In *AAAI-93, 11th National Conference on Artificial Intelligence*, Washington, D.C.: Jul. 93, 790–795.

Estimating Probability Distributions over Hypotheses with Variable Unification

Dekai Wu*
Department of Computer Science
The Hong Kong University of Science & Technology
Clear Water Bay, Hong Kong
dekai@cs.ust.hk

Abstract

We analyze the difficulties in applying Bayesian belief networks to language interpretation domains, which typically involve many unification hypotheses that posit variable bindings. As an alternative, we observe that the structure of the underlying hypothesis space permits an approximate encoding of the joint distribution based on marginal rather than conditional probabilities. This suggests an *implicit binding* approach that circumvents the problems with explicit unification hypotheses, while still allowing hypotheses with alternative unifications to interact probabilistically. The proposed method accepts arbitrary subsets of hypotheses and marginal probability constraints, is robust, and is readily incorporated into standard unification-based and frame-based models.

1 Introduction

The application of Bayesian belief networks (Pearl 1988) to natural language disambiguation problems has recently generated some interest (Goldman & Charniak 1990; Charniak & Goldman 1988, 1989; Burger & Connolly 1992). There is a natural appeal to using the mathematically consistent probability calculus to combine quantitative degrees of evidence for alternative interpretations, and even to help resolve parsing decisions.

However, to formulate disambiguation problems using belief networks requires an unusual form of hypothesis nodes. Natural language interpretation models (as well as many others) employ the unification operation to combine schemata; this is realized alternatively as slot-filling, role-binding, or attribute co-indexing in feature structures. Specifically, in this paper we are concerned with the class of problems where the input context introduces a number of possible conceptual entities but the relationships between them must be inferred. This phenomenon is ubiquitous in language, for example in prepositional and adverbial attachment, adjectival modification, and nominal compounds. The process of resolving such an ambiguity corresponds to unifying two variables (or role bindings or slot fillers).

In extending the models to Bayesian belief networks, unification operations are translated to hypothesis nodes—for example *(patient g3)=r2* in figure 1—that sit alongside "regular" hypotheses concerning the features of various conceptual entities. The incorporation of binding hypotheses introduces a modelling difficulty in the context of belief networks. The strength of the unification-based paradigm rests precisely in the relatively symmetric role binding, which is subject to no constraints other than those explicitly given by the linguist or knowledge engineer. However, we argue in section 2 that this same characteristic directly resists models based on the notion of conditional independence, in particular belief networks.

In section 3 we re-analyze the structure of the underlying hypothesis space and its joint distribution. This formulation leads to an alternative approach to approximation, proposed in section 4. A natural language application dealing with nominal compound interpretation is outlined in section 5.

2 Unification Resists Conditional Independence

In conditional independence networks, the values of some hypotheses are permitted to influence others but the paths of influence are restricted by the graph, thus providing computational leverage. In the extreme, a completely connected graph offers no computational shortcuts; instead, to improve performance a distribution should be graphed with the lowest possible connectivity. In general, conditional independence networks have been applied in highly structured domains where low-connectivity approximations can be accurate. The types of domains that invite unification-oriented representations, however, resist low-connectivity approximations, because binding hypotheses have a high inherent degree of interdependence.

Typically in such a domain, there will be some number

*Preparation of this paper was partially supported by the Natural Sciences and Engineering Research Council of Canada while the author was a postdoctoral fellow at the University of Toronto. Much of this research was done at the Computer Science Division, University of California at Berkeley and was sponsored in part by the Defense Advanced Research Projects Agency (DoD), monitored by the Space and Naval Warfare Systems Command under N00039-88-C-0292, the Office of Naval Research under N00014-89-J-3205, the Sloan Foundation under grant 86-10-3, and the National Science Foundation under CDA-8722788.

