Conclusion of chapter 4: Graphical Models

yiqing.tao

November 14, 2016

1 summary

1. Here well focus on Markov networks, chain graphs (which marry Belief and Markov networks) and factor graphs.
2. **Modelling** After identifying all potentially relevant variables of a problem environment, our task is to describe how these variables can interact. This is achieved using structural assumptions as to the form of the joint probability distribution of all the variables, typically corresponding to assumptions of independence of variables. Each class of graphical model corresponds to a factorisation property of the joint distribution.

   Inference Once the basic assumptions as to how variables interact with each other is formed (i.e. the probabilistic model is constructed) all questions of interest are answered by performing inference on the distribution. This can be a computationally non-trivial step so that coupling GMs with accurate inference algorithms is central to successful graphical modelling.

   Whilst not a strict separation, GMs tend to fall into two broad classes those useful in modelling, and those useful in representing inference algorithms. **For modelling**, belief networks, Markov networks, chain graphs and influence diagrams are some of the most popular. **For inference** one typically compiles a model into a suitable GM for which an algorithm can be readily applied. Such inference GMs include factor graphs and junction trees.
3. Belief networks intuitively describe which variables causally influence others and are represented using directed graphs.

   A Markov network is represented by an undirected graph.
4. Intuitively, linked variables in a Markov network are graphically dependent, describing local cliques of graphically dependent variables.
5. Markov networks are historically important in physics and may be used to understand how global collabo- rative phenomena can emerge from only local dependencies.
6. Factor graphs describe the factorisation of functions and are not necessarily related to probability distribu- tions.

2 Markov Networks

1. Definition 4.1 (Potential). A potential $\phi(x)$ is a non-negative function of the variable x, $\phi(x) \geq 0$. A joint potential $(x_1, ..., x_n)$ is a non-negative function of the set of variables. A distribution is a special case of a potential satisfying normalization, $\sum_x \phi(x) = 0$. This holds similarly for continuous variables, with summation replaced by integration.

   Belief networks correspond to a special kind of factorization of the joint probability distribution in which each of the factors is itself a distribution. An alternative factorization is, for example:

$$p(a, b, c) = \frac{1}{Z}\phi(a, b)\phi(b, c) \tag{1}$$

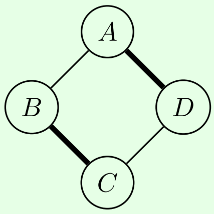
where $\phi(a,b)$ and $\phi(b,c)$ are *potentials* and Z is a constant which ensures normalisation, called *the partition function*

$$Z = \sum_{a,b,c} \phi(a,b)\phi(b,c) \tag{2}$$

- Markov Networks are defined as products of potentials defined on maximal cliques of an undirected graph.

Example 1(from the internet):

A,B,C,D four people's opinions on one thing are binary $\{0,1\}$. A,D always have the same opinion and B,C always have the same opinion, while C,D are on oppose side and A,B always doubt each other, as figure(1). Four people's opinions are all random variables, but they'll come to a consensus. Therefore, what we're interested in is the probability of each consensus.

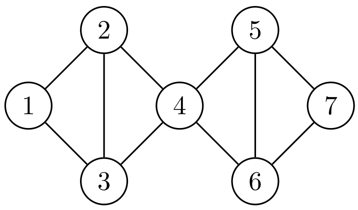


2. Definition 4.2 (Markov Network). For a set of variables $\chi = x_1, ..., x_n$ a Markov network is defined as a product of potentials on subsets of the variables $\chi_c \subseteq \chi$:

$$p(x_1, ..., x_n) = \frac{1}{Z}\prod_{c=1}^{N} \phi_c(\chi_c) \tag{3}$$

The constant Z ensures the distribution is normalised. Graphically this is represented by an undirected graph G with $\chi_c, c = 1, ..., C$ being the maximal cliques of G. For the case in which clique potentials are strictly positive, this is called a *Gibbs distribution* (*Boltzmann distribution*).

3. Definition 4.3 (Pairwise Markov network).In the special case that the graph contains cliques of only size 2, the distribution is called a pairwise Markov Network, with potentials defined on each link between two variables.



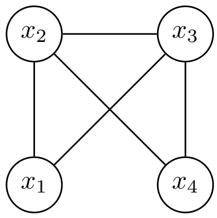


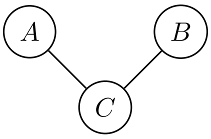
Figure 4.2: $\phi(1,2,3)\phi(2,3,4)\phi(4,5,6)\phi(5,6,7)$.By the global Markov property, since every path from 1 to 7 passes through 4, then $1 \perp\!\!\!\perp 7|4$.

