Case-Factor Diagrams

CIS 620 Spring 2005

Based on material by David McAllester, Michael Collins, and Fernando Pereira

Penn
University of Pennsylvania

Structured Probilistic Models

- Hidden Markov Models
- Dynamic Bayes Nets
- Bayes Nets
- Markov Random Fields (MRFs)
- Conditional Random Fields
- Probabilistic Context Free Grammars (PCFGs)
- Probabilistic Relational Models

Most can be viewed as special cases of MRFs. However, representing PCFGs as MRFs is problematic. We would like a unified formalism.

A Linear Boolean Model (LBM)

- A set of Boolean variables $\mathcal{V}$
- A feasible set F of truth assignments $\rho : \mathcal{V} \to \{0, 1\}$ perhaps defined by a SAT problem.
- A potential function Ψ assigning a potential $\Psi(x)$ to each Boolean variable x.

$$\Psi(\rho) = \sum_{z \in V} \Psi(z)\rho(z)$$

$$P(\rho | F, \Psi) = \frac{1}{Z(F, \Psi)}\, e^{-\Psi(\rho)} \qquad Z(F, \Psi) = \sum_{\rho \in F} \; e^{-\Psi(\rho)}$$

Penn
University of Pennsylvania

A Markov Random Field

- A set of variables (nodes).
- A set of local potential functions Ψ_k each depending on some (small) subset of the variables.

$$\Psi(\rho) = \sum_k \Psi_k(\rho)$$

$$P(\rho|\Psi) = \frac{1}{Z(\Psi)} e^{-\Psi(\rho)} \qquad Z(\Psi) = \sum_\rho e^{-\Psi(\rho)}$$

MRFs as LBMs

Given an MRF M we construct an LBM:

- A variable "$y = v$" for each variable y of M and possible value v of y.
- A constraint that exactly one of "$y = v_1$", ..., "$y = v_n$" is true.
- A Boolean variable "$k, y_1 = v_1 \wedge \cdots \wedge y_m = v_m$" for each local energy term Ψ on variables $y_1, \ldots, y_m$.
- **A constraint that "$k, y_1 = v_1 \wedge \cdots \wedge y_m = v_m$" is true if and only if each of "$y_1 = v_1$",...,"$y_m = v_m$" is true.**

$$\Psi(\text{"}k, y_1 = v_1 \wedge \cdots \wedge y_m = v_m\text{"}) = \Psi_k(v_1, \ldots, v_m)$$

Weighted Context Free Grammars

A CFG in Chomsky normal form is a set of productions of the following form.

$$X \rightarrow Y Z$$
$$X \rightarrow a$$

A parse tree is a tree each node of which is labeled by a production of the grammar in the standard way.

In a weighted CFG each production $X \rightarrow \gamma$ is assigned an energy (weight) $\Psi(X \rightarrow \gamma)$.

PCFGs as LBMs

Binary Variables:

- "$X_{i,j}$" for each nonterminal X and string positions i and j.
- "$X_{i,k} \rightarrow Y_{i,j} Z_{j,k}$" for each production $X \rightarrow YZ$ in the grammar and string positions i, j, and k.
- "$X_{i,i+1} \rightarrow a$" for each production $X \rightarrow a$ in the grammar and string position i.

$$\begin{aligned}
\Psi(\text{"}X_{i,j}\text{"}) &= 0 \\
\Psi(\text{"}X_{i,k} \rightarrow Y_{i,j} Z_{j,k}\text{"}) &= \Psi(X \rightarrow YZ) \\
\Psi(\text{"}X_{i,i+1} \rightarrow a\text{"}) &= \Psi(X \rightarrow a)
\end{aligned}$$

Recursive Conditioning (Darwiche)

- Case on the value of a variable.

- Factor into independent subproblems when possible.

- Cache answers to subproblems.

Case-Factor Diagrams (CFDs)

A case-factor diagram is defined by the following grammar.

$$D ::= \mathsf{case}(x, D_1, D_2) \mid \mathsf{factor}(D_1, D_2) \mid \mathsf{unit} \mid \mathsf{empty}$$

- In $\mathsf{case}(x, D_1, D_2)$ the variable x does not occur in D_1 or D_2.
- In $\mathsf{factor}(D_1, D_2)$ no variable occurs in both D_1 or D_2.

$$\mathsf{case}(\langle z_1, D_1\rangle, \langle z_2, D_2\rangle, \ldots, \langle z_m, D_m\rangle) \equiv \mathsf{case}(z_1, D_1, \mathsf{case}(z_2, D_2, \ldots \mathsf{case}(z_m, D_m, \mathsf{empty}) \ldots))$$

Meaning of a CFD

$$
\begin{aligned}
F(\mathsf{unit}) &= \{\overline{0}\} \\
F(\mathsf{empty}) &= \emptyset \\
F(\mathsf{case}(x, D_1, D_2)) &= F(D_1)[x := 1] \cup F(D_2) \\
F(\mathsf{factor}(D_1, D_2)) &= F(D_1) \vee F(D_2)
\end{aligned}
$$

$$
\begin{aligned}
\sigma[x := 1](y) &= \begin{cases} 1 & x = y \\ \sigma(y) & \text{otherwise} \end{cases} \\
\Sigma[x := 1] &= \{\sigma[x := 1] : \sigma \in \Sigma\} \\
(\sigma \vee \rho)(x) &= \sigma(x) \vee \rho(x) \\
\Sigma \vee \Delta &= \{\sigma \vee \rho : \sigma \in \Sigma, \rho \in \Delta\}
\end{aligned}
$$