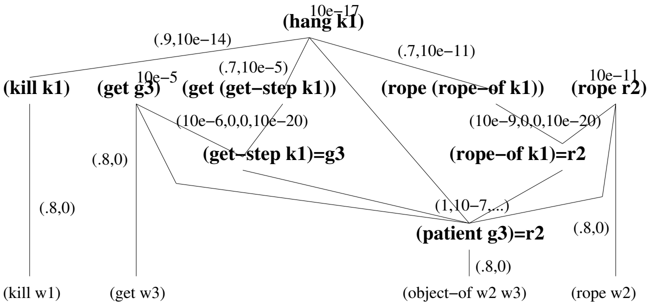


Figure 1: Example belief net from Goldman & Charniak (1991).

n of "free" variables $a, b, c \ldots$ that are potentially unifiable with others. A unification hypothesis is of the form $a = b$, and there are $m = \binom{n}{2}$ such hypotheses. *A priori* knowledge, like selectional restrictions, may help rule out some of these hypotheses, but many bindings will remain possible and we'll assume here that all unification hypotheses have nonzero probability. A *joint hypothesis* is an assignment of truth values to each of the m unification hypotheses.[1] The number of legal joint hypotheses is less than 2^m because of the dependence between hypotheses. For example, if $a = c$ and $b = c$ are true, then $a = b$ must also be true. In fact the number of legal joint hypotheses is equal to the number of possible partitionings of a set of n elements. Figure 2 shows the legal joint hypotheses when $n = 4$.

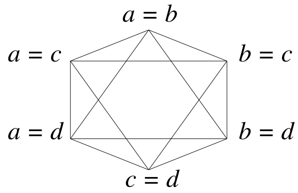
Hypotheses	*Legal assignments*														
$a = b$	0	0	0	0	0	0	1	0	0	1	0	0	1	1	1
$a = c$	0	0	0	0	0	1	0	0	1	0	0	1	0	1	1
$a = d$	0	0	0	0	1	0	0	1	0	0	0	1	1	0	1
$b = c$	0	0	0	1	0	0	0	1	0	0	1	0	0	1	1
$b = d$	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1
$c = d$	0	1	0	0	0	0	0	0	0	1	1	1	0	0	1

Figure 2: The legal joint hypotheses for $n = 4$. Each column shows a permissible truth value assignment.

Now consider the dependence relationships between unification hypotheses. The probabilities of $a = c$ and $b = c$ are not independent since they may be affected by the value of $a = b$; if $a \neq b$ then all events where $a = c$ and $b = c$ are ruled out. However, it is possible for $a = c$ and $b = c$ to be conditionally independent given $a = b$, which can be modelled by

$$a = c \;\text{———}\; a = b \;\text{———}\; b = c$$

By symmetry, all three hypotheses must be connected. This extends to larger n, so if $n = 4$, then if $a = d$ and $b = d$ are conditionally independent, it must also be conditioned on $a = b$:



In general, any pair of unification hypotheses that involve a common variable must be connected. Thus for n variables, the total number of links is

$$l = n \cdot \binom{n-1}{2} = \frac{n(n-1)(n-2)}{2} = m(n-2)$$

which is $\Theta(n^3)$ or $\Theta(m^{3/2})$. This is better than a completely connected network which would be $\Theta(n^4)$ or $\Theta(m^2)$ but there are many loops nonetheless, so evaluation will be expensive. By symmetry, each of the m hypotheses is of degree

$$\frac{2l}{m} = 2(n-2)$$

and any clustering of variables will be subject to this bound.

We conclude that in domains where unification hypotheses are relatively unconstrained, the connectivity of conditional independence networks is undesirably high. This means that it is difficult to find efficient conditional probability representations that accurately approximate the desired joint probability distributions. Therefore, in the next section we reconsider the event space that underlies the joint distribution.

3 Back to Basics

Since conditional probabilities do not lend themselves well to representations involving unification hypotheses, we now examine the structure of the joint hypothesis space. Before, we considered the unification hypotheses in explicit form because we sought conditional independence relationships between them. Having abandoned that objective, here we instead consider the feature structures (or frames) that result from assigning truth values to the unification hypotheses. In other words, the unification hypotheses are left implicit, reflected by co-indexed variables (roles) in feature structures.

