T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

Title	Topic Extraction Based on Continuous Speech Recognition in Broadcast-News Speech
Author	Katsutoshi Ohtsuki, S. Matsunaga, Tatsuo Matsuoka, Sadaoki Furui
Journal/Book name	IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU1997), Vol. , No. , pp. 527-534
Issue date	1997, 12
DOI	http://dx.doi.org/10.1109/ASRU.1997.659132
URL	http://www.ieee.org/index.html
Copyright	(c)1997 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.
Note	このファイルは著者（最終）版です。 This file is author (final) version.

Powered by T2R2 (Science Tokyo Research Repository)

Proc. 1997 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 527-534 (1997)

Topic Extraction Based on Continuous Speech Recognition in Broadcast-News Speech

K. Ohtsuki
S. Matsunaga
NTT Human
Interface Labs.
Kanagawa, Japan

T. Matsuoka
NTT Multimedia
Business Dept.
Tokyo, Japan

S. Furui
Tokyo Institute of
Technology
Tokyo, Japan

Abstract - This paper reports on topic extraction in Japanese broadcast-news speech. We studied, using continuous speech recognition, the extraction of several topic-words from broadcast-news. A combination of multiple topic-words represents the content of the news. This is more detailed and more flexible than a single word or a single category. A topic-extraction model shows the degree of relevance between each topic-word and each word in the articles. For all words in an article, topic-words which have high total relevance score are extracted from the article. We trained the topic-extraction model with five years of newspapers, using the frequency of topic-words taken from headlines and words in articles. The degree of relevance between topic-words and words in articles is calculated on the basis of statistical measures, i.e., mutual information or the χ^2-value. In topic-extraction experiments for recognized broadcast-news speech, we extracted five topic-words using a χ^2-based model and found that 75% of them agreed with topic-words chosen by subjects.

1. Introduction

Recent advances in digital technology make it possible to store a large amount of speech data. In order to use such a large amount of data effectively as 'information', classifying and indexing are essential. Particularly, in order to classify or retrieve speech data without playing it back, automatic topic identification (TID) of speech data is needed. In the TID of conversational speech or news speech, keyword spotting methods are commonly used with selected discriminative keywords [1][2][3][4]. But with those methods the number of keywords is limited, and many wrong keywords are spotted if the number of keywords increases. Although some approaches to TID employ continuous speech recognition (CSR) techniques [5][6][7], they classify speech data into at most ten categories or topics. There are not so many application areas for data classified into such a limited number of categories. Even with a large number of categories, the classification is not usable unless it matches the users' ideas.

What we attempt to extract from news speech is a set of topic-words. A combination of several topic-words represents the content of the news. This is more detailed than a single word or a single category [8]. A topic-extraction model shows the relation (degree of relevance) between each topic-word and each word in

the articles. For all words in an article, relevance scores to each topic-word are summed up. Then topic-words with high total relevance scores are extracted. We trained the topic-extraction model with five years of newspapers, using the frequency of topic-words taken from headlines and words in articles. We dealt with not all the words in the news article but content-words such as nouns and verbs, which contain more semantic information than other words from the point of view of information retrieval. Topic-words and words in articles are separate sets, therefore we can extract topic-words that have high relevance to the article whether they are in the article or not.

2. Topic Extraction

2.1 Topic-Extraction Model

Statistical measures that indicate the probability or the frequency of appearance of words in article, for a given topic-word can be calculated from a large amount of data, i.e., many articles and their topic-words. We calculated a relevance score based on two such measures in this study. One is mutual information, defined as the degree of dependence between random variables. It shows high value when the mutual dependence of a topic-word and a word in article is strong. The other measure is the χ^2-value in the χ^2-test, which represents the degree of a gap between a topic-word and an article-word. A high value shows that the word appears biasedly according to topic-words, or it appears to a particular topic-word.

