Committee-based Decision Making in Probabilistic Partial Parsing

INUI Takashi * and INUI Kentaro * †

* Kyushu Institute of Technology
† PRESTO, Japan Science and Technology Corporation
Background

- Increasing availability of large tree banks
- Success of statistical approaches to parsing

However,

Improvements appear to be getting saturated

Therefore,

- Two new directions for extending the current probabilistic parsing techniques,
  - Probabilistic Partial Parsing
  - Committee-based decision making
Overview of today’s talk

- **Probabilistic partial parsing**
  - A probabilistic extension of partial parsing

- **Committee-based probabilistic partial parsing**
  1. Probabilistic voting
  2. Standardization
  3. Multiple voting

- **Experiments**
Bunsetsu phrase (BP)

In this talk

- The target language of experiments is Japanese
  ⇒ However,
- Our proposal is not limited to Japanese
- It should be able to be applied to other languages like English
Bunsetsu phrase (BP)

- A Bunsetsu phrase (BP) is a chunk of words consisting of a content word (noun, verb, etc.) accompanied by some functional words (particle, auxiliary, etc.)

- A Japanese sentence can be analyzed as a sequence of BPs, which constitute an inter-BP dependency structure

太郎が学校に行く. (Taro goes to school.)

- 太郎/taro が/ga (noun, auxiliary)
- 学校/gakkou に/ni (noun, auxiliary)
- 行く/iku (verb)

: Bunsetsu phrase
**Bunsetsu phrase (BP)**

- A *Bunsetsu phrase (BP)* is a chunk of words consisting of a content word (noun, verb, etc.) accompanied by some functional words (particle, auxiliary, etc.).

- A Japanese sentence can be analyzed as a sequence of BPs, which constitute an inter-BP dependency structure.

---

太郎が学校に行く。
Taro goes to school.
Probabilistic partial parsing
Overview

- **Probabilistic extension** (Jensen et al., 1993)
- **Output only a part of the parse tree that are probabilistically highly reliable**

![Diagram of a parser with inputs and outputs, showing reliable and unreliable parse trees.]
Overview

- **Probabilistic extension** (Jensen et al., 1993)
- **Output only a part of the parse tree that are probabilistically highly reliable**

### Dependency probabilities (DPs)

![Diagram showing dependency probabilities]

- BP1: 0.90
- BP2: 0.65
- BP3: 1.00

Selecting only dependency relations whose estimated probability is higher than a certain threshold $\sigma$. 
Overview

- **Probabilistic extension** (Jensen et al., 1993)
- **Output only a part of the parse tree that are probabilistically highly reliable**

Selecting only dependency relations whose estimated probability is higher than a certain threshold $\sigma$. 

$$\sigma = 0.7$$
C-A curve

**Probabilistic partial parsing**

\[
\text{coverage} = \frac{\# \text{ of the decided relations}}{\# \text{ of all the relations in the test set}}
\]

\[
\text{accuracy} = \frac{\# \text{ of the correctly decided relations}}{\# \text{ of the decided relations}}
\]

You can achieve significantly higher accuracy only by sacrificing coverage very little.
Advantages

The user can make a fine-graind arbitrary choice on the trade-off between coverage and accuracy.

Such trade-off choice makes the existing parsers of wider application.

![Graph showing the trade-off between coverage and accuracy.]
Estimation of DPs

- **Bottom-up models** (Collins, 1996), (Uchimoto et al., 1999)
  - Directly estimate DPs

- **Top-down models**

You can estimate DPs, whether you have a top-down model or a bottom-up model.
Committee-based
Probabilistic partial parsing
Overview

Committee-based decision making is to combine the outputs from several different systems (e.g. parser) to make a better decision.

