Self-validated Story Segmentation of Chinese Broadcast News

Wei Feng[1], Lei Xie[2(✉)], Jin Zhang[2], Yujun Zhang[1], and Yanning Zhang[2]

[1] School of Computer Science and Technology, Tianjin University, Tianjin 300350, China
{wfeng,yujunzhang}@tju.edu.cn
[2] School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
{lxie,jzhang,ynzhang}@nwpu.edu.cn

Abstract. Automatic story segmentation is an important prerequisite for semantic-level applications. The normalized cuts (NCuts) method has recently shown great promise for segmenting English spoken lectures. However, the availability assumption of the exact story number per file significantly limits its capability to handle a large number of transcripts. Besides, how to apply such method to Chinese language in the presence of speech recognition errors is unclear yet. Addressesing these two problems, we propose a self-validated NCuts (SNCuts) algorithm for segmenting Chinese broadcast news via inaccurate lexical cues, generated by the Chinese large vocabulary continuous speech recognizer (LVCSR). Due to the specialty of Chinese language, we present a subword-level graph embedding for the erroneous LVCSR transcripts. We regularize the NCuts criterion by a general exponential prior of story numbers, respecting the principle of Occam's razor. Given the maximum story number as a general parameter, we can automatically obtain reasonable segmentations for a large number of news transcripts, with the story numbers automatically determined for each file, and with comparable complexity to alternative non-self-validated methods. Extensive experiments on benchmark corpus show that: (i) the proposed SNCuts algorithm can efficiently produce comparable or even better segmentation quality, as compared to other state-of-the-art methods with true story number as an input parameter; and (ii) the subword-level embedding always helps to recovering lexical cohesion in Chinese erroneous transcripts, thus improving both segmentation accuracy and robustness to LVCSR errors.

Keywords: Story segmentation · Self-validation · Topic detection
Chinese broadcast news · Subwords · Normalized cuts

L. Xie—This work is supported by NSFC 61671325, 61572354.

J. Ren et al. (Eds.): BICS 2018, LNAI 10989, pp. 568–578, 2018.
https://doi.org/10.1007/978-3-030-00563-4_55

1 Introduction

As the explosive growth of multimedia content, there is an urgent demand for automatic organization of the massive multimedia data to facilitate efficient topic-based retrieval and analysis [7,10,15]. Hence, a well-segmented multimedia document is clearly an important *prerequisite* for various tasks of high-level semantic browsing [10]. Story segmentation aims to partition a text, audio and/or video stream into a sequence of topically coherent segments, namely stories.

Previous efforts on story segmentation have focused on topic modeling and the selection of topical boundary cues. Such as lexical chaining [17], C99 [3], latent semantic analysis (LSA) [4], etc., detect word-level semantic variations in a document via various cohesive measures, and produce local similarity minima as story boundaries. Recently, graph-theoretic approaches have shown promising potentials in segmenting natural data, such as images [6,16] and real-world discourses [12]. It has been shown that the graph embedding of linguistic units and the normalized cuts (NCuts) criterion [16] lead to effective story segmentations of English spoken lectures [12]. Our preliminary results [13,23] also showed that the NCuts approach can obtain superior performance than previous lexical-based methods [4,22] in handling subtle and ambiguous topical boundaries of Chinese broadcast news. Indeed, we prefer an automatic story segmentation approach that meets the following four requirements. *Self-validation*: it should be able to automatically determine the number of stories in a document. *Efficiency*: it should be fast enough to be able to segment a large number of documents. *Accuracy*: the segmentation result should be reasonable and as accurate as possible. *Robustness*: since the segmentation may be based on erroneous transcripts generated by LVCSR [12], it should be robust to various recognition errors.

In this paper, we study how to segment inaccurate news transcripts, transcribed from audio via LVCSR [9]. Firstly, the inevitable Chinese LVCSR errors, resulted from adverse acoustic conditions, multiple speakers and out-of-vocabulary (OOV) words, pose significant difficulties in word-level lexical story segmentation [11,23]. Secondly, the specialty of Chinese language makes previous successful methods for English story segmentation [12], not directly applicable. We propose a simple yet effective approach, namely self-validate normalized cuts (SNCuts) using subword-level graph embedding. We demonstrate the effectiveness of the proposed approach to both error-free manual transcripts and erroneous LVCSR transcripts at different error rates using two benchmark corpora.

