International Conference on Machine Learning and Computational Intelligence-2017
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 7 | ISSN : 2456-3307

Probabilistic Modeling in Machine Learning and Artificial Intelligence

Sajal Kaushik, Pulkit Kogat, Dr. Narina Thakur
Bharati Vidyapeeth's College of Engineering, A-4, Paschim Vihar, Rohtak Road, New Delhi, India

ABSTRACT

Probabilistic modeling plays a vital role in inferencing from huge datasets with high probability of uncertainity. This paper introduces most common probabilistic models applicable in machine learning found in litreature. An extensive litreature survey on Bayesian Networks, Markov Models,Hidden markov Models and stochastic grammars is captured under this single formalism. It also discusses a generic formalism called Bayesian Programming. The following paper presents various probabilistic modeling techniques used in practical applications of machine learning and machine learning related areas.

Keywords : Bayesianprogramming(bp), pertitentvariables (P), decomposition(d), Bayesian networks(bn), searching(s).

I. INTRODUCTION

The idea of probabilistic framework in machine learning provides models to explain the observed data set .Therefore, a machine can exploit such models to make better predictions about future datasets, and take decisions accordingly that are rational to the given predictions. Uncertainty plays a major role in inferencing. Observed data can be consistent with many models, and therefore which model is appropriate, can be thought of. Similarly, predictions about future data and the future consequences of actions are uncertain. Probability theory provides a framework for modelling uncertainty.

Machine learning is playing a vital role in every industry.machine learning involves training a model on particular dataset.With the huge generation of data, there is uncertainity in data which can be resolved to a larger extent with the help of probabilistic models.

II. LITREATURE REVIEW

Probabilistic modeling [1] is a mainstay of modern machine learning research, providing essential tools for analyzing the vast amount of data that have become available in cognitive science. [2]With the increasing amount of data, uncertainity of predicting the result from the given data has increased. Therefore,[3] requirement of probabilistic models to predict the output for test data. This section discusses various types of probabilistic models used in machine learning.

III. TYPES OF MODELS

A. BAYESIAN MODELS

Bayesian Models(composed of bayesian networks) have been evoloved as a basic approach for dealing with large datasets of probabilistic and uncertain information. These are the outcome of the integration between graph theory and various theories of probability. They are defined by the Bayesian algorithm of

$$
P \left\{ \begin{array}{l} D \left\{ \begin{array}{l} S \left\{ \begin{array}{l} \text{Variables} \\ \quad X^1, \ldots, X^N \\ \text{Decomposition} \\ \quad P(X^1 \ldots X^N) = \prod_{i=1}^{N} P(X^i |\; Pa^i), \\ \quad \text{with } Pa^i \subseteq \{X^1, \ldots, X^{i-1}\} \\ \text{Forms: any} \end{array} \right. \\ \text{Identification: any} \end{array} \right. \\ \text{Question: } P(X^i \mid Known) \end{array} \right.
$$

Figure 3: The BN formalism rewritten in BP.

Figure 3. The P(pertitent) variables have no constrains and have no specific meaning. The D(decomposition) variables, on the contrary, are clearly identified: it is a product of distributions of one variable Xi , conditioned by a conjunction of other variables, called its "parents", P a i , P a i ⊆ {X 1 , . . . , X i−1 }.This presumes that variables are ordered, and fully certains that applying Bayes' rule correctly defines the joint distribution .Also note that Xi indicates that one and only one variable. Thus, the above model, which can fit in a BP, does not in a BN:

P (A B C D) = P (A B)P (C | A)P (D | B).

In a BN, if A and B are to sight together on the LHS of a term, as in P (A B), then they have to be merged into a single variable Ai, Bi, and cannot be consequently separated, as in P (C | A) and P (D | B).

A bijection exists within joint probability distributions defined by such a decomposition and directed acyclic graphs: nodes are associated to variables, and edges are associated to conditional dependencies. Using graphs in probabilistic models leads to an efficient way to define hypotheses over a set of variables, an economic representation of a joint probability distribution and, most importantly,an easy and efficient way to do probabilistic inference.

The parametric forms are not constrained theoretically, but in BN commercial softwares they are very often restricted to probability tables, or tables and constrained Gaussians.