- **POS tagging** (Halteren et al., 1998; Brill et al., 1999)
- ** Parsing** (Henderson and Brill, 1999)
- **Word sense disambiguation** (Pedersen, 2000)
- **Machine translation** (Frederking and Nirenburg, 1994)
- **Speech recognition** (Fiscus, 1997)

These works empirically demonstrated that combining different systems often achieved significant improvements over the previous best system.
A basic scheme

Committee-based
Probabilistic partial parsing

Models (parsers)

Simple Majority function

CF : Combining Function
A basic scheme

\[ \text{Committee} \]

\[ \text{Combining Function (CF)} \]

\[ \text{Committee-based Probabilistic partial parsing} \]

To realize partial parsing on this scheme,

\[ \Rightarrow \]

the committee would need to accept probabilistically annotated votes
Most statistical parsers can be members of such a probabilistic voting committee.
Extension (2) : Standardization

- Reliability of dependency probabilities (DPs)
  - Equally reliable?

Reliability of DPs may differ depending on parsers

Standardization of DPs
To standardize input DPs, we add weighting functions.
Committee-based
Probabilistic partial parsing

Committee

Models (parsers)

Inputs

Extension (2) : Standardization

Models

WFM1

0.90

0.65

1.00

BP1

BP2

BP3

BP4

WFM1

WF : Weighting Function

BP1

BP2

BP3

BP4

Output

Committee

WFMk

1.00

0.70

1.00

BP1

BP2

BP3

BP4

WFm

0.86

0.48

1.00

BP1

BP2

BP3

BP4

M1

BP1

BP2

BP3

BP4

Mk

BP1

BP2

BP3

BP4

Mm

BP1

BP2

BP3

BP4

Committee

BP1

BP2

BP3

BP4

CF

BP1

BP2

BP3

BP4

Committee

BP1

BP2

BP3

BP4

BP1

BP2

BP3

BP4

BP1

BP2

BP3

BP4
Each member is allowed to cast (probabilistically parameterized) multiple votes for all the potential candidates.

Probabilistically annotated dependency structure

Committee-based Probabilistic partial parsing

DP: Dependency Probability
Extension (3) : Multiple voting

Committee-based
Probabilistic partial parsing

Committee

Models (parsers)

Inputs

Outputs
Extension (3) : Multiple voting

Committee-based Probabilistic partial parsing

Generalized Committee-based Probabilistic Partial Parsing
Weighting functions

- A bare DP may not be a precise estimation of the actual accuracy.

Committee-based Probabilistic partial parsing

- Underestimate accuracy
- Overestimate accuracy
You can standardize input DPs by referring to P-A curves acquired from some training data.
Weighting functions: Normal

Committee-based
Probabilistic partial parsing

Ma → BP1 → 0.80 → BP2 → BP3 → BP4 → WFMa → BP1 → 0.90 → BP2 → BP3 → BP4

Mb → BP1 → 0.80 → BP2 → BP3 → BP4 → WFMb → BP1 → 0.60 → BP2 → BP3 → BP4

Normal standardization
Weighting functions: Class

Training data

Adverbial dep. relations

Class-based standardization

Adnominal dep. relations

P-A curves may significantly differ depending on problem classes.

You could prepare weighting functions for each problem class.
Combining function

Committee-based Probabilistic partial parsing

Averaging of the given weight matrices
Summary

Our committee-based scheme:

(a) accepts probabilistic parameterized votes as its input
(b) accepts multiple voting
(c) considers the standardization of original input votes
(d) outputs a DP matrix as a final decision

Committee-based Probabilistic partial parsing

DP: Dependency Probability
Related works

Our voting scheme =

Generalization of existing voting techniques for NLP:
  - Probabilistic multiple voting
  - Standardization
  - DP matrix output (coverage/accuracy trade-off)

Previous voting techniques

- POS tagging (Halteren et al., 1998)
- Parsing (Henderson and Brill, 1999)
- . . .