2 Self-validated Story Segmentation

In this section, we show how to realize self-validated story segmentation for erroneous LVCSR transcripts of Chinese broadcast news. The core of our approach is: (i) a subword-level graph embedding, and (ii) a new self-validated SNCuts graph partitioning criterion.

2.1 Subword-Level Graph Embedding

The LVCSR transcript $\mathcal{T}$ of a Chinese broadcast news stream is constituted by a sequence of recognized words $\{w_1 w_2 \dots w_M\}$. Due to the inevitable LVCSR errors, we use subwords (i.e., characters/syllables subsequences), rather than the raw recognized words, to build the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

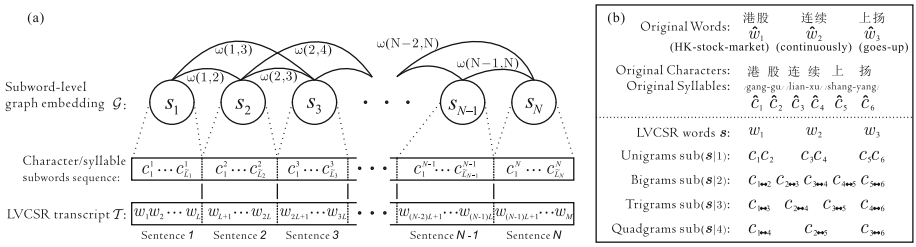

Fig. 1. Subword-level graph embedding: (a) is the subword-level graph embedding of an LVCSR transcript with cutoff distance $\tau = 2$; (b) shows an example of a Chinese sentence and the corresponding LVCSR recognized word sequence $\mathbf{s} = \{w_1 w_2 w_3\}$ and the subword-level n-gram representations sub($\mathbf{s}|n$) with $1 \le n \le 4$.

Node Extraction. Instead of relying on automatic Chinese sentence segmentation, we extract sentences from an LVCSR transcript as fixed number of consecutive word sequences. As shown in Fig. 1(a), we split the input LVCSR transcript $\mathcal{T} = \{w_1 w_2 \dots w_M\}$ into $N = \lceil \frac{M}{L} \rceil$ sentences $\{s_1 \dots s_N\}$ with the same number of words L. In our experiments, the sentence length L was empirically tuned based on training datasets.

For sentence $\mathbf{s}_i = \{w_1^i \dots w_L^i\}$, let $\text{comp}(\mathbf{s}_i) = \{c_1^i \dots c_{\tilde{L}_i}^i\}$ be its component characters/syllables sequence. We define the subword-level representation $\text{sub}(\mathbf{s}_i|n)$ of sentence $\mathbf{s}_i$ as the overlapping n-gram subsequence of characters/syllables:

$$\text{sub}(\mathbf{s}_i|n) = \{c_{p \leftrightarrow p+n-1}^i\}_{p=1}^{\tilde{L}_i - n+1} = \{c_{1 \leftrightarrow n}^i, c_{2 \leftrightarrow n+1}^i, c_{3 \leftrightarrow n+2}^i, \dots\}, \tag{1}$$

where $\tilde{L}_i$ is the number of subwords in sentence $\mathbf{s}_i$. $c_{p \leftrightarrow p+n-1}^i$ denotes the subwords subsequence in $\text{comp}(\mathbf{s}_i)$ starting from the pth to the $p+n-1$th subwords, can be viewed as a *subword* representation, where n refers to the number of local components used to compose a subword. The purpose of overlapping is to reduce the possibility of missing useful information and to provide more chances for partial matching. In order to maintain the finer granularity of the representation, n should not be very large. As shown in Fig. 1(b), we restricted $n \le 4$.

Edge Cutoff. To construct the weighted edge set $\mathcal{E}$, we need to choose a proper edge link range. The same topic, e.g., a breaking news, may be intermittently reported from different angles for several times in a program. In story segmentation, these discontinuous reoccurrences of the same topic should be labeled

as different stories, otherwise those inbetween stories would be falsely missed. Therefore, an appropriate edge *cutoff*, properly balancing lone-term correlation and short-term discrimination, is more applicable to news story segmentation. In practice, we set up an edge cutoff value τ and simply discard those nodes-links whose distances exceed the threshold, See Fig. 1(a) for an example of the graph embedding with cutoff value $\tau = 2$.

