AUTOMATIC DETECTION OF SENTENCE AND CLAUSE UNITS USING LOCAL SYNTACTIC DEPENDENCY

Tatsuya Kawahara Masahiro Saikou Katsuya Takanashi

Kyoto University, Academic Center for Computing and Media Studies
Sakyo-ku, Kyoto 606-8501, Japan
kawahara@i.kyoto-u.ac.jp

ABSTRACT

For robust detection of sentence and clause units in spontaneous speech such as lectures and meetings, we propose a novel cascaded chunking strategy which incorporates syntactic and semantic information. Application of general syntactic parsing is difficult for spontaneous speech having ill-formed sentences and disfluencies, especially for erroneous transcripts generated by ASR systems. Therefore, we focus on the local syntactic dependency of adjacent words and phrases, and train binary classifiers based on SVM (Support Vector Machines) for this purpose. An experimental evaluation using spontaneous talks of the CSJ (Corpus of Spontaneous Japanese) demonstrates that the proposed dependency analysis can be robustly performed and is effective for clause/sentence unit detection in ASR outputs.

***Index Terms*—** spontaneous speech, chunking, sentence unit, clause unit, dependency analysis, SVM

1. INTRODUCTION

Automatic transcription of spontaneous human-to-human speech is expected to expand the applications of speech technology, for example, enabling efficient access to spoken documents such as broadcast programs, lectures and meetings. To organize a spoken document in a structured and readable form, the transcript should be segmented into appropriate units like sentences. Actually, most of the conventional natural language processing (NLP) systems, including parsers and machine translation systems, assume that the input is segmented by sentence units. Sentence segmentation is also an essential step to key sentence indexing and summary generation for effective presentation of spontaneous speech. However, utterances in spontaneous speech are ill-formed, and sentence boundaries are indistinct. Output text by automatic speech recognition (ASR) systems is just a sequence of words and has no explicit sentence boundaries, so the further step of segmenting the ASR output is required for these applications.

Therefore, a number of works have been conducted on this problem. The most successful approach so far is a machine learning-based classification. For a given point in a word sequence, a classifier inputs neighboring lexical features in combination with prosodic features, and determines if the current position is a sentence boundary or not. The classification is done word by word, assuming the lexical features as independent, so called "bag-of-words" model. On the contrary, when we read texts, syntactic and semantic information is apparently useful in detecting sentence/clause boundaries. Recently, these kinds of high-level linguistic information, provided by NLP parsers, are being exploited for rescoring an N-best list of ASR output, e.g. [1][9].

In this paper, we propose the incorporation of syntactic and semantic information into sentence/clause unit detection. Specifically, we focus on the local syntactic dependency and semantic case. These provide effective constraint for the possible boundaries. Moreover, they are defined locally without parsing the whole sentence, thus, robust against disfluencies and ASR errors. The proposed method is evaluated in comparison with the conventional methods in sentence/clause unit detection in the transcripts of real lectures and speeches included in the Corpus of Spontaneous Japanese (CSJ).

2. OVERVIEW OF SENTENCE/CLAUSE UNIT DETECTION

Automatic detection of sentence units (SU) was addressed in one of meta-data extraction (MDE) tasks in DARPA EARS program, and studied mainly on broadcast news (BN) and conversational telephone speech (CTS)[2]. In [3], Liu et al. reported the sentence boundary detection results in Rich Transcription evaluation (RT-04F). They combined prosodic features with linguistic features, and trained HMM, MaxEnt and CRF (Conditional Random Fields) methods, which gave similar performance, and then combined them by obtaining their majority vote for further improvement.

In Japanese, similar efforts have been made using the CSJ[4]. In spontaneous Japanese, in which subjects and verbs can be omitted, the sentence unit is not so evident. In the CSJ, therefore, the clause unit is primarily defined using morphological information. The sentence unit is then annotated by human judgment by considering syntactic and semantic in-

1-4244-0728-1/07/$20.00 ©2007 IEEE

formation. This annotation was given for the "core" 199 talks or 0.5M words[5].

