

## Lecture for Statistical Natural Language Processing

Keh-Yih Su and Jing-Shin Chang  
kysu@bdc.com.tw, shin@nlp.csie.ncnu.edu.tw

(2002/8/17-18)

Behavior Design Corporation (BDC)  
No. 5, 2F, Industrial East Road IV, Science-Based Industrial Park  
Hsinchu, Taiwan 30077, R.O.C.

Department of Computer Science and Information Engineering  
National Chi-Nan University  
Puli, Nantou, Taiwan 545, R.O.C.

## Preface

- This two-day lecture series is not a short course. It is not aimed to replace those courses that you are expected to learn from the school (e.g., Probability, Statistics and Stochastic Process, Information Theory, Pattern Recognition, and Natural Language Processing, etc.)
- It neither has the intention to provide the technical details of any specific technique or approach that you could read and learn from those related papers by yourself, unless we feel that it is essential for explaining the concept that we try to convey to the audience.
- This talk is mainly designed to cover, we hope, those basic concepts that are seldom touched in those technical papers that focus in attacking a specific problem, and also show how those concepts are applied to NLP (which are usually not mentioned in the general textbooks that are written for wide range readers).

## Preface (Cont.)

- No specific background in probability or NLP is assumed for the audience. However, it definitely helps you to quickly grasp the idea if you have already known something. We wish everybody can get something from our talk regardless of their background.
- The same topic might be touched in various parts with different depths, as the learning process will be more effective if it is conducted in an incremental way. (You cannot fully understand a topic without the background from other related areas.)
- Last, thanks Microsoft Research Asia (particularly Dr. Ming Zhou) for inviting us to give this short lecture series. Also thanks all participants for your interest in attending this lecture series. Hope this talk is conducted in the way that you expected.

2002/8/17-18

Keh-Yih Su & Jing-Shin Chang

Statistical NLP

3

## Table of Contents

### Day-1: Introduction to Statistical Natural Language Processing (mainly on Supervised Learning)

- Part I: Introduction (1)
  - ◆ Problems and Characteristics of Natural Language Processing
- Part II: Introduction (2)
  - ◆ What, When and Why Statistical Approach
- Part III: Basic Concepts and Background
  - ◆ Feature Space, Probability, Estimator, Stochastic Process, Data Set Classification, and Performance Measure
- Part IV: Typical Applications
  - ◆ Word Segmentation, Tagging, Selecting Parse Tree, Aligning Bilingual Corpus
- Part V: Techniques for Improving Performance
  - ◆ Smoothing, Class-Based Model, Adaptive Learning, Tips for Checking
- Part VI: Advanced Topics: SVM, ME
  - ◆ Support Vector Machine, Maximum Entropy Models
- Appendix: Related Techniques
  - ◆ Parameter Estimation, Fractional Factorial Experiment Design, Decision Tree

2002/8/17-18

Keh-Yih Su & Jing-Shin Chang

Statistical NLP

4

## Table of Contents (cont.)

### Day-2: Unsupervised Learning for Natural Language Processing

- Part I: Introduction
  - ◆ What and When for Unsupervised Learning, Why it is getting popular
- Part II: Basic Concepts and Background (using EM as an example)
  - ◆ Incomplete Data Space
  - ◆ Learnability
- Part III: Typical Unsupervised Learning Algorithms: Viterbi & EM
  - ◆ Procedures, Characteristics
- Part IV: Potential Traps & Source of Problems
  - ◆ Various Mismatches, Model Deficiencies, Local Maximum, and Over-fitting
- Part V: Suggested Strategies for Better Performance
  - ◆ Lessons Learned from Past Experience
  - ◆ Recommended Procedures for Unsupervised Learning
- Part VI: Advanced Topic: Co-Training
  - ◆ Basic Principles
  - ◆ Example: Chinese New Word Extraction
- References

2002/8/17-18

Keh-Yih Su & Jing-Shin Chang

Statistical NLP

5