

論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	Linguistic Model Adaptation for Automatic Speech Summarisation
著者(和文)	ウィッタッカー エドワード, 古井 貞熙
Authors(English)	Pierre Chatain, Edward Whittaker, Joanna Mrozinski, Sadaoki Furui
出典(和文)	日本音響学会2006年春季講演論文集, Vol. , No. , pp. 35-36
Citation(English)	, Vol. , No. , pp. 35-36
発行日 / Pub. date	2006, 3

Linguistic Model Adaptation for Automatic Speech Summarisation *

© Pierre Chatain, Edward W.D. Whittaker, Joanna A. Mrozinski and Sadaaki Furui (Tokyo Institute of Technology)

1 Introduction

In this paper we extend the work done on the two-stage summarisation method described in [1] by focusing on adapting the linguistic component to make it more suited for the summarisation task. In particular we examine methods for adapting the linguistic models (LiM) automatically to improve performance, using either unigram, bi-gram or trigram information from different sources of data. Experiments were performed both on spontaneous speech, using 9 talks taken from the Translanguage English Database (TED) corpus [2], and speech read from text, using 5 talks from CNN broadcast news from 1998. For each of those talks, human (TRS) and speech recogniser (ASR) transcriptions along with human summaries were used. The talks were used for both development and evaluation with a rotating form of cross-validation [3]. The objective measure of summary quality used in this paper is summarisation accuracy (SumACCY) [4]. The full process is described in [5].

2 Experiences and Results

2.1 Reference Results

Reference results, consisting of the results for random summarisation (Rnd sum) and the baseline, evaluated using a baseline LiM (LiM_B), are given for appropriate summarisation ratios (Sum Ratio) in Tables 1 and 2 for the TED and CNN data respectively. The LiM_B is a trigram model built on a corpus consisting of around ten years of conference proceedings (17.8M words) on the subject of speech and signal processing.

2.2 TED

Nine talks from the TED corpus were used in this paper, each transcribed and manually summarised

Table 1 Reference results for TED.

Sum Ratio			SumACCY
10%	TRS	Rnd sum	34.4
		Baseline	63.1
	ASR	Rnd sum	33.9
		Baseline	48.6
30%	TRS	Rnd sum	71.2
		Baseline	81.6
	ASR	Rnd sum	56.1
		Baseline	66.7

* 音声自動要約のための言語モデル適応

シャタン ビエール, エドワード ウィッタッカー, ヨアンナ ムロジンスキ, 古井貞熙 (東工大)

Table 2 Reference results for CNN.

Sum Ratio			SumACCY
40%	TRS	Rnd sum	80.7
		Baseline	81.1
	ASR	Rnd sum	68.2
		Baseline	71.3

Table 3 TED adaptation results on TRS.

Sum Ratio	LiM	N-gram		
		1	2	3
10%	LiM_{TN}	67.6	61.6	61.4
	LiM_{SN}	62.0	62.0	65.9
30%	LiM_{TN}	82.6	83.7	82.5
	LiM_{SN}	83.2	82.1	83.7

by nine different humans for both 10% and 30% summarization ratios.

Two different types of component LiM are built, coming from two different sources of data, and are combined with the LiM_B for adaptation using linear interpolation.

The first type of component linguistic models (LiM_{S1} , LiM_{S2} and LiM_{S3} for unigrams, bigrams and trigrams, respectively, where S denotes Summary) are built on the small corpus of hand-made summaries described above, made for the same summarisation ratio as the one we are generating. For each talk the hand-made summaries of the other eight talks (i.e. 72 summaries) were used as the training corpus. This type of LiM is expected to help generate automatic summaries in the same style as those made manually.

The second type of component linguistic model (LiM_{T1} , LiM_{T2} and LiM_{T3} for unigrams, bigrams and trigrams, respectively, where T denotes Talk) are built from the paper in the conference proceedings for the talk we want to summarise. This type of LiM, used for topic adaptation, is investigated because key words and important sentences that appear in the associated paper should have a high chance of being extracted by the summarisation process.

Experiments were first made on the TRS, investigating interpolation of the baseline language model with each component. We also investigated a three way interpolation, combining LiM_B , LiM_{T1} and LiM_{S3} for both TRS and ASR.

