I propose a method for automatic extraction of future-reference sentences (FRS). In the method I apply both morphological and semantic information to represent sentences in morphosemantic structure and extract frequent patterns from FRS. Then, I perform a series of experiments, in which I firstly train fourteen classifier versions and compare them to choose the best one. I conclude that the proposed method is capable to automatically classify future-reference sentences, significantly outperforming state-of-the-art, and reaching 76% of F-score.

Proposed Method

1. I propose a method for automatic extraction of future-reference sentences (FRS). In the method I apply both morphological and semantic information to represent sentences in morphosemantic structure and extract frequent patterns from FRS.
2. Then, I perform a series of experiments, in which I firstly train fourteen classifier versions and compare them to choose the best one. I conclude that the proposed method is capable to automatically classify future-reference sentences, significantly outperforming state-of-the-art, and reaching 76% of F-score.

Previous Research

In my previous work I distinguished a variety of FRS.

- Future-reference expressions: 270 sentences from newspapers.
- Time expressions: 70 / verbs: 141
- Probability of occurrence of future reference words in sentences one time: 45%, two times or more: 55%

Classification of FRS with morphosemantic patterns (MoPs)

- Morphosemantic: morphology + semantic
- Collect data FRS non-FRS
- Learning data FRS non-FRS
- Test data with MoPs
- Frequent MoPs
- Classify FRS

Outline

I am glad to hear your suggestions and comments.

Proposed Method

Morphosemantic Structure:

J: “Nihon unagi ga zetsumetsu kigushu ni shitei sare, kanzen yoshoku ni yoru unagi no ryosan ni kitai ga takamatte iru.”

E: “As Japanese eel has been specified as an endangered species, the expectations grow towards mass production of eel in full aquaculture.”

MoPs: predicate argument structure semantic role labelling: SRL

Argument Structure Analyzer etc.

Thesaurus of predicate argument structure for Japanese verbs words: 4400 semantic labels: 80

Use SPEC for training and classification:

- Sophisticated patterns (with disjoint elements)
  - awarding length (LA)
  - awarding length and occurrence (LOA)
  - awarding none (normalized weight, NW)
  - using all patterns (ALL)
  - erasing all ambiguous patterns (AMS)
  - erasing only those ambiguous patterns which appear in the same number in both sides (zero patterns, 0P)
  - patterns (PAT)
  - only n-grams (NGR)
- n-fold cross validation
- Results calculated in F-score, Precision, Recall
- Choose the most useful pattern

Results of experiments

Extracting frequent MoPs

- learning data: set50, set130
- 10-fold cross validation

Compare to F-scores set130 and set50

F-score with set50

F-score with set130

Extracted FRS

1. Score = 2.27

Conclusion

- We extracted FRS with morphosemantic.
- We verified validity of FRS patterns by two experiments.
1. using only 5-15 frequent patterns (F-score = 0.43)
2. validated classification for fully optimized model (F-score = 0.78, P = 0.65, R = 0.98)

Future work

- apply to real world events on large data.
- validate supporting to decide “yes” or “no” a future event with FRS.
  - ranking FRS
  - ordering timeline (short span, long span)