Subwords Similarity. For two connected sentences $\mathbf{s}_i$ and $\mathbf{s}_j$ in the graph embedding, we assign the edge weight $\omega(i, j)$ as the exponential cosine similarity at subword level:

$$\omega(i,j) = \exp\big(\cos(\mathbf{f}_i, \mathbf{f}_j)\big) = \exp\Big(\frac{\mathbf{f}_i \cdot \mathbf{f}_j}{\|\mathbf{f}_i\|\|\mathbf{f}_j\|}\Big). \tag{2}$$

Note that, the n-gram representation of subwords may exponentially increase the subword vocabulary size. To make the similarity computation tractable, in practice, the subwords frequency vectors $\mathbf{f}_i$ and $\mathbf{f}_j$ are derived based on the local vocabulary $\mathcal{D}_{ij}$ instead of the global one, where the local vocabulary $\mathcal{D}_{ij}$ is composed of all the subwords occurred in $\mathrm{sub}(\mathbf{s}_i|n)$ and $\mathrm{sub}(\mathbf{s}_j|n)$.

Sentence similarities are inclined to be high within the same story and low at story boundaries. To alleviate this, in Eq. (2), we use *temporally smoothed* frequency vectors instead of the original ones to compute the sentence similarity $\tilde{\mathbf{f}}_i = \frac{1}{Z}\sum_{p=i-\frac{T}{2}}^{i+\frac{T}{2}} \exp\Big(-\frac{|p-i|}{\sigma}\Big)\mathbf{f}_p$, where σ controls the degree of smoothing, T is the size of sliding window, and Z is the constant normalization factor.

2.2 Self-validated Normalized Cuts

Dealing with multi-class tasks with different misclassification costs of classes is harder than dealing with two-class ones [5]. For a particular story number K, the dynamic programming normalized cuts(DP-NCuts) solution can efficiently produce a globally optimal K-partitioning to the input news transcript. In the next, we show how to enable the DP-NCuts method to self-validated story segmentation using the general principle of Occam's razor with reasonable complexity.

A Probabilistic Formulation. In order to seek the best segment number $\hat{K}$ and an optimal linear $\hat{K}$-labeling $\hat{X} = \{\hat{x}_1, \ldots, \hat{x}_N\}$ to each node of $\mathcal{G}$ with $\hat{x}_i \in \{1, \ldots, \hat{K}\}$, by maximizing the following posterior probability:

$$(\hat{K}, \hat{X}) = \arg\max_{K,X} \Pr(X, K \,|\, \mathcal{G}) = \arg\max_{K,X} \Pr(X \,|\, \mathcal{G}, K)\Pr(K), \tag{3}$$

where $\Pr(X, K \,|\, \mathcal{G})$ is the joint posterior likelihood of labeling X and segment number K given the observation; $\Pr(X \,|\, \mathcal{G}, K)$ measures the segmentation goodness; $\Pr(K)$ is the prior preference of story numbers. From Eq. (3), the self-validated story segmentation converts to a joint optimization problem. Due to the efficiency and efficacy of the non self-validated DP-NCuts algorithm, we simplify the formulation of self-validated labeling as:

$$(\hat{K}, \hat{X}) = \arg\max_{K} \Pr(K)\big[\arg\max_{X} \Pr(X \,|\, \mathcal{G}, K)\big] \tag{4}$$

$$= \arg\max_{K,X} \Pr(\hat{X}(K) \,|\, \mathcal{G}, K)\Pr(K), \tag{5}$$

where $\hat{X}(K) = \arg\max_X \Pr(X \mid \mathcal{G}, K)$ is the optimal K-labeling of $\mathcal{G}$, and $\Pr(\hat{X}(K) \mid \mathcal{G}, K)$ is the corresponding maximum K-labeling likelihood. Note that the joint optimization of K and X in Eq. (3) is decoupled in Eqs. (4)–(5).

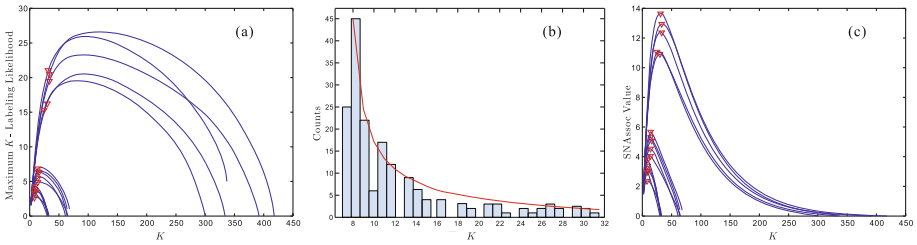

Fig. 2. Exemplar curves of K-labeling likelihood, prior of story numbers and the SNCuts score: (a) the maximum K-labeling likelihood $\Pr(\hat{X}(K) \mid \mathcal{G}, K)$ curves of TDT2 dataset; (b) the empirical histogram of TDT2 corpus and the fitted exponential distribution (red curve); (c) the SNAssoc curves of the transcripts shown in (a). The red triangles in (a) and (c) indicate the real story number for each transcript. (Color figure online)