Figure 3 depicts the qualitative structure of the joint hypothesis space, which forms a semilattice hierarchy. We now take into consideration not only the implicit unification hypotheses, but also implicit hypotheses that specialize the features on the variables; for example, a variable of type a may be specialized to the subtype b. Each box denotes the feature structure that results from some combination of truth values over a subset of unification hypotheses and specialization hypotheses. Each small shaded box denotes a joint hypothesis specifying the truth values over *all* unification and specialization hypotheses. Thus the distinction between the shaded and non-shaded boxes is a kind of type-token distinction where the shaded boxes are tokens. Notice furthermore that role specialization and unification are intertwined: a role of type z results when a type x role and a type y role are conjoined by unifying their fillers.

[1] We ignore all other types of hypotheses in this section's analysis.

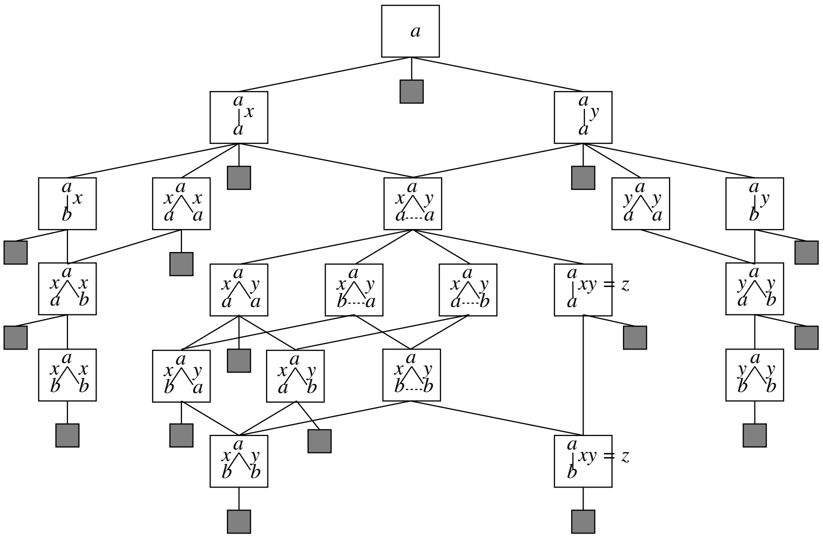


Figure 3: Simplified partial abstraction lattice for feature structures. The type b is a subtype of a; the role type z is a composite role equivalent to the conjunction xy. A dashed line indicates that the variables are "free" to be unified.

In principle, the joint distribution would be completely specified if we could enumerate the probabilities over the (shaded) tokens. We saw in the previous section that conditional probabilities are not well suited for approximately summarizing distributions over this space, because there is no way to discard large numbers of binding dependencies in the general case. However, there is another straightforward way to store distributional information, namely to record *marginal* probabilities over the abstract (non-shaded) types, i.e., the sums of probabilities over all descendant leaves. To summarize the distribution approximately, a selected subset of the marginal probabilities can be stored. Theoretically, a set of marginal probabilities induces an equivalent set of conditional probabilities over the same lattice, though it may be an unreasonably large set. If there are any independence relationships to be exploited, equivalently a subset of marginal probabilities can be omitted and the maximum-entropy principle (Jaynes 1979) can be applied to reconstruct the joint distribution.

The advantages of this formulation are: (1) fewer parameters are required since it does not encode redundant distributional information in multiple dependent conditional probabilities, (2) consistency is easier to maintain because the interdependent unification hypotheses are not explicit, (3) it facilitates an alternative *structural* approximation method for computing a conditional distribution of interest, as discussed in the next section.

4 An Approximation Based on Marginal Constraints

By itself, the marginal probability formulation can reduce probability storage requirements but does not improve computation cost. Computing maximum entropy distributions subject to large numbers of marginal constraints is infeasible in the general case. However, in many applications, including language interpretation, the input cues are sufficient to eliminate all but a relatively small number of hypotheses. Only the distribution over these hypotheses is of interest. Moreover, the input cues may suffice to preselect a subset of relevant marginal probability constraints.