Whilst a Markov network is formally defined on maximal cliques, in practice authors often use the term to refer to non-maximal cliques. For example, in the graph on the right, the maximal cliques are x_1, x_2, x_3 and x_2, x_3, x_4, so that the graph describes a distribution $p(x_1, x_2, x_3, x_4) = \phi(x_1, x_2, x_3)\phi(x_2, x_3, x_4)/Z$. In a pairwise network though the potentials are assumed to be over two-cliques, giving

$$p(x_1, x_2, x_3, x_4) = \phi(x_1, x_2)\phi(x_1, x_3)\phi(x_2, x_3)\phi(x_2, x_4)\phi(x_3, x_4)/Z$$

.

4. Definition 4.4 (Properties of Markov Networks).



$p(A, B, C) = \phi_{AC}\phi_{BC}/Z$

Marginalising over C makes A and B (graphically) dependent. In general $p(A, B) \neq p(A)p(B)$.

Conditioning on C makes A and B independent: $p(A, B|C) = p(A|C)p(B|C)$.

5. Definition 4.5 (Separation). A subset S separates a subset A from a subset B (for disjoint A and B) if every path from any member of A to any member of B passes through S. If there is no path from a member of A to a member of B then A is separated from B. If $S = \phi$ then provided no path exists from A to B, A and B are separated.

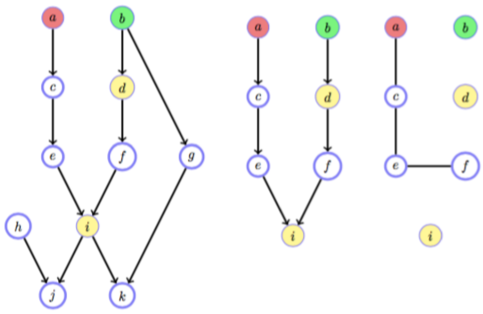
6. Definition 4.6 (Global Markov Property). For disjoint sets of variables, (A, B, S) where S separates A from B in G, then $A \perp\!\!\!\perp B|S$.

7. Definition 4.7 (Markov Random Field, A MRF is a set of conditional distributions, one for each indexed location.) A MRF is defined by a set of distributions $p(x_i|ne(x_i))$ where $i \in 1, ..., n$ indexes the distributions and $ne(x_i)$ are the neighbours of variable x_i, namely that subset of the variables $x_1, ..., x_n$ that the distribution of variable xi depends on. The term Markov indicates that this is a proper subset of the variables. A distribution is an MRF with respect to an undirected graph G if

$$p(x_i|X_{\setminus i}) = p(x_i|ne(x_i)) \tag{4}$$

where $ne(x_i)$ are the neighbouring variables of variable x_i, according to the undirected graph G. The notation $x_{\setminus i}$ is shorthand for the set of all variables X excluding variable x_i, namely $X \backslash x_i$ in set notation.

8. Conditional independence using Markov networksFor X, Y, Z each being collections of variables, in section(3.3.4) we discussed an algorithm to determine if $X \perp\!\!\!\perp Y|Z$ for belief networks. An alternative and more general method (since it handles directed and undirected graphs) uses the procedure below. See fig(4.4) for an example.



(a) (b) (c)

Figure 4.4: **(a)**: Belief network for which we are interested in checking conditional independence $a \perp\!\!\!\perp b | \{d, i\}$. **(b)**: Ancestral graph. **(c)**: Ancestral, moralised and separated graph for $a \perp\!\!\!\perp b | \{d, i\}$. There is no path from a red to green node so a and b are independent given d, i.

Procedure 4.2 (Ascertaining independence in Markov and belief networks). For Markov Networks only the final separation criterion needs to be applied:

Ancestral Graph Identify the ancestors A of the nodes $X \cup Y \cup Z$. Retain the nodes $X \cup Y \cup Z$ but remove all other nodes which are not in A, together with any edges in or out of such nodes.

Moralisation Add a link between any two remaining nodes which have a common child, but are not already connected by an arrow. Then remove remaining arrowheads.

Separation Remove links neighbouring Z. In the undirected graph so constructed, look for a path which joins a node in X to one in Y. If there is no such path deduce that $X \amalg Y | Z$.

Note that the ancestral step in procedure(4.2) for belief networks is intuitive since, given a set of nodes X and their ancestors A, the remaining nodes D form a contribution to the distribution of the form $p(D|X, A)p(X, A)$, so that summing over D simply has the effect of removing these variables from the DAG.