Using this corpus, we have studied sentence boundary detection. A comprehensive evaluation was reported in [6], in which we combined lexical features with pause information, and trained a statistical language model and SVM (Support Vector Machines). The SVM proved to be more discriminative and robust against ASR errors.

In these works, a simple machine learning scheme treating all listed features as a single vector was adopted, though a separate classifier is designed to integrate possible prosodic features and compute a likelihood[3][7], which is then used as one of the features in the final classifier. Moreover, the linguistic features are usually defined in a "bag-of-words" model, which usually consists of baseforms or surface forms of respective words together with their POS (Part-Of-Speech) tags. Structural information is not extracted in this process.

On the other hand, when humans read transcripts, it is almost apparent that "chunking" process is performed to concatenate a sequence of words based on syntactic and semantic units. These units provide predictive markers on possible larger units of clauses and sentences, that is, sentence/clause boundaries appear only on the boundaries of these syntactic and semantic units. In this work, we explore the effective use of syntactic and semantic information in the sentence/clause unit detection. Previously, we investigated the use of syntactic parsing information in the sentence boundary detection[8], but obtained only a slight improvement in the manual transcripts, mainly because the parsing is not easy in spontaneous speech, which is often ill-formed and disfluent. In [9], Roark et al. also proposed the use of syntactic features generated by parsers in the sentence unit detection, and obtained a significant improvement for reference transcripts but only a marginal gain for ASR outputs. In order to realize robustness against ill-formed utterances as well as ASR errors, in this work, we focus on the local structure of syntactic dependency rather than conducting full parsing of sentences.

3. CHUNKING BASED ON LOCAL SYNTACTIC DEPENDENCY

3.1. Local Syntactic Dependency

Dependency structure is defined by dependency relations between words or compound nouns. The dependency structure is widely used for syntactic analysis of Japanese language, because (1) the relation is concisely defined based on a minimal grammatical unit named *bunsetsu*, which consists of one content word plus adjacent functional words, (2) the order of *bunsetsu*s is relatively free, (3) the dependency is usually from left to right *bunsetsu*s as the predicate is placed in the final position of the sentence. An example of the dependency analysis for a Japanese sentence is shown in Figure 1. In this example, each line corresponds to a *bunsetsu* unit, and its dependency relation is represented with lines. Note that all relations are left-to-right (downward in the figure).

English translation of the sentence:
"Currently, (they) do only three times in a summer, I suppose, because of financial reasons."

```
ima wa————————————————+  (currently)
yosaN no–+            |  (financial)
  kaNkei da to–+      |  (because of)
        omoi masu ga  |  (I suppose)
  hito natsu ni–+     |  (in a summer)
        saNkai gurai shika—-+  (only three times)
                 yari mase N  ((they) do)
```

Fig. 1. Example of dependency structure

ima wa |
yosaN no | *kaNkei da to* |
omoi masu ga |
hito natsu ni | *saNkai gurai shika* |
yari mase N |

Fig. 2. Example of chunking

Typical categories of the dependency relations are listed below.

- noun modifier (ex.) fine → day
- adverb (ex.) often → rains
- topic marker (ex.) rains ← today
- case noun (ex.) have ← rain

Among these, the first relation of noun modifier is local, in that dependency is always defined with adjacent units. The relations defined by adverbs and topic markers are also often local, and in those cases, they are closely connected. On the other hand, the last relation of case-noun is not local, because other phrases can be put between them. Thus, they should not be connected even if they are placed adjacently. In this work, we focus on the local syntactic dependency defined above. This relation is expected to be robust against spontaneous utterances, which often contains ill-formed sentence structures and disfluency phenomena. Here, filler words are disregarded in the dependency analysis.

3.2. Chunking Algorithm

In order to detect the local syntactic dependency, we introduce SVM-based chunkers. Chunking is performed in a cascaded manner as follows.

1. Chunk words into *bunsetsu* units

   A *bunsetsu* unit consists of a content word or a noun compound with adjacent functional words. We prepare a chunker based on SVM [10], which determines if the two adjacent words are to be combined into this unit. Features for the chunker include surface forms and POS tags. We used the 3rd-order polynomial kernel for SVM and the IOE labeling scheme with left-to-right analysis.