The proposed method takes advantage of these factors by dynamically creating a secondary marginal probability formulation of the same form as that above, but with far fewer constraints and hypotheses, thereby rendering the entropy maximization feasible. In the secondary formulation, only details within the desired hypothesis and constraint space are preserved. Outside this space, the minimum possible number of "dummy" events are substituted for multiple hypotheses that are not of interest. It turns out that one dummy event is required for each marginal constraint. Let $\mathcal{Q}$ be the set of token feature structures and $\mathcal{G}$ is the set of type feature structures, and $\mathcal{F} \stackrel{\text{def}}{=} \mathcal{G} \cup \mathcal{Q}$. Suppose $\mathcal{H} = \{h_1, \ldots, h_i, \ldots, h_H\} \subset \mathcal{Q}$ are the candidate hypotheses, and suppose $\mathcal{M} = \{m_1, \ldots, m_j, \ldots, m_M\} \subset \mathcal{G}$ are the abstract class types that have been preselected as being relevant, with associated marginal probabilities $P_{m_j} = P(m_j)$. Denote by $\sqsubset$ the partial ordering induced on $\mathcal{H} \cup \mathcal{M}$ by the subsumption semilattice on f-structure space.

Then we define the secondary formulation as follows. Let the set of dummy events be $\mathcal{D} = \{d_1, \ldots, d_j, \ldots, d_M\}$, one for each marginal constraint. Define $\dot{\mathcal{F}} \stackrel{\text{def}}{=} \mathcal{H} \cup \mathcal{M} \cup \mathcal{D}$ to be the *approximate event space*, and define $\dot{\mathcal{H}} \stackrel{\text{def}}{=} \mathcal{H} \cup \mathcal{D}$ to be the *approximate hypothesis space*. We construct the *approximate ordering relation* $\dot{\sqsubset}$ over $\dot{\mathcal{F}}$ according to:

$$\begin{cases} a \mathrel{\dot{\sqsubset}} b, & \text{if} \begin{cases} a \sqsubset b; a, b \in \mathcal{F} \\ a = m_j; b = d_j \\ a \sqsubset c; c = m_j; b = d_j \end{cases} \\ a \mathrel{\dot{\not\sqsubset}} b, & \text{otherwise} \end{cases}$$

Let $\dot{P}_{m_j}$ be the marginal probability constraints on $\mathcal{F}$. We use P_{m_j} as estimators for $\dot{P}_{m_j}$. (Of course, since the event space has been distorted by the structural dummy event approximation, actually $P_{m_j} \neq \dot{P}_{m_j}$.)

To estimate the distribution over the hypotheses of interest, along with the dummy events, we compute $\hat{P}_{h_i}$ and $\hat{P}_{d_i}$ such that

(1) $$\sum_{q \in \dot{\mathcal{H}}} \hat{P}_q = 1$$

while maximizing the entropy

(2) $$E = -\sum_{q \in \dot{\mathcal{H}}} \hat{P}_q \log \hat{P}_q$$

subject to the marginal constraints

(3) $$\sum_{q : q \in \dot{\mathcal{H}}, m_j \sqsubseteq q} \hat{P}_q = \dot{P}_{m_j}$$

Technical details of the solution are given in Appendix A.

Note that unlike methods for finding maximum *a posteriori* assignments (Charniak & Santos Jr. 1992) which returns the probability for the most probable joint assignment, the objective here is to evaluate the conditional distribution over a freely chosen set of joint hypothesis assignments and marginal constraints.

One of the strengths of AME is robustness when arbitrarily chosen marginals are discarded. Arithmetic inconsistencies do not arise because the dummy events absorb any discrepancies arising from the approximation. For example, if C through F are discarded from figure 4(a), then $P(A) + P(B) < P(G)$, but the remaining probability weight is absorbed by G's dummy event in (b). The ability to handle arbitrary subpartitions of the knowledge base is important in practical applications, where many different heuristics may be used to preselect the constraints dynamically. In contrast, when there are dependent unification hypotheses in a belief network, discarding conditional probability matrices can easily lead to networks that have no solution.