Maximum K-Labeling Likelihood. $\Pr(\hat{X}(K) \mid \mathcal{G}, K)$ In Eq. (5), $\Pr(\hat{X}(K) \mid \mathcal{G}, K)$ indeed measures the goodness of the optimal K-segmentation of $\mathcal{G}$. We can naturally define $\Pr(\hat{X}(K) \mid \mathcal{G}, K)$ as the sum of normalized intra-sentence associations, so smaller NCuts value corresponds to better K-labeling to the $\mathcal{G}$. Thus,

$$\Pr(\hat{X}(K) \mid \mathcal{G}, K) \propto \sum_{k=1}^{K} \frac{\text{assoc}(\hat{\mathbf{s}}_k)}{\text{vol}(\hat{\mathbf{s}}_k)} = K - \text{NCuts}\big(\hat{X}(K)\big), \tag{6}$$

where $\text{assoc}(\hat{\mathbf{s}}_k)$, $\text{vol}(\hat{\mathbf{s}}_k)$ indicate the association and volume of the optimal sentence $\hat{\mathbf{s}}_k$. In Eq. (6), $\text{NCuts}(\hat{X}(K)) = \sum_{k=1}^{K} \text{NCuts}(\hat{\mathbf{s}}_k)$ denotes the minimum NCuts value of K-segmentations of $\mathcal{G}$. There are two important properties of $\Pr(\hat{X}(K) \mid \mathcal{G}, K)$. First, a better K-labeling X to graph $\mathcal{G}$ has larger likelihood value $\Pr(X \mid \mathcal{G}, K)$. Second, as shown in Fig. 2(a), the value of $\Pr(\hat{X}(K) \mid \mathcal{G}, K)$ is quickly increases first as K becomes larger, then slowly goes down after some critical point to penalize fragmental segments in the labeling $\hat{X}(K)$.

General Exponential Prior. $\Pr(K)$ The prior probability of story number K should reflect the empirical distribution of story numbers in real data, and respect the general principle of Occam's razor [6]. As shown in Fig. 2(b), in real-world transcripts with unfixed lengths, the story number in a transcript approximately follows an exponential distribution:

$$\Pr(K) \propto \alpha^K, \quad \text{with } 0 < \alpha < 1, \tag{7}$$

where α is the scaling parameter that controls the suppression strength to the possibility of choosing larger K. We believe that such exponential prior reflects

the similar fact that described by the well-known power-law distribution. The exponential prior defined in Eq. (7) has similar property of the power-law, and is empirically more suitable to the task of news story segmentation. On the other hand, the rationale of the exponential prior of K can also be explained as a natural respect to the general principle of Occam's razor, since it clearly favors smaller K and suppresses larger ones.

SNCuts Score. From Eqs. (3)–(7), we can define a new graph partitioning criterion, namely self-validated NCuts, which takes accounts of both the segmentation goodness and the labeling cost. We use the posterior energy to measure the segmentation quality. Accordingly, we define the SNCuts score of labeling X as $\mathrm{SNCuts}(X \mid \mathcal{G}) = -\log(\Pr(X, K(X) \mid \mathcal{G}))$, thus yielding

$$\begin{aligned} \mathrm{SNCuts}(X \mid \mathcal{G}) &= -\log\big(\mathrm{SNAssoc}(X \mid \mathcal{G})\big) \\ &= -\log\Big(\big[K(X) - \mathrm{NCuts}(X)\big]\,\alpha^{K(X)}\Big), \end{aligned} \tag{8}$$

where $\mathrm{SNAssoc}(X \mid \mathcal{G}) = \Pr(X \mid \mathcal{G}, K(X))\Pr(K(X))$ indicates the posterior likelihood of labeling X. Clearly, the optimal labeling $\hat{X}$ to graph $\mathcal{G}$ corresponds to the minimum SNCuts and maximum SNAssoc value. Note that, α balances the relative importance of segmentation goodness and the log labeling cost in the energy function of Eq. (8). Figure 2(c) shows that with an appropriate scaling parameter α, the real story numbers $\hat{K}$ approximately coincide with the points of maximum SNAssoc and minimum SNCuts values.