The mutual information based method The mutual information based relevance score between word w_i and topic-word t_j is expressed as

$$I(w_i;t_j) = \log \frac{P(w_i,t_j)}{P(w_i)P(t_j)}. \tag{1}$$

If there is no concurrence of word w_i and topic-word t_j in training data, $P(w_i,t_j)=0$ and this arises a problem in summing up scores. Then we consider that no information is obtained when no concurrence is observed. The mutual information is set to 0 under this condition as

$$I'(w_i;t_j) = \begin{cases} I(w_i;t_j), & \text{if}\left(P(w_i,t_j)\neq 0\right) \\ 0, & \text{if}\left(P(w_i,t_j)=0\right) \end{cases}. \tag{2}$$

Mutual information is a measure based on conditional probability, and frequencies of word w_i and topic-word t_j are not considered. We also consider a relevance score based on the mutual information weighted with a joint probability as

$$I''(w_i;t_j) = P(w_i,t_j)\cdot \log \frac{P(w_i,t_j)}{P(w_i)P(t_j)}. \tag{3}$$

The χ^2-value based method The χ^2-value based relevance score between word w_i and topic-word t_j is expressed as

$$\chi_{ij}^2 = \frac{\left(f_{ij} - F_{ij}\right)^2}{F_{ij}}, \tag{4}$$

where f_{ij} is the frequency of word w_i appearing in a news item with topic-word t_j, and F_{ij} is the theoretical frequency of word w_i, i.e., when it appears with equal probabilities for all topic-words. It is given by

$$F_{ij} = \frac{\sum_{l=1}^{M} f_{il}}{\sum_{k=1}^{N}\sum_{l=1}^{M} f_{kl}} \cdot \sum_{k=1}^{N} f_{kj}, \tag{5}$$

where N is a distinct number of words and M is a distinct number of topic-words. If the gap between the actual frequency and the theoretical frequency, i.e., if $f_{ij} - F_{ij}$ is positive and big, the word should tend to appear with the topic-word. However, eq. (4) yields the same value when $f_{ij} - F_{ij}$ is either positive or negative. Then we calculate the χ^2-value based relevance score as

$$\chi_{ij}^{2'} = \frac{\left(f_{ij} - F_{ij}\right) \cdot \left|f_{ij} - F_{ij}\right|}{F_{ij}} \tag{6}$$

considering the sign of $f_{ij} - F_{ij}$.

2.2 Training Data

The topic-extraction model contains all relevance scores between each word in the content and each topic-word. We trained the topic-extraction model with newspaper articles and headlines extending back about five years from January 1990 to September 1994. It contains about 900k news articles. Japanese sentences have no spaces between words, so we segmented the articles and the headlines into words with a morphological analyzer. Words in headlines were trained as topic-words. To reduce the enormous number of combinations of words and topic-words, we didn't use infrequently appearing words and function words. In the end, the distinct number of topic-words was about 70k.

2.3 Topic Extraction Method

Topic-words are extracted from news articles on the basis of relevance scores between topic-words and articles. The relevance score R_{aj} between topic-word t_j and article a is calculated as

$$R_{aj} = \sum_{k=1}^{N_a} s_k r_{kj}, \tag{7}$$

where N_a is the number of words in the article a and r is the relevance score between a topic-word and a word in an article. s is a weighting factor for a word

and we calculated with $s = 1$ in these experiments. For each article, topic-words with high relevance scores are extracted in order of their score.

3. Evaluation Data

3.1 Speech Data

The broadcast-news speech was segmented into sentences for evaluation. The evaluation data set were 29 articles which had 142 utterances. An article had from 2 to 14 utterances and an average of 5 utterances. The data sets consisted of utterances of 15 male speakers: 8 anchor speakers and 7 other speakers. The speech contained spontaneous speech phenomena, such as 'uh' at the beginning of a sentence or the correction of slips and also included background noise or music.

3.1 Topic Data

We transcribed the evaluation speech data and gave topic-words to each transcribed article. We were aiming at extracting topics that match the users' ideas well. We asked three subjects to give topic-words to use as keys for retrieving the article and evaluated extracted topics with them. Each of the subjects was instructed to give more than 4 and an average of 10 topic-words to each article. We made two evaluation topic sets, an AND set, which contains any topic-words that all the subjects gave in common, and an OR set, which contains all topic-words that at least one subject gave. We segmented given topic-words into morphemes with a morphological analyzer for the evaluation. The average number of the morphemes per article was 10.4 for the AND set and 35.7 for the OR set.