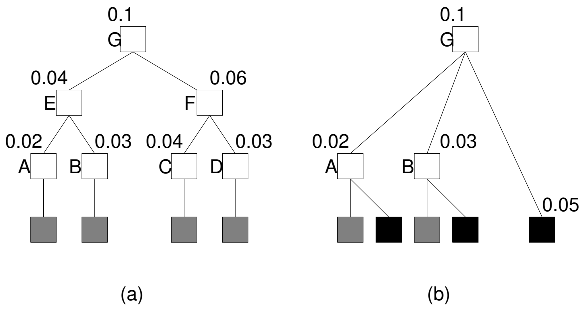


Figure 4: Robust handling of discarded marginal constraints. (a) Original KB fragment. (b) Dummy event (black) absorbing discrepancy caused by discarding marginals.

5 A Nominal Compound Interpretation Example

In this section we summarize an example from the language interpretation domain that drove the development of AME, a more detailed discussion of which is found in the companion paper (Wu 1993a). Space does not permit a description of the semantics and lexical models; see Wu (1992, 1993b). Although our modelling objectives arise solely from disambiguation problems, we believe the foregoing discussion applies nonetheless to other structured domains involving highly interdependent variable bindings with uncertainty.

The example task here is to interpret the nominal compound *coast road*,[2] which in null context most likely means a *road in coastal area* but, particularly in other contexts, can also mean other things including a *road leading to coastal area*, a *coasting road* amenable to coasting, and *Highway 1*. As is typical with novel nominal compounds, interpretation requires a wide range of knowledge. Figure 5 shows the fairly standard feature-structure notation we use to encode such knowledge; the marginal probabilities in (a) and (b) are the primary representational extension.

During interpretation, a hypothesis network as in figure 6 is dynamically constructed. Each node corresponds to a marginal constraint from the knowledge base, of the form figure 5(a)—(b). Ignoring the boldface marginals for now, the probabilities $P($*coast and road*$)$ and $P($*coast and coastal road*$)$ indicate that when thinking about roads, it is the subcategory of roads running along the coast that is frequently thought of. Similarly $P($*coastal road*$)$ and $P($*Highway 1*$)$ model a non-West Coast resident who does not frequently specialize coastal roads to Highway 1. Together, $P($*L:coast*$)$, $P($*C:coast:seacoast*$)$, and $P($*C:coast:coasting accomplishment*$)$ indicate that the noun *coast* more frequently designates a seacoast rather than an unpowered movement. Finally, $P($*C:NN:containment*$)$ indicates that the noun-noun construction signifies containment twice as often as $P($*C:NN:linear order locative*$)$.

Figure 6 summarizes the results of the baseline run and four variants, from a C implementation of AME. In the base run labelled "**0:**", the AME estimate of the conditional distribution assigns highest probabilities to *road in coastal area* and *road along coastline* (features distinguishing these two hypotheses have been omitted). The next run "**1:**" demonstrates what would happen if "*coast*" more often signified *coasting accomplishment* rather than *seacoast*: the *coasting road* hypothesis dominates instead. In "**2:**" the noun-noun construction is assumed to signify linear order locatives more frequently than containment. The marginals in "**3:**" effectively reduce the conditional probability of thinking of roads along the seacoast, given one is thinking of roads in the context of seacoasts. The West Coast res-

[2]From the Brown corpus (Kučera & Francis 1967). Our approach to nominal compounds is discussed in Wu (1990), which proposes the use of probability to address long-standing problems from the linguistics literature (e.g., Lees 1963; Downing 1977; Levi 1978; Warren 1978; McDonald 1982).