3 Chain Graphical Model

1. Chain Graphs (CGs) contain both directed and undirected links. To develop the intuition, consider fig(4.6a).

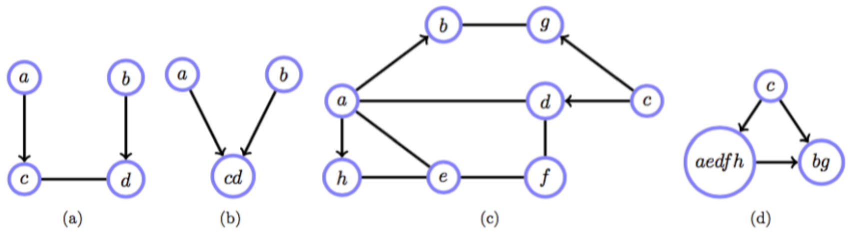


Figure 4.6: Chain graphs. The chain components are identified by deleting the directed edges and identifying the remaining connected components. **(a)**: Chain components are (a),(b),(c, d), which can be written as a BN on the cluster variables in (b). **(c)**: Chain components are (a, e, d, f, h), (b, g), (c), which has the cluster BN representation (d).

4 Factor Graphs

1. Definition 4.10 (Factor Graph).Given a function

$$f(x_1, ..., x_n) = \prod_i \psi_i(\chi_i) \tag{5}$$

The FG has a node (represented by a square) for each factor ψ_i, and a variable node (represented by a circle) for each variable x_j. For each$x_j \in \chi_i$an undirected link is made between factor ψ_i and variable x_j.

When used to represent a distribution

$$p(x_1, ..., x_n) = \frac{\prod_i \psi_i(\chi_i)}{Z} \tag{6}$$

a normalisation constant $Z = \sum_X \prod_i \psi_i(\chi_i)$ is assumed. Here X represents all variables in the distribution.

For a factor $\psi_i(\chi_i)$ which is a conditional distribution $p(x_i|pa(x_i))$, we may use directed links from the parents to the factor node, and a directed link from the factor node to the child x_i. This has the same structure as an (undirected) FG, but preserves the information that the factors are distributions.

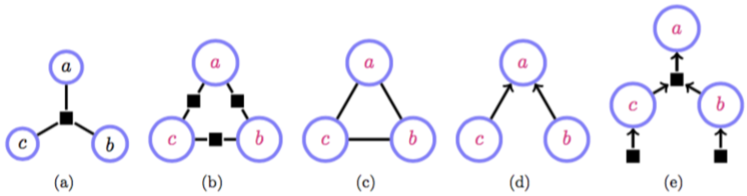


Figure 4.8: **(a)**: $\phi(a, b, c)$. **(b)**: $\phi(a, b)\phi(b, c)\phi(c, a)$. **(c)**: $\phi(a, b, c)$. Both (a) and (b) have the same undirected graphical model, (c). **(d)**: (a) is an undirected FG of (d). **(e)**: Directed FG of the BN in (d). A directed factor represents a term $p(\text{children}|\text{parents})$. The advantage of (e) over (a) is that information regarding the marginal independence of variables b and c is clear from graph (e), whereas one could only ascertain this by examination of the numerical entries of the factors in graph (a).

2. Conditional independence in factor graphs:

   Conditional independence questions can be addressed using a rule which works with directed, undirected and partially directed FGs. To determine whether two variables are independent given a set of conditioned variables, consider all paths connecting the two variables. If all paths are blocked, the variables are conditionally independent. A path is blocked if one or more of the following conditions is satisfied:

   - One of the variables in the path is in the conditioning set.
   - One of the variables or factors in the path has two incoming edges that are part of the path (variable or factor collider), and neither the variable or factor nor any of its descendants are in the conditioning set.

3. Expressiveness of Graphical Models:

   It is clear that directed distributions can be represented as undirected distributions since one can associate each (normalised) factor of the joint distribution with a potential. For example, the distribution $p(a|b)p(b|c)p(c)$ can be factored as $\phi(a, b)\phi(b, c)$, where $\phi(a, b) = p(a|b)$ and $\phi(b, c) = p(b|c)p(c)$, with $Z = 1$. Hence every belief network can be represented as some MN by simple identification of the factors in the distributions. However, in general, the associated undirected graph (which corresponds to the moralised directed graph) will contain additional links and independence information can be lost. For example, the MN of $p(c|a, b)p(a)p(b)$ is a single clique $\phi(a, b, c)$ from which one cannot graphically infer that $a \perp\!\!\!\perp b$.