COM4511 Speech Technology - Practical Exercise - Graphical Models

Anton Ragni

1 Introduction

Graphical models and dynamic Bayesian networks (DBN) in particular provide a convenient methodology for displaying complex sequence models. Numerous DBNs were presented this semester to describe sequence models ranging from acoustic and language models to dialogue systems modelled by Markov decision processes. There is a wide range of literature discussing not only the methodology of describing statistical models using the elements of graphical models but also the unified framework for performing inference and parameter estimation. These aspects, however, have not been covered in lecture materials and are more suitable for an independent study. The recommended literature section contains references to some related works.

A less restricted form of DBNs, pseudo DBNs, are often used to describe and display complex forms of neural networks that have become popular in speech technology and beyond. These networks are becoming more complex each year. As such, an accurate and succinct methodology for describing these approaches is of crucial importance from both engineering and pedagogical standpoints. Although there is much less, if any, literature on pseudo DBNs, they largely follow the notation of standard DBNs. Given that few dependencies in a typical neural network architecture are probabilistic, the distinction between probabilistic and non-probabilistic dependencies in such pseudo DBNs is not normally maintained. Instead, dependencies represent the flow of information (computation) through the network and hence have a clear directional characteristic.

The lecture materials were meant to illustrate the flexibility offered by standard and pseudo DBNs in describing both simple and complex sequence models. This task will test your understanding of these approaches and your ability to define, describe and utilise the elements of methodology to discuss, criticise and improve two known forms of sequence models.

2 Task

The main task consists of two subtasks. Please take a close look at each respective figure and answer given questions. Your report should include a detailed answer to each question. It may require you to typeset equations, draw diagrams and embed references to published works. If you are not sure how to best proceed please ask a question on Blackboard Forum or e-mail Anton Ragni directly using `a.ragni@sheffield.ac.uk`.

2.1 Dynamic Bayesian Networks

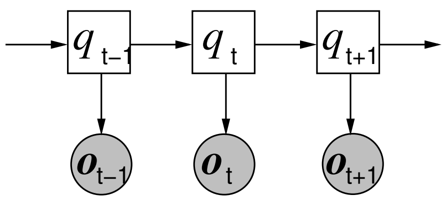

Figure 1: An example of dynamic Bayesian networks

1. Name sequence models described by Figure 1.
2. Describe all elements of Figure 1.
   (a) If variables are present then describe their domain and range.
   (b) If dependencies are present then describe their nature.

3. For any dependency present in Figure 1
   (a) Give a mathematical description of each dependency. Make sure to name each term.
   (b) Describe one, parametric or non-parametric, approach to model the dependency. Write down the corresponding mathematical expression and name each term.
4. Describe three types of dependencies that can be added to Figure 1.
   (a) Update Figure 1 to show each dependency type. Add 3 newly created figures to your written report.
   (b) Describe what impact each dependency would have had if they were added.
   (c) Update, if needed, the mathematical expressions in point 3(b) to incorporate new dependencies.
5. If there are latent (unobserved) variables in Figure 1
   (a) Describe how unobserved variables can be transformed into observed variables.
   (b) Describe how unobserved variables can be marginalised over.
   (c) What is the main difference between 5(a) and 5(b) above. Discuss under which conditions 5(a) is either an adequate or inadequate approximation to 5(b).
   (d) If each unobserved variable can take one of N values and the length of sequence is T, what are time complexities of processes 5(a) and 5(b).
6. If there are observed variables in Figure 1.
   (a) Describe how the joint probability density/mass function can be computed.
   (b) Describe suitable choice for modelling observed variables if these are D-dimensional vectors for all t which are *i)* uncorrelated and *ii)* correlated.
   (c) Describe the issue arising from the variables in 6(b) being correlated across time and a possible solution.