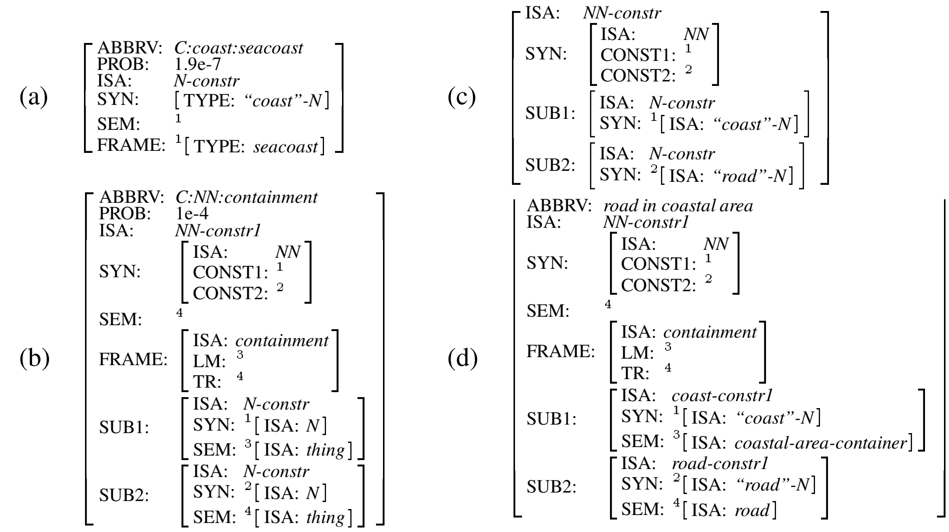
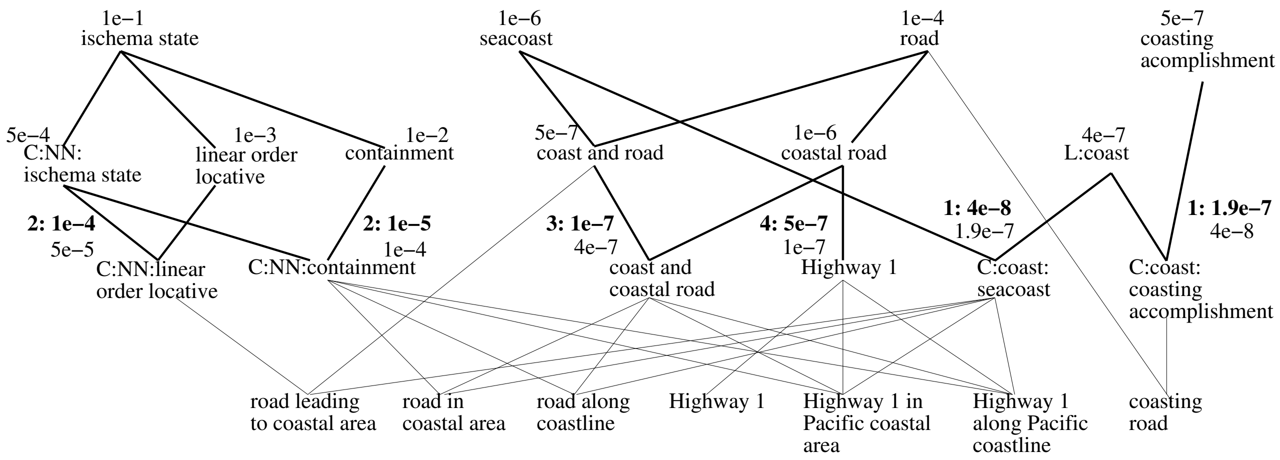


Figure 5: Feature-structures for (a) the noun *coast* signifying a seacoast, (b) a noun-noun construction signifying a containment schema, (c) an input form, and (d) a full interpretation hypothesis (the floor brackets indicate a token as opposed to type).