4. Experiments

4.1 Large Vocabulary Continuous Speech Recognition

Our large vocabulary continuous speech recognition (LVCSR) system has context-dependent phoneme HMMs and statistical n-gram language models [9][10]. The phoneme HMMs are triphone models designed using tree-based clustering [11]. They were trained with phonetically-balanced sentences and dialogue read by 53 male speakers. The total number of utterances was 13,270 and the total volume of training data was approximately 20 hours. The n-gram language models were trained using broadcast-news manuscripts extending over five years, which had about 500k sentences, or 24M words. We estimated unseen n-gram probabilities using Katz's back-off smoothing method [12]. The vocabulary size of the system was 20k. Table 1 shows the test-set perplexity and LVCSR results for the above-mentioned testing data. The improvement of recognition performance with the trigram language model is rather small because of insufficiency of training data.

Language Model	Test-set Perplexity	Word Error Rate
bigram	98	28.2%
trigram	56	24.6%

Table 1: LVCSR results for news speech data

4.2 Topic Extraction Experiments

Figure 1 shows the results of fifty topic-words extracted from transcribed news speech, which were evaluated with the OR set. Recall and precision are defined as

$$Recall = \frac{C}{T} \cdot 100 \tag{8}$$

$$Precision = \frac{C}{H} \cdot 100, \tag{9}$$

where C is the number of correct topic-words retrieved, T is the total number of topic-words in the evaluation topic sets, and H is the total number of topic-words retrieved. Recall and precision tend to have a trade-off relation, and achieving high recall and high precision, i.e., the upper-right region in the charts, is required for an information retrieval system. In Fig. 1, the χ^2-value based method [CHI: eq. (6)] achieved better performance than the mutual information method [MI: eq. (2)]. Although the mutual information method that was weighted with joint probability [wMI: eq. (3)] achieved better results than the MI method, it was still worse than CHI. The topic-words extracted by the wMI method are rather general and such general topic-words don't match given topic-words well.

Then we applied the topic extraction models based on χ^2-value to speech recognition results. Figure 2 shows topic extraction results for recognized news speech (tg), evaluated with both the AND and the OR evaluation sets. Results for transcription (txt) are also shown for comparison. The performance for speech-recognition results is inferior to that for transcription because of recognition error.

For information retrieval with keywords, using five keywords is reasonable. We extracted five topic-words, or five plot symbols from the upper-left of a line graph in the charts, and evaluated them with the OR set. 86% precision for transcription and 75% for recognition results were obtained. The given topic-words vary according to the subjects. The overlapping rate of the topic-words given by two subjects is 74 % on average. That is to say, picking up five topic-words from a subject's topic, they will have about seventy-percent precision for another subject's topic. Therefore 75% precision is applicable for practical use.

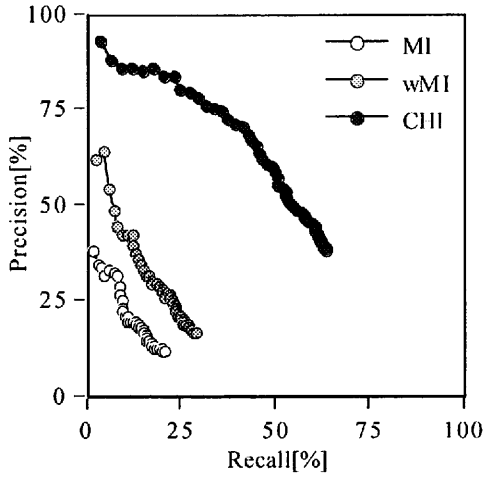

Figure 1: Topic extraction results for transcribed news speech (evaluated with OR set)

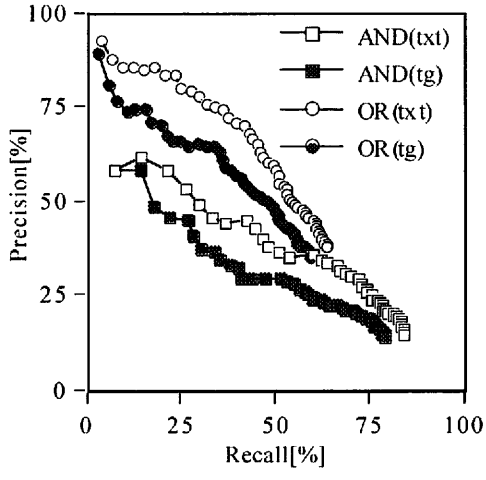

Figure 2: Topic extraction results for speech-recognized news speech (extracted with CHI method)