2.2 Pseudo Dynamic Bayesian Networks

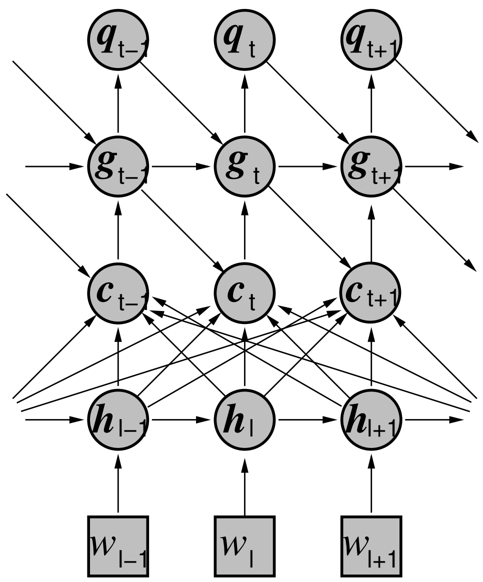

Figure 2: An example of pseudo dynamic Bayesian networks

1. Name sequence models described by Figure 2.
2. Describe all elements of Figure 2.
   (a) If variables are present then describe their domain and range.

(b) If dependencies are present then describe their nature.

3. For any variable present in Figure 2
   (a) Write down the most general expression linking all dependent variables
   (b) Give a particular expression for each variable in 3(a).

   (Make sure to name all terms and define all functions.)

4. Group all variables into meaningful groups
   (a) Name each group.
   (b) Describe the role each group is playing.
   (c) Discuss an alternative choice of variables (implementation) for each group.

5. For one model from point 1
   (a) Describe inputs and outputs.
   (b) Describe criteria for estimating model parameters and performance assessment.
   (c) If criteria are different, discuss the rationale, potential issues and possible solutions.
   (d) If criteria identical, discuss why ”perfect” performance has not been achieved, what can be done about that.

   (Make sure to provide proper citations of any publications used.)

6. Pick an alternative to the model in point 5.
   - Draw a reasonably accurate pseudo dynamic Bayesian network.
   - Provide mathematical expressions for all variables.
   - Discuss main differences compared to the model in point 5.

3 Marking Scheme

Two critical elements are assessed: dynamic Bayesian networks and pseudo dynamic Bayesian networks. *Note:* Even if you cannot complete this task as a whole you can certainly provide a description of what you were planning to accomplish.

1. **Dynamic Bayesian Networks.**
   - **Figure 1 models (1 mark for each correct and -1 mark for each incorrect, maximum 5 marks and minimum -5 marks).** Name as many statistical sequence models described by Figure 1 as you can. Any model not covered in lecture materials must be supplied with a reference to publication.
   - **Elements (5 marks).** List all elements of DBNs present in Figure 1. If there are implied yet not shown elements then acknowledge their existence. For each variable present in Figure 1 state whether it is discrete or continuous, scalar or vector; also describe a typical range. For each dependency present in Figure 1 describe whether it is probabilistic or general (unconstrained), which elements of Figure 1 it is concerned with.
   - **Dependencies (5 marks).** Give a functional form for each dependency. Describe all arguments and the chosen functional form. If it is probabilistic then make an explicit statement whether it is a probability mass function or a probability density function. Describe one approach to model each dependency.
   - **Additional dependencies (20 marks).** Create a new diagram for each added dependency or draw on top of Figure 1. (If you cannot crop Figure 1 from the supplied PDF file please email Anton Ragni <`a.ragni@sheffield.ac.uk`>). Add three created diagrams to your report. Discuss each dependency following the previous subsection **Dependencies (5 marks).** Describe the impact of each added dependency.
   - **Unobserved variables (if present maximum 20 and minimum 0 marks, if not maximum 0 and minimum -20 marks).**
     Describe one general approach to transform unobserved variables into observed variables. Provide a pseudo-code of an algorithm that implements this approach. (Alternatively, name the underlying algorithm.) Describe one useful application of transforming unobserved variables into observed variables.

Describe one general approach to marginalise (or sum) over unobserved variables. Provide a pseudo-code of an algorithm that implements this approach. (Alternatively, name the underlying algorithm.) Describe one useful application of marginalising over unobserved variables.