Penn
University of Pennsylvania

PCFGs as CFDs

We define the CFD $D(\text{"}X_{i,k}\text{"})$ such that $F(D(\text{"}X_{i,k}\text{"}))$ represents the parse trees of the span from i to $k-1$ with root nonterminal X.

$$D(\text{"}X_{i,k}\text{"}) = \mathsf{case}(\text{"}X_{i,k}\text{"}, B(\text{"}X_{i,k}\text{"}), \mathsf{empty})$$

$$B(\text{"}X_{i,k}\text{"}) = \mathsf{case}(\langle b_1,\ B(b_1)\rangle, \ldots, \langle b_n,\ B(b_n)\rangle)$$

where the variables b_p are all possible branch variables of the form "$X_{i,k} \to Y_{i,j} Z_{j,k}$", and $B(\text{"}X_{i,k} \to Y_{i,j} Z_{j,k}\text{"}) = \mathsf{factor}(D(\text{"}Y_{i,j}\text{"}), D(\text{"}Z_{j,k}\text{"}))$.

$$B(\text{"}X_{i,i+1}\text{"}) = \begin{cases} \mathsf{case}(\text{"}X_{i,i+1} \to a_i\text{"}, \mathsf{unit}, \mathsf{empty}) & \text{if } X \to a_i \in G \\ \mathsf{empty} & \text{otherwise} \end{cases}$$

CFDs for MRFs

Here we define a CFD representation of the feasible set for the LBM constructed in section 3. Consider the problem of computing $Z(M)$ for an MRF M. We assume that the variables of M have been given in a fixed order $y_1, y_2, \ldots, y_n$. The assignments to these variables form a tree whose root has a branch for each value of y_1, the next level branches for each value of y_2 and so on. As variables are assigned, however, the residual hypergraph defined by the energy terms often factors into disjoint sets of terms on disjoint sets of variables. So one can compute $Z(M)$ by factoring the residual problem when possible and, if no factoring is possible, casing out on the value of the next variable (after which more factoring may be possible). This "case-factor process" determines a set of subproblems. The nodes (subexpressions) in the CFD representation of the MRF correspond to the subproblems that arise in this way. Each such subproblem is defined by a subset Σ of the energy terms and a partial assignment ρ to (some of) the variables occurring in Σ.

CFDs and Tree Width

Let M be an MRF with N local potential functions, d possible values for each node, and tree width w.

Let $\langle D(M), \Psi(M) \rangle$ be the case-factor diagram representation of M.

Theorem: The number of nodes in $D(M)$ is $O(Nd^w)$.

Viterbi for CFDs

$$\Psi^*(D, \Psi) = \min_{\rho \in F(D)} \Psi(\rho)$$

We can compute $\Psi^*(D, \Psi)$ using the following equations.

$$\Psi^*(\mathsf{case}(z, D_1, D_2), \Psi) = \min \begin{pmatrix} \Psi(z) + \Psi^*(D_1, \Psi), \\ \Psi^*(D_2, \Psi) \end{pmatrix}$$

$$\Psi^*(\mathsf{factor}(D_1, D_2), \Psi) = \Psi^*(D_1, \Psi) + \Psi^*(D_2, \Psi)$$

$$\Psi^*(\mathsf{unit}, \Psi) = 0$$

$$\Psi^*(\mathsf{empty}, \Psi) = +\infty$$

Inside values for CFDs

$$Z(\text{case}(x, D_1, D_2), \Psi) = e^{-\Psi(x)} Z(D_1, \Psi) + Z(D_2, \Psi)$$
$$Z(\text{factor}(D_1, D_2), \Psi) = Z(D_1, \Psi) Z(D_2, \Psi)$$
$$Z(\text{unit},\ \Psi) = 1$$
$$Z(\text{empty},\ \Psi) = 0$$

Penn
University *of* Pennsylvania

Outside for CFDs

The outside value $O(D, C, \Psi)$ is the total weight of the "contexts" in which D appears.

$$O(D', C, \Psi) = \sum_{\mathsf{case}(z, D', D'') \preceq C} O(\mathsf{case}(z, D' D''), C, \Psi) e^{-\Psi(z)}$$

$$+ \sum_{\mathsf{case}(z, D'', D') \preceq C} O(\mathsf{case}(z, D'', D'), C, \Psi)$$

$$+ \sum_{\mathsf{factor}(D', D'') \preceq C} O(\mathsf{factor}(D', D''), C, \Psi) Z(D'', \Psi)$$

$$+ \sum_{\mathsf{factor}(D'', D') \preceq C} O(\mathsf{factor}(D'', D'), C, \Psi) Z(D'', \Psi)$$

Conditional Probabilities for CFDs

$$P(x = 1 | D, \Psi) \;=\; \frac{Z(D, \Psi, x = 1)}{Z(D, \Psi)}$$

$$Z(D, \Psi, x = 1) \;=$$

$$\sum_{\mathsf{case}(x, D', D'') \preceq D} \left(\begin{array}{l} O(\mathsf{case}(x, D', D''), D, \Psi) \\ e^{-\Psi(x)} \\ Z(D', \Psi) \end{array} \right)$$

Conclusions and Future Work

- Case-Factor Diagrams provide an abstract simple model subsuming MRFs of low tree width and PCFGs.

- A Unification of modeling formalisms allows greater generality in the statement of theorems and the construction of code.

  LBMs with feasible sets defined by SAT formulas can model arbitrary MRFs and have some advantages.

  Future work is planned for Loopy propagation and perhaps other graphical model algorithms for LBMs.

Penn
UNIVERSITY of PENNSYLVANIA