Results for the adapted LiMs are given in Table

Table 4 Three way interpolation for TED.

Sum Ratio		SumACCY
10%	TRS	67.8
	ASR	49.8
30%	TRS	83.3
	ASR	68.7

3 for the combinations of the different component LiMs with the baseline one by one. Results using LiM_{T_1} seem to indicate that in the case of talk-based adaptation, the unigram adaptation model has the greatest effect. This is because the unigram information coming from such a small amount of data is more robust than bigram or trigram information. This unigram information helps selecting important key and technical words present in the lecture which are uttered during the talk.

However, with summary-based adaptation, trigram models yield the best results, which confirms our idea of realising stylistic adaptation: human-made summaries used in the model actually help us select utterances likely to be used by human subjects when making summaries. The unigram information in this case is not as important as the trigram one, which helps selecting frequently used expressions and phrases likely to appear in multiple summaries.

The results of the three way interpolation combining LiM_B , LiM_{S_3} and LiM_{T_1} are shown in Table 4, for both TRS and ASR. For TRS, relative improvements of 7.4% and 2.1% were obtained for the 10% and 30% summarisation ratios respectively. With the ASR, relative improvements of 2.5% and 3% are observed for the 10% and 30% summarisation ratios.

2.3 CNN

The CNN data consists of five talks from broadcast news of 1998. The five talks were transcribed and manually summarised by sixteen different humans for a 40% summarization ratio, and used for both development and evaluation.

For the CNN news stories we only built one type of component linguistic model, using hand made summaries in the same manner as for the TED data: for each talk the hand-made summaries of the other four talks (i.e. 64 summaries) were used as the training corpus. LiM_{C_1} , LiM_{C_2} and LiM_{C_3} (for unigram, bigram, and trigram, respectively, C standing for CNN) were interpolated separately with LiM_B as above.

Results for LiM adaptation on the CNN data are given in Table 5 for both TRS and ASR.

LiM adaptation did not yield any improvement on the CNN TRS. The baseline and random summarisation results were already very high, and adaptation had almost no effect.

With the ASR though, 3.1% relative improvement was obtained using LiM adaptation, the greatest in-

Table 5 CNN adaptation results.

		N-gram		
		1	2	3
LiM_{CN}	TRS	80.9	80.8	80.8
LiM_{CN}	ASR	72.5	73.4	73.5

crease coming from the trigram adaptation model, as was also observed on the TED data.

3 Conclusions

In this paper we have investigated combinations of models used to compute the linguistic score in an automatic speech summarisation system. It was found that summarisation performance was improved by at least 2.5% relative increase over the baseline on the ASR output of both spontaneous speech data coming from the TED corpus and speech read from text from CNN broadcast news.

Topic adaptation was performed using unigram information from a small source of data related to the talk we are trying to summarise, and stylistic adaptation was realised using the trigram information coming from a small corpora of hand-made summaries.

Acknowledgements The authors would like to thank Matthias Woelfel for the recognizer transcriptions and Chiori Hori for her previous work on two stage summarization and for gathering data from the TED corpus. This work is supported in part by the 21st Century COE Program.

References

- [1] T. Kikuchi, S. Furui, and C. Hori, "Automatic speech summarization based on sentence extraction and compaction," *Proc. ICASSP, Hong Kong, China*, vol. 1, pp. 236–239, 2003.
- [2] L. Lamel, F. Schiel, A. Fourcin, J. Mariani, and H. Tillmann, "The translanguage english database (ted)," *Proc. ICSLP, Yokohama, Japan*, vol. 4, pp. 1795–1798, 1994.
- [3] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.
- [4] C. Hori, T. Hori, and S. Furui, "Evaluation method for automatic speech summarization," *Proc. Eurospeech, Geneva, Switzerland*, vol. 4, pp. 2825–2828, 2003.
- [5] P. Chatain, E.W.D. Whittaker, J.A. Mrozinski, and S. Furui, "Topic and stylistic adaptation for speech summarisation," *Accepted at ICASSP, Toulouse, France*, 2006.