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 uncertainty. This paper introduces most common probabilistic models applicable in machine learning found in literature. An extensive literature 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: Bayesian programming (bp), pertinent variables (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 uncertainty in data which can be resolved to a larger extent with the help of probabilistic models.

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, uncertainty 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.

II. LITREATURE REVIEW

III. TYPES OF MODELS

A. BAYESIAN MODELS

Bayesian Models (composed of bayesian networks) have been evolved 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
Figure 3: The P(pertinent) 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 ⊆ {X1 , . . . , 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

Variables
\[ S_0, \ldots, S_T, O_0, \ldots, O_T \]

Decomposition
\[ P(S_0 \ldots S_T O_0 \ldots O_T) = P(S_0)P(O_0 | S_0) \prod_{i=1}^{T} [P(S_i | S_{i-1})P(O_i | S_i)] \]

Forms
- \[ P(S_i) \equiv \text{Matrix} \]
- \[ P(S_i | S_{i-1}) \equiv \text{Matrix} \]
- \[ P(O_i | S_i) \equiv \text{Matrix} \]

Identification: Baum-Welch algorithm

Question: \[ P(S_0 S_1 \ldots S_{t-1} | S_t O_0 \ldots O_t) \]

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 intractable. 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
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>
<thead>
<tr>
<th>OBJECTIVE</th>
<th>TECHNIQUES USED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving PILCO (a reinforcement learning model) [12]</td>
<td>✓</td>
</tr>
<tr>
<td>Improving ensemble learning [11]</td>
<td>✓</td>
</tr>
<tr>
<td>Analysis of neuroimaging data using fsl [13]</td>
<td>✓</td>
</tr>
<tr>
<td>To improve exploratory behaviour in student’s model [14]</td>
<td>✓</td>
</tr>
<tr>
<td>Sampling and Bayes’ Inference in Scientific Modelling and Robustness [15]</td>
<td>✓</td>
</tr>
<tr>
<td>Bayesian exponential random graph modelling of interhospital patient referral networks [16]</td>
<td>✓</td>
</tr>
<tr>
<td>Estimation of distributive algorithms [17]</td>
<td>✓</td>
</tr>
<tr>
<td>Opportunities for Personalization in Modeling Students [18]</td>
<td>✓</td>
</tr>
<tr>
<td>Towards a probabilistic formalisation of case based inference [19]</td>
<td>✓</td>
</tr>
<tr>
<td>Probabilistic modelling of joint orientation [20]</td>
<td>✓</td>
</tr>
<tr>
<td>Survey of probabilistic models, using the Bayesian Programming methodology as a unifying framework [21]</td>
<td>✓</td>
</tr>
<tr>
<td>Topic</td>
<td>✓</td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>---</td>
</tr>
<tr>
<td>Global behaviour patterns using probabilistic latent semantic analysis[22]</td>
<td>✓</td>
</tr>
<tr>
<td>2-dimension dynamic Bayesian network for large-scale degradation modelling with an application bridges network[23]</td>
<td>✓</td>
</tr>
<tr>
<td>Complexity of inference of probabilistic models[24]</td>
<td>✓</td>
</tr>
<tr>
<td>Constrained Bayesian networks [25]</td>
<td>✓</td>
</tr>
<tr>
<td>Reliability modelling with dynamic Bayesian networks[26]</td>
<td>✓</td>
</tr>
<tr>
<td>A probabilistic approach to modelling uncertain arguments[27]</td>
<td>✓</td>
</tr>
<tr>
<td>Probabilistic modelling , inference and learning using logical theories[28]</td>
<td>✓</td>
</tr>
<tr>
<td>Probabilistic models of cognition[29]</td>
<td></td>
</tr>
<tr>
<td>Probabilistic sensitivity analysis of complex models[30]</td>
<td>✓</td>
</tr>
<tr>
<td>Statistical inference and probabilistic modelling for constrained based NLP[31]</td>
<td></td>
</tr>
<tr>
<td>Recognition of patterns[32]</td>
<td>✓</td>
</tr>
<tr>
<td>Speech recognition[33]</td>
<td></td>
</tr>
<tr>
<td>Learning and detecting activities from movement trajectories using the hierachial hidden markov model[34]</td>
<td></td>
</tr>
<tr>
<td>Large margin hidden markov models for automatic speech recognisation[35]</td>
<td></td>
</tr>
<tr>
<td>Hidden markov models in computer vision[36]</td>
<td></td>
</tr>
</tbody>
</table>
IV. DESCRIPTION OF TABLE-1 AND PIE-CHART

The above table describes real life problems in which probabilistic modelling is used ranging from speech recognition[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 uncertainty in the data.

![Pie Chart of Techniques Used]

V. RECOMMENDATIONS

☑ it has immense work in computer vision where there is lot of uncertainty 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