	road leading to coastal area	road in coastal area	road along coastline	Highway 1	Highway 1 in Pacific coastal area	Highway 1 along Pacific coastline	coasting road
0:	0.046524	**0.37215**	**0.37215**	0.00025822	0.074419	0.074419	0.060089
1:	0.015757	0.12605	0.12605	0.00061625	0.02521	0.02521	**0.6811**
2:	**0.38339**	0.15336	0.15336	0.0010636	0.030672	0.030672	0.24748
3:	**0.40849**	0.20422	0.20422	0.00056666	0.025527	0.025527	0.13145
4:	0.010205	0.081579	0.081579	0.00028371	**0.40657**	**0.40657**	0.01321

Figure 6: Estimated conditional distributions for five runs on *coast road* with varying marginal constraints. Dummy events have been omitted.

ident is modelled in "**4:**" by an increase in the marginal $P(\mathit{Highway\ 1})$.

6 Conclusion

We have discussed the difficulties encountered in applying Bayesian belief networks to domains like language interpretation, which involve unification hypotheses over "free" variables. We observed that the structure of the underlying joint hypothesis space permits an alternative approximate encoding based on marginal rather than conditional probabilities. This *implicit binding* formulation facilitates a structural approximation method. For many applications, language interpretation in particular, the structural approximation is adequate and flexibility in handling unification hypotheses is quite important, whereas exact probability distribution computation is unnecessary. The method is robust and incorporates readily into unification- or frame-based models.

Acknowledgements

I am indebted to Robert Wilensky, Jerome Feldman, and the members of the BAIR and L_0 groups for many valuable discussions, as well as Graeme Hirst, Geoff Hinton, and their respective groups.

A Details of the Entropy Maximization

To solve the constrained maximization problem in equations (1)—(3), we define a new energy function with Lagrange multipliers, J, to be maximized:

$$
\begin{aligned}
J &\stackrel{\text{def}}{=} E + \sum_{j=1}^{M} \lambda_j (\dot{P}_{m_j} - \sum_{q:q \in \dot{\mathcal{H}}, m_j \dot{\sqsubseteq} q} \hat{P}_q) \\
&= -\sum_{q \in \dot{\mathcal{H}}} \hat{P}_q \log \hat{P}_q + \sum_{j=1}^{M} \lambda_j (\dot{P}_{m_j} - \sum_{q:q \in \dot{\mathcal{H}}, m_j \dot{\sqsubseteq} q} \hat{P}_q)
\end{aligned}
$$

This method is a modified version of Cheeseman's (1987) method, which applied only to feature vectors. Observe that setting the gradients to zero gives the desired conditions:

$$
\begin{aligned}
\nabla_\lambda J = 0 &\Rightarrow \tfrac{\partial J}{\partial \lambda_j} = 0; 1 \le j \le M \\
&\Rightarrow \text{expresses all marginal constraints} \\
\nabla_{\hat{\mathbf{P}}} J = 0 &\Rightarrow \tfrac{\partial J}{\partial \hat{P}_q} = 0; q \in \dot{\mathcal{H}} \\
&\Rightarrow \text{maximizes entropy}
\end{aligned}
$$

Since the partials with respect to $\hat{\mathbf{P}}$ are

$$\frac{\partial J}{\partial \hat{P}_q} = -\log \hat{P}_q - \sum_{j:m_j \dot{\sqsubseteq} q} \lambda_j$$

then at $\nabla_{\hat{\mathbf{P}}} J = 0$,

$$\log \hat{P}_q = -\sum_{j:m_j \dot{\sqsubseteq} q} \lambda_j$$

Defining $\omega_j \stackrel{\text{def}}{=} e^{-\lambda_j}$,

$$\hat{P}_q = \prod_{j:m_j \dot{\sqsubseteq} q} \omega_j$$

the original marginal constraints become

$$\dot{P}_{m_j} = \sum_{q:m_j \dot{\sqsubseteq} q} \prod_{k:m_k \dot{\sqsubseteq} q} \omega_k$$

which can be rewritten

$$\dot{P}_{m_j} - \sum_{q:m_j \dot{\sqsubseteq} q} \prod_{k:m_k \dot{\sqsubseteq} q} \omega_k = 0$$