B. HIDDEN MARKOV MODELS

Hidden markov models are the advance models containing bayesian filters

$$
P \left\{ \begin{array}{l} D \left\{ \begin{array}{l} S \left\{ \begin{array}{l} \text{Variables} \\ \quad S_0, \ldots, S_T, O_0, \ldots, O_T \\ \text{Decomposition} \\ \quad P(S_0 \ldots S_T\; O_0 \ldots O_T) = \\ \quad\quad P(S_0) P(O_0 \mid S_0) \\ \quad\quad \prod_{i=1}^{T} [P(S_i \mid S_{i-1}) P(O_i \mid S_i)] \\ \text{Forms} \\ \quad P(S_0) \;\equiv\; \text{Matrix} \\ \quad P(S_i \mid S_{i-1}) \;\equiv\; \text{Matrix} \\ \quad P(O_i \mid S_i) \;\equiv\; \text{Matrix} \end{array} \right. \\ \text{Identification: Baum-Welch algorithm} \end{array} \right. \\ \text{Question: } P(S_0\; S_1 \ldots S_t - 1 \mid S_t\; O_0 \ldots O_t) \end{array} \right.
$$

Figure 2

Variables are assumed to be discrete. Therefore, the transition model P (S i | S i−1) and the observation model P (O i | S i) are both specified using probability matrices (conditional probability tables).Variants exist about this particular point: when the observation variables are continuous, the formalism becomes known as "semi-continuous HMMs" .In this case, the observation model is associated either with a Gaussian form, or a Mixture of Gaussian form.

C. MARKOV MODELS

In probability theory, a Markov model is a stochastic model used to model randomly and site changing systems where it is assumed that future states are dependent on the present state not on the events that occurred previously (it assumes the Markov property). Generally, this assumption with the hardcoded markov model helps us to improve inferencing and computatability otherwise it is intracable. For this reason, in the fields of predictive modelling and probabilistic forecasting, it is necessary to use markov models for better predictions.

D. STOCHASTIC GRAMMARS

PCFGs extend context-free grammars similar to how hidden Markov models extend regular grammars.Probability is assigned to each production. The probabilities derivation uses the product of product derivations.These probabilities can be viewed as parameters of the model, and for large problems it is convenient to learn these parameters via machine learning.The validity of probabilistic grammar is limited by the context of its training dataset.

PCFGs have application ranging from nlp to rna molecules to design of programming languages. Weighing factors of scalability and generality result in designing of efficient PCFG.

Table 1

OBJECTIVE	TECHNIQUES USED				
	Generic probabilistic modelling	**Bayesian Model**	**Markov models**	**Hidden Markov modes**	**Stochastic grammars**
Improving PILCO(a reinforcement learning model)[12]		✓			
Improving ensemble learning[11]	✓				
Analysis of neuroimaging data using fsl[13]		✓			
To improve exploratory behaviour in student's model[14]	✓				
Sampling and Bayes' Inference in Scientific Modelling and Robustness[15]		✓			
Bayesian exponential random graph modelling of interhospital patient referral networks[16]		✓			
Estimation of distributive algorithms[17]		✓			
Opportunities for Personalization in Modeling Students[18]		✓			
Towards a probabilistic formalisation of case based inference[19]	✓				
Probabilistic modelling of joint orientation[20]	✓				
survey of probabilistic models, using the Bayesian Programming methodology as a unifying framework[21]		✓			

Global behaviour patterns using probabilistic latent semantic analysis[22]	✓				
2-dimension dynamic Bayesian network for large-scaledegradation modelling with an application bridges network[23]		✓			
Complexity of inference of probabilistic models[24]	✓				
Constrained Bayesian networks [25]		✓			
Reliabilty modelling with dynamic Bayesian networks[26]		✓			
A probabilistic approach to modelling uncertain arguments[27]	✓				
Probabilistic modelling ,inference and learning using logical theories[28]	✓				
Probabilistic models of cognition[29]					✓
Probabilistic sensitivity analysis of complex models[30]		✓			
Statistical inference and probabilistic modelling for constained based NLP[31]					✓
Recognisition of patterns[32]			✓		
Speech recognisition[33]				✓	
Learning and detecting activities from movement trajectories using the hierachial hidden markov model[34]				✓	
Large margin hidden markov models for automatic speech recognisition[35]				✓	
Hidden markov models in computer vision[36]				✓	

IV. DESCRIPTION OF TABLE-1 AND PIE-CHART

The above table describesreal life problems in which probabilistic modelling is used ranging from speech recognisition[35] ,computer vision[36],nlp to distributed algorithms specifically in all machine learning related fields.It provides a abstract of the problem and technique of probabilistic modelling used.pie-chart simply presents the percentage of technique used in various applications mentioned in the literature review of the paper.table-1 presents the wide-ranging applications in which probabilistic modelling can be used in all machine learning related fields. Further proving that there is a huge scope in on improving the probabilistic models for removing the uncertainity in the data.