Discuss the main difference between the above two approaches in one or two sentences. Discuss under which conditions the probability distribution over original and newly observed variables is either an adequate or inadequate approximation to the probability distribution over the original observed variables. For instance, if x is an observed variable, y is an unobserved discrete variable and $\hat{y}$ is the unobserved variable transformed into an observed variable, the question above asks you to compare the adequacy of using $p(x, \hat{y})$ to approximate $p(x) = \sum_y p(x, y)$.

Discuss complexities of the above two approaches if each unobserved variable can take on of N discrete values and the length of sequence is T. Use standard complexity theory terminology (constant, linear, quadratic, etc.) or nomenclature (o, $\mathcal{O}$, Ω, etc.).

- **Observed variables (if present maximum 20 and minimum 0 marks, if not maximum 0 and minimum -20 marks).**

  Describe how the joint probability density/mass function can be computed over observed variables. Provide a pseudo-code of an algorithm that implements your approach. (Alternatively, name the underlying algorithm.)

  Assume each observed variable is a D-dimensional continuous vector. Which probability density function would you use if elements of those vectors were *i)* uncorrelated and *ii)* correlated, and why?

  Additionally assume that these vectors are correlated in time. Describe one issue arising from this and one solution to address it.

2. **Pseudo Dynamic Bayesian Networks.**

- **Figure 2 (1 mark for each correct and -1 mark for each incorrect, maximum 5 marks and minimum -5 marks).** Name as many sequence models described by Figure 2 as you can. Any model not covered in lecture materials must be supplied with a reference to publication.
- **Elements (5 marks).** List all elements of pseudo-DBNs present in Figure 2. If there are implied yet not shown elements then acknowledge their existence. For each variable present in Figure 2 state whether it is discrete or continuous, scalar or vector; also describe a typical range. For each dependency present in Figure 1 describe whether it is probabilistic or general (unconstrained), which elements of Figure 2 it is concerned with.
- **Variables (5 marks).** Write down the most general functional form for each variable that links it with all dependent variables. Provide a particular form (whether used in the literature or not) for each variable.
- **Variable groups (5 marks).** If some variables form meaningful groups then name all such groups you can find in Figure 2. Describe the role each such group is playing and one alternative group you can think of to replace it.
- **Model description (20 marks).** Pick one of the models from the first subsection (point 1). Describe inputs and outputs (name the corresponding variables and describe what do they represent). Describe how model parameters are estimated (objective function). Describe how performance is evaluated (objective function). If objective functions used in training and evaluation are different then discuss one rationale for such choice, one potential issue and one possible solution. If objective functions used in training and evaluation are the same then discuss why the "perfect" performance has not been achieved and what can be done about that.
- **Alternative model description (20 marks).** Pick an alternative to the model described in the previous subsection **Model description**. This can be either a model discussed in lectures or not. The model must be applied to the same sequence modelling problem as the model in the previous section. Use any graphics editor to produce a pseudo DBN diagram or copy any existing diagram (make sure to reference it). Write down the most general functional form for each variable. Ensure that all dependent variables have been included. Discuss main differences compared to the model discussed in the previous subsection.

4 Hand-in procedure

All outcomes, however complete, are to be submitted jointly in a form of a package file (`zip/tar/gzip`) that includes directories for each task which contain the associated required files. Submission will be performed via MOLE.

5 Reading List

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*, Chapter 8: Graphical Models, Springer, pp. 359–422. Available online: `https://www.microsoft.com/en-us/research/wp-content/uploads/2016/05/Bishop-PRML-sample.pdf`

[2] M. I. Jordan, ”Graphical Models”, Statistical Science, pp. 140–155, 2004. Available online: `https://projecteuclid.org/download/pdfview_1/euclid.ss/1089808279`

[3] J. Bilmes, ”Graphical Models and Automatic Speech Recognition”. In R. Rosenfeld, M. Ostendorf, S. Khudanpur, and M. Johnson (ed.), *Mathematical Foundations of Speech and Language Processing*, Springer-Verlag, 2003. Available online: `https://people.ece.uw.edu/bilmes/p/mypubs/bilmes2003-gmasr-book.pdf`