Committee-based Probabilistic partial parsing

Not accept multiple voting

Not accept probabilistic voting
Experiments
committee members (parsers)

**KANA** (Ehara, 1998) : a bottom-up model based on maximum entropy estimation

**CHAGAKE** (Fujio et al., 1998) : an extension of the bottom-up model proposed Collins (Collins, 1996)

**Kanayama’s parser** (Kanayama et al., 1999) : a bottom-up model coupled with a HPSG

**Shirai’s parser** (Shirai et al., 1998) : a top-down model incorporating lexical collocation statistics

**Peach Pie Parser** (Uchimoto et al., 1999) : a bottom up model based on maximum entropy estimation
Training / test sets

- **Kyoto corpus (ver 2.0)** (Kurohashi et al., 1997)
  - collection of Japanese newspaper articles
  - annotated in terms of:
    - POS tags
    - BP boundaries
    - Inter-BP dependency relations

Experiments

- Rejected by at least one parser
- Assigned inconsistent BP boundaries by different parsers

- 19,956 sentences
- 13,990 sentences

- Five-fold cross-validation (for open test)
## Performance of each individual model

<table>
<thead>
<tr>
<th>Model (parser)</th>
<th>Total accuracy</th>
<th>11-point accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.8974</td>
<td>0.9607</td>
</tr>
<tr>
<td>B</td>
<td>0.8551</td>
<td>0.9281</td>
</tr>
<tr>
<td>C</td>
<td>0.8586</td>
<td>0.9291</td>
</tr>
<tr>
<td>D</td>
<td>0.8470</td>
<td>0.9266</td>
</tr>
<tr>
<td>E</td>
<td>0.7885</td>
<td>0.8567</td>
</tr>
</tbody>
</table>

*Total accuracy and 11-point accuracy are both given by C-A curve*
11-point accuracy is a summary of the C-A curve, which is given by the average of the accuracy of 11 points.
### Accuracy of each individual model

<table>
<thead>
<tr>
<th>Model (parser)</th>
<th>Total accuracy</th>
<th>11-point accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.8974</td>
<td>0.9607</td>
</tr>
<tr>
<td>B</td>
<td>0.8551</td>
<td>0.9281</td>
</tr>
<tr>
<td>C</td>
<td>0.8586</td>
<td>0.9291</td>
</tr>
<tr>
<td>D</td>
<td>0.8470</td>
<td>0.9266</td>
</tr>
<tr>
<td>E</td>
<td>0.7885</td>
<td>0.8567</td>
</tr>
</tbody>
</table>

- **Model A** is significantly better than other models
Issue (1) : Probabilistic voting

- Can we easily gather committee members?
  - Shirai’s parser (Shirai et al., 1998):
    - a top-down model (not provide DPs directly)
    - By using n-best dependency structure candidates, we were able to estimate DPs reasonably correctly

Experiments

![Graph showing dependency probability vs. accuracy]

Most statistical parsers can be committee members
Issue (2) : Standardization

Is standardization actually effective?  

Yes!

11-point accuracy

Standardization actually improved the performance

A included  
B included  
C included

committee
Issue (3) : Multiple voting

Does multiple voting improve the performance?

Yes!

At least when the size of a committee is small, multiple voting significantly outperformed single voting.

Multiple voting

Experiments

Committee size

Small

Large
Does combining parsers actually improve

- Including the optimal model A, the performance?
  - Not very visible improvement.

- Including the comparable members such as BC or BD
  - Extensive improvement

Experiments
Issue (4) : Contributions of a committee

- Does combining parsers actually improve
  - Including the optimal model A, the performance?
    - Not very visible improvement.

If we have another optimal parser that was comparable to parser A, then we might achieve significant improvements even in case where parser A participates.
Conclusion

- We proposed a general committee-based framework that can be coupled with probabilistic partial parsing

Findings through experiments

- (a) Both multiple voting and vote standardization effectively work in committee-based partial parsing
- (b) If more than two comparably competent optimal models are available, it is likely to be worthwhile to combine them
- (c) Our scheme also enables a non-parametric rule-based parser to make a good contribution