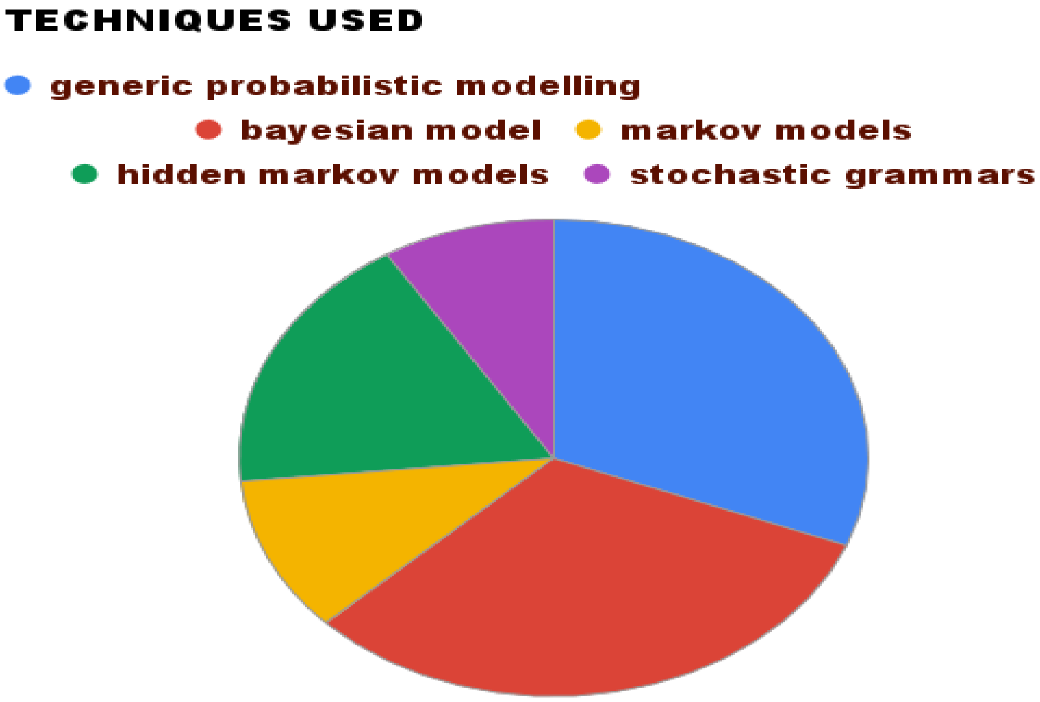

Figure 4

V. RECOMMENDATIONS

- ✓ it has immense work in computer vision where there is lot of uncertainity in the data.
- ✓ Stochastic grammars are mostly used in recommender systems where they usually generate a sentence.
- ✓ In concepts like pca in dimension reduction, probabilistic modelling is used.
- ✓ Probabilistic modeling will play a important role in nlp related field.

VI. CONCLUSION

Probabilistic approaches in machine learning is a very active research area with wide-ranging impact beyond conventional pattern-recognition problems .The key difference between problems in which a probabilistic approach is important and problems that can be solved using non-probabilistic machine-learning approaches is whether uncertainty has a central role. Moreover, most conventional optimization-based machine-learning approaches have probabilistic analogues that handle uncertainty in a more principled manner.

VII. REFERENCES

[1]. Bishop, C. M. Pattern Recognition and Machine Learning (Springer,2006).

[2]. Murphy, K. P. Machine Learning: A Probabilistic Perspective (MIT Press, 2012).

[3]. Ghahramani, Z. Bayesian nonparametrics and the probabilistic approach to modelling. Phil. Trans. R. Soc. A 371, 20110553 (2013).

[4]. Jaynes, E. T. Probability Theory: the Logic of Science (Cambridge Univ. Press, 2003).

[5]. Koller, D. & Friedman, N. Probabilistic Graphical Models: Principles and Techniques (MIT Press, 2009).

[6]. Doya, K., Ishii, S., Pouget, A. & Rao, R. P. N. Bayesian Brain: Probabilistic Approaches to Neural Coding (MIT Press, 2007).

[7]. Jordan, M., Ghahramani, Z., Jaakkola, T. & Saul, L. An introduction to variational methods in graphical models. Mach. Learn. 37, 183–233 (1999).

[8]. Pfeffer, A. Practical Probabilistic Programming (Manning, 2015).

[9]. Bishop, C. M. Model-based machine learning. Phil. Trans. R. Soc. A 371, 20120222 (2013).

[10]. Mansinghka, V., Selsam, D. & Perov, Y. Venture: a higher-order probabilistic programming platform with programmable inference. Preprint at http://arxiv. org/abs/1404.0099 (2014).

[11]. MacKay, David JC. "Developments in probabilistic modelling with neural networks-ensemble learning." In Neural Networks: Artificial Intelligence and Industrial Applications. Proc. of the 3rd Annual Symposium on Neural Networks,pp.191-198. 1995.

[12]. Gal, Yarin, Rowan Thomas McAllister, and Carl Edward Rasmussen. "Improving PILCO with bayesian neural network dynamics models." In Data-Efficient Machine Learning workshop, vol. 951, p. 2016. 2016.

