Jiawei Yong, Shintaro Kawamura, Katsumi Kanasaki, Shoichi Naitoh, and Kiyohiko Shinomiya
Ricoh Company, Ltd.
index

Segmentation subtask

- Overall thought for segmentation
- Cue-phrase-based idea
  - Semi-supervised segmentation
- Results and conclusion

Classification subtask

- Research challenges
- Research methods
- Results and conclusion
Segmentation subtask
Segmentation subtask in 2 steps

input: Date, Speaker, Summary

Date:
Speaker:
xxxxxxx
xxxxxxx
xxxxxxx
xxxxxxx
xxxxxxx
xxxxxxx
xxxxxxx
xxxxxxx

minutes

segments

contiguous segments that correspond to the input

search

 segmentation

初めに、xxx
xxxxxxx
xxx見解を求めます。

次に、xxx
xxxxxxx
xxx見解を求めます。

最後に、xxx
xxxxxxx
xxx質問を終わります。

次に、xxx
xxxxxxx
xxx見解を求めます。

最後に、xxx
xxxxxxx
xxx質問を終わります。
Data sets for the segmentation subtask

- **data sets provided by the task organizer**
  - training data: used as development data
  - test data

- **annotated by ourselves**
  - training data: 4804 utterances, 995 segments
  - development data: 3438 utterances, 683 segments

**Diagram:***
- **Segmentation**
  - minutes
  - segments

- **Search**
  - contiguous segments that correspond to the input
Cue-phrase-based idea (segmentation step)

- Hints for topical segmentation
  - Lexical cohesion
    TextTiling was tried in the dry run
    not reliable
  - Cue phrases
    used in the formal run
    effective for speech in the assembly
Models for segmentation step (formal run)

Submitted 5 Runs

- Rule-based Model (string pattern matching) … Run 1
- Supervised Model
  - BoW ⇒ SVM … Run 2
  - pre-trained word2vec ⇒ LSTM … Run 5
  - *word embeddings ⇒ HAN (unsubmitted)
- Semi-supervised Model (Original method)… Run 3
- No segmentation Model (each utterance is a segment) … Run 4
Semi-supervised model (Segmentation step)

- Segment boundaries are learned through bootstrapping.

84905 utterances

10 words at the head and the tail

boundary

iteration

BoW

compressed with LSI

logistic regression

speaker boundary

the first line of a segment

the last line of a segment

estimated segment boundary
**Search step**

- maximize $\sum_{i=1}^{k} \text{idf}(t_i) - \lambda k \log(n)$

- Coverage of weighted words $t_i(i=1,...,k)$ in the summary

- Penalty for the length $n$ utterances

- Hyperparameter $\lambda$ is tuned by the development data. (0.4 for questions and 0.7 for answers)
### Evaluation results

The performance of the methods when applied to the test data set (mean values of 5 runs)

<table>
<thead>
<tr>
<th>Segmentation method</th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>rule-based</td>
<td>0.851</td>
<td>0.913</td>
</tr>
<tr>
<td>SVM</td>
<td>0.819</td>
<td>0.851</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.916</td>
<td>0.690</td>
</tr>
<tr>
<td>HAN</td>
<td>0.871</td>
<td>0.874</td>
</tr>
<tr>
<td>semi-supervised</td>
<td>0.836</td>
<td>0.760</td>
</tr>
<tr>
<td>no segmentation</td>
<td>0.828</td>
<td>0.715</td>
</tr>
</tbody>
</table>

- The rule-based segmentation was the best during the formal run (**Top 1 in F1**). The method using a hierarchical attention network (unsubmitted one) also shows good performance.
Assembly speeches can be effectively segmented by cue phrases.

A rule-based segmentation and a neural network segmentation combined with a simple search model give good results. They can be baselines for more advanced methods that take syntactic or semantic features into account.

A semi-supervised segmentation that does not require training data is also feasible.
Classification subtask
The kappa statistics among annotators are quite low to the same sentence labelling.

Challenge 1: Low Kappa Statistic

The quantity of labelled utterances for each topic are insufficient.

Challenge 2: Underfitting

The volume of different labels in different topics are in a great disparity.

Challenge 3: Imbalanced Learning
Challenge 1: Low Kappa Statistic

Fact Checkability Subtask

1. Unanimous training data (4710)
   LSTM
   ① F1 score: 0.91

2. Majority training data (10291)
   LSTM
   ② F1 score: 0.81

News Detection Support for Fact Check (NLP2018)

Challenge 2: Underfitting

Stance Classification Subtask

Integrated model

The underfitting problem has been alleviated.
Research methods in classification

Challenge 3: Imbalanced Learning

Relevance & Stance Classification Subtask

Outlier detection

We regard Majority class as normal data, minority class as outlier value.

Relevance ("1") : irrelevance ("0") = 9390 : 901 ≈ 10 : 1

Isolation Forest

One class SVM

The F1 score of Minority Class
Evaluation results

The performance of the methods when applied to the test data set for classification

<table>
<thead>
<tr>
<th>Classification Subtasks</th>
<th>Top Values of RICT Runs for each criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
</tr>
<tr>
<td>1. Relevance</td>
<td>0.857 (rank 7)</td>
</tr>
<tr>
<td>2. Fact-checkability</td>
<td>0.729 (rank 3)</td>
</tr>
<tr>
<td>3. Stance</td>
<td>0.808 (rank 1)</td>
</tr>
<tr>
<td></td>
<td>2-Recall</td>
</tr>
</tbody>
</table>
Conclusions on classification subtask

- We have showed the assembly utterances can be classified by supervised learning methods with a high accuracy.

- The selection of training data acts an important role for supervised learning method. We shall select out the training data in consideration of quality, quantity, and balance.

---

1. Low Kappa Statistic Challenge
2. Underfitting Challenge
3. Imbalanced Learn Challenge

- Unanimous training data
- Integrated model
- Isolation Forest
Thank you for your attention.