5. Summary

We reported topic extraction from broadcast news speech, based on continuous speech recognition. We proposed topic-extraction models that have mutual information or the χ^2-value as the degree of relevance between topic-words and words in news articles. For each article, topic-words that have high relevance score

for the words in the article are extracted. In our experiments the χ^2-value based topic-extraction model achieved better performance than the mutual information based model. Extracting five topic-words using a topic extraction model based on the χ^2-value, 75% precision was obtained for speech-recognized news speech. The precision is practically applicable, since it is almost the same as the overlap rate of topic-words given by each subject. But the results are inferior to results for transcribed news speech; therefore, robustness to recognition error is needed for the topic extraction models.

Acknowledgments

The authors would like to thank Mr. Kazuo Tanaka of NTT Human Interface Laboratories for providing us with the morphological analyzer. The authors are also grateful to Nihon Keizai Shimbun Incorporated for allowing us to use the newspaper text database (Nikkei CD-ROM 90-94) for our research. The authors greatly thank NHK (Japan Broadcasting Corporation) for providing us with the broadcast-news database.

References

[1] Rose, R.C., Chang, E.I., and Lippmann, R.P. "Techniques for Information Retrieval from Voice Messages," Proc. ICASSP, pp. 317-320, 1991.

[2] McDonough, J., and Gish, H. "Issues in Topic Identification on the Switchboard Corpus," Proc. ICSLP, pp. 2163-2166, 1994.

[3] Foote, J.T., Jones, G.J.F., Sparck Jones, K., and Young, S.J. "Talker Independent Keyword Spotting for Information Retrieval," Proc. EUROSPEECH, pp. 2145-2148, 1995.

[4] James, D.A. "A System for Unrestricted Topic Retrieval form Radio News Broadcasts," Proc. ICASSP, pp. 279-282, 1996.

[5] Rohlicek, J.R., Ayuso, D., Bates, M., Bobrow, R., Boulanger, A., Gish, H., Jeanrenaud, P., Meteer, M., and Siu, M. "Gisting Conversational Speech," Proc. ICASSP, Vol. II, pp. 113-116, 1992.

[6] Jeanrenaud, P., Siu, M., Rohlicek, J.R., Meteer, M., and Gish, H. "Spotting Events in Continuous Speech," Proc. ICASSP, Vol. I, pp. 381-384, 1994.

[7] Peskin, B., Connolly, S., Gillick, L., Lowe, S., McAllaster, D., Nagesha, V., Mulbregt, P., Wegmann, S. "Improvements in Switchboard Recognition and Topic Identification," Proc. ICASSP, pp. 303-306, 1996.

[8] Imai, T., Schwartz, R., Kubala, F., and Nguyen, L. "Improved Topic Discrimination of Broadcast News using a Model of Multiple Simultaneous Topics," Proc. ICASSP, Vol. 1, pp. 727-730, 1997.

[9] Matsuoka, T., Ohtsuki, K., Mori, T., Yoshida, K., Furui, S., and Shirai, K. "Japanese Large-Vocabulary Continuous-Speech Recognition using a Business-Newspaper Corpus," Proc. ICASSP, Vol. 3, pp. 1803-1806, 1997.

[10] Matsuoka, T., Taguchi, Y., Ohtsuki, K., Furui, S., and Shirai, K. "Toward Automatic Transcription of Japanese Broadcast News," Proc. EUROSPEECH, Vol. 2, pp. 915-918, 1997.

[11] Young, S.J., Odell, J.J. and Woodland, P.C. "Tree-based State Tying for High Accuracy Acoustic Modeling," Proc. DARPA Human Language Technology Workshop, pp. 307-312, 1994.

[12] Katz, S.M. "Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer," IEEE Trans. ASSP-35, pp. 400-401, 1987.