2. Determine if the adjacent *bunsetsu* units have dependency relation

   For this purpose, we prepare another SVM that makes binary decision (related or not). Features for the classifier include morphological information (baseform and POS tag) of the content word and the last functional word in each unit. The two adjacent units are chunked if their dependency relation is identified.

3. Determine if the *bunsetsu* is a predicate in order to detect the case-noun relation

   A simple rule based on the POS tag is enough to detect predicate verbs. If the unit is a predicate, it is not chunked with the adjacent unit regardless of the previous step.

An example of the resultant chunks is shown in Figure 2.

4. DETECTION OF CLAUSE/SENTENCE UNITS

4.1. Definition of Clause/Sentence Boundaries

In spontaneous Japanese, the definition of sentences is not distinct. In the CSJ, the clause unit is primarily defined. The clause boundaries are classified into following three types. These three types differ in their degree of completeness as a syntactic and semantic unit, and independence from their subsequent clauses.

- absolute boundary... complete sentences
- strong boundary... independent and parallel clauses (ex.) "... and ...", "... but ..."
- weak boundary... subordinate or conditional clauses (ex.) "if...", "because..."

Among these, absolute boundaries and strong boundaries are basically defined as sentence boundaries. Some of weak boundaries are also sentence boundaries, but this annotation is done by human judgement considering the meaning of sentences.

4.2. Clause Unit Detection

In the previous work[6], we have demonstrated that the SVM-based chunker realizes the best performance in sentence boundary detection, both for manual transcripts and ASR outputs, compared with statistical language models. Thus, we adopt this approach in this work.

The classification of four (=N) categories (three boundary types and non-boundary case) is realized by pair-wise binary classifiers based on SVMs, and the final decision is done by voting of these $N * (N - 1)/2$ classifiers. Features given to SVMs are surface forms and POS tags of the preceding and following three words of the current boundary candidate (=every boundary of chunks).

Table 1. Accuracy of *bunsetsu* unit chunking

	recall	precision	F-measure
text	97.9%	98.4%	98.2%
ASR	80.3%	78.4%	79.3%

5. EXPERIMENTAL EVALUATION

5.1. Experimental Setup

An experimental evaluation was done using the CSJ, which is a collection of academic presentations and extemporaneous speeches[4]. Clause/sentence boundaries are manually annotated for the core set of talks. As a test-set, we used 30 talks, which are also used as the ASR evaluation set[11]. The text size of the test-set is 71K words. Automatic transcription was made using the baseline speaker-independent ASR system[11], and the average word error rate is 30.2%. The remaining 168 core talks were used for training of all the SVM-based classifiers. The text size of the training data is 424K words.

5.2. Chunking Accuracy

First, performance of the two chunking processes was evaluated. Performance of the initial *bunsetsu* chunking is given in Table 1 for both manual transcripts (text) and ASR outputs (ASR), with respect to the recall of correct boundaries, precision of detected boundaries, and their mean (F-measure). A quite high accuracy is obtained for the transcripts of spontaneous speech which contains disfluencies. In a preliminary experiment, we also tested the case where fillers were removed before chunking, and found that the filler removal had adverse effect in chunking. This suggests that fillers are more likely to be inserted in the phrase boundaries and they are useful markers for chunking. The accuracy is degraded for the automatic transcripts by ASR, but the degradation (19%) is much smaller than the word error rate (30%).

Next, performance of the local syntactic dependency analysis is shown in Table 2. For the manual transcripts, an accuracy over 90% is achieved. In the previous work[8], we conducted the general dependency analysis for the same corpus with an accuracy of 80.6%. Most of the errors were caused by long-distance dependencies. In this work, by focusing on the dependency of adjacent units with a binary classifier, we obtain a much higher accuracy. The degradation for the ASR outputs (17%) is much smaller than the word error rate (30%), demonstrating the robustness of the chunking.