The last expression is solved using a numerical algorithm of the following form:

1. Start with a constraint system $X \leftarrow \{\}$ and an estimated ω vector $\langle\rangle$ of length zero.
2. For each constraint equation,
   (a) Add the equation to X and its corresponding ω_i term to $\langle \omega_1, \ldots, \omega_{i-1}, \omega_i \rangle$.
   (b) Repeat until $\langle \omega_1, \ldots, \omega_i \rangle$ settles, i.e., the change between iterations falls below some threshold:
      1. For each equation in X constraining $\dot{P}_{m_j}$, solve for the corresponding ω_j assuming all other ω values have their current estimated values.

References

BURGER, JOHN D. & DENNIS CONNOLLY. 1992. Probabilistic resolution of anaphoric reference. In *AAAI Fall Symposium on Probabilistic NLP*, Cambridge, MA. Proceedings to appear as AAAI technical report.

CHARNIAK, EUGENE & ROBERT GOLDMAN. 1988. A logic for semantic interpretation. In *Proceedings of the 26th Annual Conference of the Association for Computational Linguistics*, 87–94.

CHARNIAK, EUGENE & ROBERT GOLDMAN. 1989. A semantics for probabilistic quantifier-free first-order languages, with particular application to story understanding. In *Proceedings of IJCAI-89, Eleventh International Joint Conference on Artificial Intelligence*, 1074–1079.

CHARNIAK, EUGENE & EUGENE SANTOS JR. 1992. Dynamic MAP calculations for abduction. In *Proceedings of AAAI-92, Tenth National Conference on Artificial Intelligence*, 552–557, San Jose, CA.

CHEESEMAN, PETER. 1987. A method of computing maximum entropy probability values for expert systems. In *Maximum-entropy and Bayesian spectral analysis and estimation problems*, ed. by Ray C. Smith & Gary J. Erickson, 229–240. Dordrecht, Holland: D. Reidel. Revised proceedings of the Third Maximum Entropy Workshop, Laramie, WY, 1983.

DOWNING, PAMELA. 1977. On the creation and use of English compound nouns. *Language*, 53(4):810–842.

GOLDMAN, ROBERT P. & EUGENE CHARNIAK. 1990. A probabilistic approach to text understanding. Technical Report CS-90-13, Brown Univ., Providence, RI.

JAYNES, E. T. 1979. Where do we stand on maximum entropy. In *The maximum entropy formalism*, ed. by R. D. Levine & M. Tribus. Cambridge, MA: MIT Press.

KUČERA, HENRY & W. NELSON FRANCIS. 1967. *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

LEES, ROBERT B. 1963. *The grammar of English nominalizations*. The Hague: Mouton.

LEVI, JUDITH N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

MCDONALD, DAVID B. 1982. Understanding noun compounds. Technical Report CMU-CS-82-102, Carnegie-Mellon Univ., Dept. of Comp. Sci., Pittsburgh, PA.

PEARL, JUDEA. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

WARREN, BEATRICE. 1978. *Semantic patterns of noun-noun compounds*. Gothenburg, Sweden: Acta Universitatis Gothoburgensis.

WU, DEKAI. 1990. Probabilistic unification-based integration of syntactic and semantic preferences for nominal compounds. In *Proceedings of the Thirteenth International Conference on Computational Linguistics*, volume 2, 413–418, Helsinki.

WU, DEKAI, 1992. *Automatic inference: A probabilistic basis for natural language interpretation*. University of California at Berkeley dissertation. Available as UC Berkeley Computer Science Division Technical Report UCB/CSD 92/692.

WU, DEKAI. 1993a. Approximating maximum-entropy ratings for evidential parsing and semantic interpretation. In *Proceedings of IJCAI-93, Thirteenth International Joint Conference on Artificial Intelligence*, Chamberry, France. To appear.

WU, DEKAI. 1993b. An image-schematic system of thematic roles. In *Proceedings of PACLING-93, First Conference of the Pacific Association for Computational Linguistics*, Vancouver. To appear.