[13]. Woolrich, Mark W., Saad Jbabdi, Brian Patenaude, Michael Chappell, Salima Makni, Timothy Behrens, Christian Beckmann, Mark Jenkinson, and Stephen M. Smith."Bayesian analysis of neuroimaging data in FSL." Neuroimage 45, no. 1 (2009): S173-S186.

[14]. Bunt, Andrea, and Cristina Conati. "Probabilistic student modelling to improve exploratory behaviour." User Modeling and User-Adapted Interaction 13, no. 3 (2003): 269-309.

[15]. Box, GEORGE EP. "Sampling and Bayes' inference in scientific modelling and robustness." Journal of the Royal Statistical Society A 143 (1980): 383-430.

[16]. Caimo, Alberto, Francesca Pallotti, and Alessandro Lomi. "Bayesian exponential random graph modelling of interhospital patient referral networks." Statistics in Medicine (2017).

[17]. Gallagher, Marcus, Ian Wood, Jonathan Keith, and George Sofronov. "Bayesian inference in estimation of distribution algorithms." In Evolutionary Computation, 2007. CEC 2007. IEEE Congress on, pp. 127-133. IEEE, 2007.

[18]. Garcia, Patricio, Analia Amandi, Silvia Schiaffino, and Marcelo Campo. "Evaluating Bayesian networks' precision for detecting students' learning styles." Computers & Education 49, no. 3 (2007): 794-808.

[19]. Hullermeier, Eyke. "Toward a probabilistic formalization of case-based inference." In IJCAI, pp. 248-253. 1999.

[20]. Kulatilake, Pinnaduwa HSW, Tien H. Wu, and Deepa N. Wathugala. "Probabilistic modelling of joint orientation." International Journal for Numerical and Analytical Methods in Geomechanics 14, no. 5 (1990): 325-350.

[21]. Diard, Julien, Pierre Bessiere, and Emmanuel Mazer. "A survey of probabilistic models using the bayesian programming methodology as a unifying framework." (2003).

[22]. Li, Jian, Shaogang Gong, and Tao Xiang. "Global Behaviour Inference using Probabilistic Latent Semantic Analysis." In BMVC, vol. 3231, p. 3232. 2008.

[23]. Kosgodagan, Alex, et al. "A 2-dimension dynamic Bayesian network for large-scale degradation modelling with an application to a bridges network." (2017).

[24]. Jaeger, Manfred. "On the complexity of inference about probabilistic relational models." Artificial Intelligence 117, no. 2 (2000): 297-308.

[25]. Beaumont, Paul, and Michael Huth. "Constrained Bayesian Networks: Theory, Optimization, and Applications." arXiv preprint arXiv:1705.05326(2017).

[26]. Weber, Philippe, and Lionel Jouffe. "Reliability modelling with dynamic bayesian networks." IFAC Proceedings Volumes 36, no. 5 (2003): 57-62.

[27]. Hunter, Anthony. "A probabilistic approach to modelling uncertain logical arguments." International Journal of Approximate Reasoning 54, no. 1 (2013):47-81.

[28]. Ng, Kee Siong, John W. Lloyd, and William TB Uther. "Probabilistic modelling, inference and learning using logical theories." Annals of Mathematics and Artificial Intelligence 54, no. 1-3 (2008): 159-205.

[29]. Chater, Nick, Joshua B. Tenenbaum, and Alan Yuille. "Probabilistic models of cognition: Conceptual foundations." Trends in cognitive sciences 10, no. 7 (2006): 287-291.

[30]. Oakley, Jeremy E., and Anthony O'Hagan. "Probabilistic sensitivity analysis of complex models: a Bayesian approach." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66, no. 3 (2004): 751-769.

[31]. Riezler, Stefan. "Statistical inference and probabilistic modelling for constraint-based nlp." arXiv preprint cs/9905010 (1999).

[32]. Bishop, Christopher M. Pattern recognition and machine learning. springer, 2006.

[33]. Nilsson, Mikael, and Marcus Ejnarsson. "Speech recognition using hidden markov model." (2002).

[34]. Nguyen, Nam Thanh, Dinh Q. Phung, Svetha Venkatesh, and Hung Bui. "Learning and detecting activities from movement trajectories using the hierarchical hidden Markov model." In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2, pp. 955-960. IEEE, 2005.

[35]. Sha, Fei, and Lawrence K. Saul. "Large margin hidden Markov models for automatic speech recognition." In Advances in neural information processing systems, pp. 1249-1256. 2007.

[36]. Bunke, H. and Caelli, T. eds., 2001. Hidden Markov models: applications in computer vision (Vol. 45). World Scientific.