M. Dynamic Bayesian Networks: A State of the Art. Technical Report, Computer Science Department, University of Twente, Netherlands; 2001

[6]   Premebida C, Faria D, Nunes U. Dynamic Bayesian network for semantic place classification in mobile robotics. Autonomous Robots (AURO), Springer; 2016

[7]   Rottmann A, Mozos OM, Stachniss C, Burgard W. Semantic place classification of indoor environments with mobile robots using boosting. In: Proceeding of the 20th National Conference on Artificial Intelligence (AAAI'05); 9-13 July 2005; Pittsburgh, Pennsylvania: AAAI Press; 2005

[8]   Faria DR, Premebida C, Nunes C. A probabilistic approach for human everyday activities recognition using body motion from RGB-D images. In: Proceedings of the IEEE RO-MAN'14: International Symposium on Robot and Human Interactive Communication; 25-29 August. 2014; Edinburgh, UK. IEEE; Cambridge, MA, USA; 2014

[9]   Sung J, Ponce C, Selman B, Saxena A. Unstructured human activity detection from RGBD images. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Saint Paul, MN, New York, NY, USA, May 2012; pp. 842-849

[10] Thrun S, Burgard W, Fox D. Probabilistic Robotics. MIT Press; New Jersey, NJ, USA; 2005

[11] Li T, Prieto J, Corchado JM, Bajo J. On the use and misuse of Bayesian filters. In: Proceeding of the IEEE 18th Int. Conference on Information Fusion (Fusion); 6-9 July 2015; Washington, DC, USA. IEEE; 2015

[12] Chen Z. Bayesian filtering: From Kalman filters to particle filters and beyond. Statistics. 2003;**182**(1):1-69

[13] Bishop CM. Pattern recognition. Machine Learning. 2006;**128**:1-58

[14] Duda RO, Hart PE, Stork DG. Pattern Classification. John Wiley & Sons; New Jersey, NJ, USA

[15] Neapolitan RE. Learning Bayesian Networks. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.; 2003

[16] Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 3rd ed. Prentice Hall; New Jersey, NJ, USA; 2010

[17] Cornell Activity Datasets CAD-60 [Internet]. Available from: http://pr.cs.cornell.edu/humanactivities/data.php [Accessed: January 2017

[18] Faria DR, Vieira M, Premebida C, Nunes U. Probabilistic human daily activity recognition towards robot-assisted living. In: Proceeding of the IEEE RO-MAN'15: IEEE International Symposium on Robot and Human Interactive Communication; Kobe, Japan; New York, NY, USA; September 2015