The Equivalence of the Model of McCallum and Wellner (2003) to Minimum Cut and Maximum Inter-Cluster Similarity

Jason D. M. Rennie
jrennie@alum.mit.edu

October 18, 2004

1 Introduction

Likelihood Maximization of the binary indicator variables, $\{y_{ij}\}$, in the model of McCallum and Wellner (2003) can equivalently be viewed as (1) Minimum Cut and (2) Maximum Inter-Cluster Similarity.

2 Minimum Cut Equivalence

Let C be the set of edges involed in the "cut." If C is empty, then the log-likelihood of the model is

$$\log P(C = \emptyset | \vec{x}) = \sum_{i,j,k} \lambda_k f_k(x_i, x_j, 1) - \log Z_x. \tag{1}$$

If C is not empty, we get

$$\log P(C|\vec{x}) = \sum_{i,j,k} \lambda_k f_k(x_i, x_j, 1) + \sum_{(i,j)\in C} \sum_k \lambda_k \Big[f_k(x_i, x_j, 0) - f_k(x_i, x_j, 1) \Big] - \log Z_x. \tag{2}$$

The maximum likelihood objective maximizes this second quantity. However, we can subtract the likelihood with an empty cut without harm since it is constant. We get:

$$\max_C \log P(C|\vec{x}) \equiv \max_C \sum_{(i,j)\in C} \sum_k \lambda_k \Big[f_k(x_i, x_j, 0) - f_k(x_i, x_j, 1) \Big]. \tag{3}$$

In other words, maximum likelihood is equivalent to finding the minimum cut on the graph with edge weights $W_{ij} = \sum_k \lambda_k \Big[f_k(x_i, x_j, 1) - f_k(x_i, x_j, 0) \Big]$.

3 Maximum Inter-Cluster Similarity Equivalence

Similarly, we can show that maximum likelihood is equivalent to maximizing inter-cluster similarity on the graph with particular edge weights. Let E be the set of edges conntecting objects within the same cluster; i.e. two objects are in the same cluster if and only if the edge between them, e, is in E. If E is empty, then the log-likelihood of the model is

$$\log P(E = \emptyset|\vec{x}) = \sum_{i,j,k} \lambda_k f_k(x_i, x_j, 0) - \log Z_x. \tag{4}$$

If E is not empty, we get

$$\log P(E|\vec{x}) = \sum_{i,j,k} \lambda_k f_k(x_i, x_j, 0) + \sum_{(i,j)\in E} \sum_k \lambda_k \Big[f_k(x_i, x_j, 1) - f_k(x_i, x_j, 0) \Big] - \log Z_x. \tag{5}$$

The maximum likelihood objective maximizes this second quantity. However, we can subtract the likelihood of $E = \emptyset$ without harm since it is constant. We get:

$$\max_E \log P(E|\vec{x}) \equiv \max_E \sum_{(i,j)\in E} \sum_k \lambda_k \Big[f_k(x_i, x_j, 1) - f_k(x_i, x_j, 0) \Big]. \tag{6}$$

In other words, maximum likelihood is equivalent to maximum inter-cluster similarity on the graph with edge weights $W_{ij} = \sum_k \lambda_k \Big[f_k(x_i, x_j, 1) - f_k(x_i, x_j, 0) \Big]$. Note that these are the same edge weights as we found for the Minimum Cut Equivalence.

References

McCallum, A., & Wellner, B. (2003). Toward conditional models of identity uncertainty with application to proper noun coreference. *Proceedings of the IJCAI Workshop on Information Integration on the Web.*