INSIDE-OUTSIDE ESTIMATION MEETS DYNAMIC EM

Detlef Prescher

DFKI Language Technology Lab
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
prescher@dfki.de

October 17-19, 2001

INSIDE-OUTSIDE ESTIMATION

- The standard method for **unsupervised training** of a given PCFG.
- In contrast to supervised training (relying on tree banks), IO estimation is an **iterative procedure** (on free text), but
- **dynamic programming techniques** make IO estimation efficient (in contrast to state-of-the-art training methods for stochastic UBGs)

Problem: The theoretical behaviour of the **log-likelihood** (corpus probabilities resp.), while iteratively applying IO estimation:

- **Monotonicity** has been observed in many experiments, but as well as
- **convergence to stationary points** (especially to local maxima) has never been formally proved.

INSIDE-OUTSIDE ESTIMATION MEETS DYNAMIC EM

Solve the problems with IO trusting **"common wisdom"**:

"**IO is a dynamic programming-variant of EM**".

I.e. put IO estimation into the context of the theoretically well-known **expectation-maximization algorithm** and formally prove the following

Theorem: The results of EM, when applied to PCFGs, equal the results of IO estimation. Thus:

- IO inherits the **well-known convergence behaviour of EM**,
- EM for PCFGs inherits the **dynamic behaviour of IO**.

PROOF OF THE THEOREM

IO estimation (for a training corpus of sentences y):

$$\hat{p}(r) := \frac{\sum_y C_y(r)}{\sum_y C_y(A)} \qquad (A = \mathsf{lhs}(r))$$

using rule and category counts $C_y(r)$ and $C_y(A)$ based on **recursively defined inside and outside probabilities** associated to sentence spans.

EM estimation for PCFGs:

$$\hat{p}(r) = \frac{\sum_y p(.|y)\,[\ f_r\ ]}{\sum_y p(.|y)\,[\ f_A\ ]} \qquad (A = \mathsf{lhs}(r))$$

using **expected frequencies**: $p(.|y)\,[\ f_r\ ] := \sum_x p(x|y) \cdot f_r(x)$.

> **Key step:** resolve non-dynamics of EM proving $p(.|y)\,[\ f_r\ ] = C_y(r)$.

CONCLUSION

We have presented:

- a brief review of IO and EM estimation for PCFGs,
- a formal proof confirming “common wisdom” that **IO is a dynamic-programming variant of EM**,
- **a formal proof of the convergence behaviour of IO estimation** which is omitted in the original literature.