Table 2. Accuracy of local dependency analysis

	recall	precision	F-measure
text	91.6%	88.8%	90.2%
ASR	75.5%	74.3%	74.9%

Table 3. Accuracy of clause unit detection (F-measure)

clause type		baseline	proposed
absolute	text	96.2%	96.3%
	ASR	74.7%	74.6%
strong	text	97.4%	97.4%
	ASR	73.6%	76.8%
weak	text	97.2%	93.2%
	ASR	63.4%	62.2%
overall	text	95.4%	94.5%
	ASR	**69.9%**	**74.0%**

5.3. Clause Unit Detection Performance

Then, clause unit detection was conducted and evaluated. For reference, a method without the proposed chunking was implemented. The baseline method is also based on pair-wise SVMs using the same features from neighboring six words (three words in both sides). The difference is that the proposed method limits possible clause boundaries to those of the generated chunks. The results (F-measure) are listed in Table 3 for all types of clause boundaries. In the overall evaluation, confusions between three boundary types are disregarded to evaluate the clause unit detection itself. The total number of clauses in the test-set is 7046.

For the manual transcripts, both methods obtain much the same accuracy. A degradation is observed for the weak boundaries by the proposed method, which cannot detect boundaries of clauses modifying adjacent nouns. On the other hand, for the ASR outputs, the proposed method realizes a significantly higher accuracy (by 4% absolute). The chunking provides robustness in the clause unit detection although there are some confusions among three boundary types.

6. CONCLUSIONS

We have proposed a novel chunking strategy for robust detection of clause/sentence units in spontaneous speech. The proposed method focuses on the local syntactic dependency and introduces binary classifiers based on SVM for improved robustness. Whereas it is known in the NLP research community that chunking into compound nouns and phrases is useful in clause/sentence detection, the proposed method is to generate larger chunks based on their dependencies, and was evaluated in spontaneous speech having ill-formed sentences and disfluencies. The experimental evaluation shows that the method is much more effective for ASR outputs rather than manual transcripts. The method will also be useful in ASR-based transcript correction and summary generation since it provides semantically meaningful units.

7. REFERENCES

[1] M.Balakrishna, D.Moldovan, and E.Cave. N-best list reranking using higher level phonetic, lexical, syntactic and semantic knowledge sources. In *Proc. IEEE-ICASSP*, volume 1, pages 413–416, 2006.

[2] Y.Liu, E.Shriberg, A.Stolcke, B.Peskin, J.Ang, D.Hillard, M.Ostendorf, M.Tomalin, P.Woodland, and M.Harper. Structural metadata research in the EARS program. In *Proc. IEEE-ICASSP*, volume 5, pages 957–960, 2005.

[3] Y.Liu, E.Shriberg, A.Stolcke, D.Hillard, M.Ostendorf, and M.Harper. Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Trans. Audio, Speech & Language Process.*, 14(5):1526–1540, 2006.

[4] S.Furui. Recent advances in spontaneous speech recognition and understanding. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pages 1–6, 2003.

[5] K.Maekawa. Corpus of Spontaneous Japanese: Its design and evaluation. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pages 7–12, 2003.

[6] Y.Akita, M.Saikou, H.Nanjo, and T.Kawahara. Sentence boundary detection of spontaneous Japanese using statistical language model and support vector machines. In *Proc. INTERSPEECH*, pages 1033–1036, 2006.

[7] J.H.Kim and P.C.Woodland. The use of prosody in a combined system for punctuation generation and speech recognition. In *Proc. EUROSPEECH*, pages 2757–2760, 2001.

[8] K.Shitaoka, K.Uchimoto, T.Kawahara, and H.Isahara. Dependency structure analysis and sentence boundary detection in spontaneous Japanese. In *Proc. COLING*, pages 1107–1113, 2004.

[9] B.Roark, Y.Liu, M.Harper, R.Stewart, M.Lease, M.Snover, I.Shafran, B.Dorrand, J.Hale, A.Krasnyanskaya, and L.Yung. Reranking for sentence boundary detection in conversational speech. In *Proc. IEEE-ICASSP*, volume 1, pages 545–548, 2006.

[10] T.Kudo and Y.Matsumoto. Chunking with support vector machines. In *Proc. NAACL*, 2001.

[11] T.Kawahara, H.Nanjo, T.Shinozaki, and S.Furui. Benchmark test for speech recognition using the Corpus of Spontaneous Japanese. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pages 135–138, 2003.