Information Extraction in Diverse Settings

Sandeep Satpal (05329004)
Guide : Prof. Sunita Sarawagi

July 26, 2006

Outline

- Introduction
- Motivation
- Problem Description
- Literature Survey
- Probable Approaches
- Experiments and Conclusion

Introduction

- Information Extraction
  - Identifying entities from the given sequence
- Named Entity Recognition
  - Finding label for each entity in sequence
- Part of Speech Tagging
  - Tags part of the text

Motivation and Problem Description

- Not enough training sequence
- Standard CRF fails
- Generalised CRF: save training time
- **Problem Description**
  - Building domain independent model

Literature Survey

- **Conditional Random Fields**
  - Probabilistic model for segmentation and labeling
  - $P(y|x) = \frac{1}{Z(x)} e^{WF(y,x)}$
  - $F(y,x) = \sum_i \mathbf{f}(\mathbf{y},\mathbf{x},i)$
  - $Z(x) = \sum_{y'} e^{WF(y',x)}$ is a global normalization factor
- **Semi-Markov CRF**
  - $P(s|x) = \frac{1}{Z(x)} e^{WG(s,x)}$
  - $G(x,s) = \sum_j^{|s|} g(j,x,s)$
  - $s_i =< t_j, u_j, y_j >$
  - $Z(x) = \sum_{y'} e^{WG(y',x)}$

Probable Approaches

- Semi-Supervised CRF
- Multiple CRFs
- Regularisation in CRFs
- Exploiting Local Regularities
- Other Approaches

Semi-Supervised CRFs

- Semi-supervised using Markov Random Fields
  - When instances of differing class label found near and same class label are far apart
  - Additinal feature and weights on pairs of examples
  - $f_{k'}(x_i, x_j, y_i, y_j)$
  - $P(y_l, y_u | x_l, x_u) = \frac{1}{Z(x)} e^{(\sum_i \sum_k w_k f_k(x_i, y_i) + \sum_{i<k} \sum_{k'} w_{k'} f_{k'}(x_i, x_j, y_i, y_j))}$

Semi-Supervised CRFs (contd...)

- Semi-supervised CRF for Improved Segmentation and Labeling
  - Log-likelihood for Standard CRF
    - $LL(W) = \sum_{i=1}^{N} logp_w(y_i|x_i) - U(W)$
  - Maximize overlap between $\tilde{p}(x)$ and $\sum_y p_w(y|x)\tilde{p}(x)$
  - $RL(W) = LL(W) + \gamma \sum_{i=N+1}^{M} \sum_y p_w(y|x_i) logp_w(y|x_i)$
- Semi-Supervised Technique partially solves the problem

Multiple CRFs

- Train Multiple CRFs: one for each domain
- Problems :
  - Time and Space Complexity
  - Domain dependent
- Single CRFs for multiple domain
- Problems:
  - Same set of features for different labels

Multiple CRFs (contd...)

- Cascading and Joint CRFs

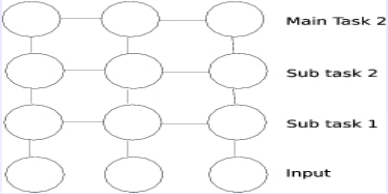


Figure: Graphical model for jointly decoded CRF

- New feature function : $f_k^i(s_{t-1}^i, s_t^i, s_t^{i-1}, x, t)$
- Cascading technique does not work for our problem

Regularisation in CRFs

- Log-likelihood
  - $LL(W) = \sum_k [W \cdot F(y_k, x_k) - log Z_w(x_k)] - \frac{||W||^2}{2 * \sigma^2}$
- Parameter-dependent
- Parameter free
- Minimax Problem Formulation
  - Minimize hinge loss : $\sum_i h(w, y_i, x_i) = \sum_i [1 - y_i w \cdot x_i]$
  - $h(w, y_i, x_i) = max[1 - y_i w \cdot (x_i * (1 - \alpha_i))]$
  - $\alpha_i \in \{0, 1\}$ and $\sum_j \alpha_{ij} = K$

Exploiting Local Regularities

- Special case : Same domain but different sources
- Learn Global Regularities
- Add features by learning "Locale" specific regularities
- Extension for our problem
  - Delete features from global learned model

Other Approaches

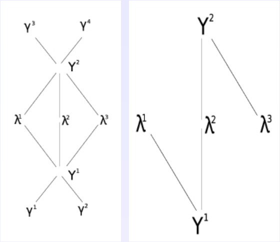
- Learn the model
- Find relevant features
    - Expected value on test sequence
    - Difference of Expected value
    - Weightage Difference (weight * difference)
- Remove irrelevant features
- Re-learn the model

Methods	*Token Level*			*Span Level*		
	Pre	Rec	F1-meas	Pre	Rec	F1-meas
Standard CRF	95.92	39.66	56.12	75.00	27.00	39.71
testExpected > 0	99.24	54.85	70.65	87.27	48.00	61.94
avgDifference ≥ 0.9	78.60	94.51	85.82	63.11	77.00	69.37
avgDifference ≥ 0.8	78.20	95.36	85.93	67.21	82.00	73.87
weightDifference ≥ 0.9	95.83	97.05	96.44	86.87	86.00	86.43

Conclusion and Future Plan

- **Conclusion**
  - Standard CRF does not directly applicable when applied to different domain
- **Work Plan**
  - Starts with simple case where whole sequence has one label
  - Then we will explore more complicated models like semi-markov CRF

Semi-Supervised CRFs (contd...)



(a) (b)

- Learning on Test Data: Leveraging "Unseen Features
  - $P_S(y|x, z, \gamma, \lambda^S)\alpha e^{(\gamma.yx+\lambda^S.yz)}$