3 Experiments

3.1 Corpus and Experimental Setup

We carry out the experiments on two benchmark Mandarin broadcast news corpora, TDT2 [19] and CCTV [1]. The TDT2 Mandarin corpus [19] contains about 53 h of VOA Chinese broadcast news audio (177 recordings in total) from Feb to June, 1998. We separate the corpus into two non-overlapping subsets: a training set of 90 recordings (1321 boundaries) for parameter tuning, and a test set of 87 recordings (1262 boundaries) for evaluation. The CCTV corpus records 71 news episodes of 27 h of CCTV (i.e., China Central Television) Mandarin broadcast news from July to Dec, 2007. Due to the particular news production rules of CCTV, we further label CCTV news stories as either detailed ('-f') or brief ('-s') ones. Similar to TDT2, we separate the CCTV corpus into a training set with 40 audio files (1209 story boundaries) and a test set with 31 audio files (892 story boundaries). Accord with the TDT2 convention [14], we consider a detected story boundary on CCTV corpus as being correct if it lies in a K-word-length tolerance window on each side of the exact boundary position ($K = 10$ for brief stories, and $K = 30$ for detailed stories).

In all our experiments, we assess story segmentation accuracy using the F1-measure, i.e., $\frac{2 \cdot \mathrm{Recall} \cdot \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}}$. For a particular word or subword level, we use

two forms to represent a news transcript $\mathcal{T}$: (1) the sequence of Chinese characters (denoted by char for short) and (2) the sequence of base-syllables (denoted by syll).

3.2 Comparison to State-of-the-Art Methods

We use TDT2 corpus to compare the proposed SNCuts approach with nine state-of-the-art story segmentation methods: (1) TextTiling (TT) [8]; (2) latent semantic analysis (LSA) [4]; (3) LSA-TextTiling (LSA-TT) [22]; (4) lexical chains (LC) [2]; (5) conditional random field (CRF) [20]; (6) maximum lexical cohesion (MLC) [11]; (7) LE-TextTiling (LE-TT) [21]; (8) spectral clustering (SC) [21]; (9) LE-DP [21]. To maintain the fairness of comparison, for all competing methods, we compare the best segmentation results reported by their authors. On CCTV corpus, besides comparing the best segmentation accuracy, we further investigate the behavior and sensitivity of different methods for erroneous transcripts with increasing ASR error rates.

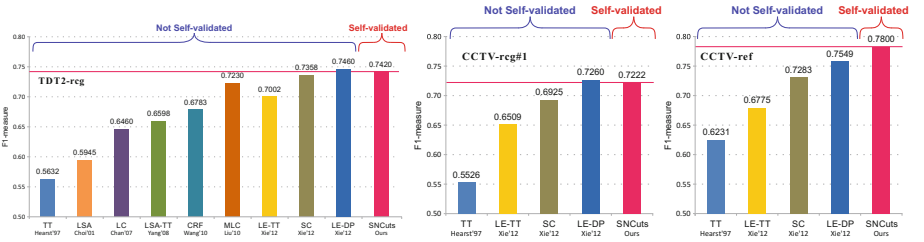

Fig. 3. Comparative best performance of the proposed SNCuts approach and state-of-the-art story segmentation methods on (a) TDT2-rcg dataset, (b) CCTV-rcg#1 dataset (left) and CCTV-ref dataset (right).

Figure 3(a) shows that almost all existing story segmentation methods are not self-validated, which is a major drawback. Besides self-validation, the proposed SNCuts approach has achieved the second highest accuracy and is only 0.004 less than the highest one on TDT2-rcg dataset. As shown in Fig. 3(b), on CCTV-rcg#1, SNCuts has also obtained the second highest accuracy with 0.0038 disparity to the best one; and on CCTV-ref, SNCuts has achieved the highest F1-measure 0.78 that is 0.0251 higher than the second best one.

In Table 1, we compare the relative degradation ratio of different methods (with valid reported performance) for transcripts with increasing ASR errors. We can see that for all methods, increasing ASR errors may degrade their segmentation accuracy. On CCTV-rcg#1, both SNCuts and NCuts exhibit more robustness to ASR errors than TextTiling (TT) [8]. LE-DP [21] has the lowest degradation ratio on CCTV-rcg#1. But due to the lack of results on CCTV-rcg#2 and CCTV-rcg#3 [21], we cannot further check its robustness for higher ASR error rates. SNCuts has obtained comparable degradation ratio to NCuts.

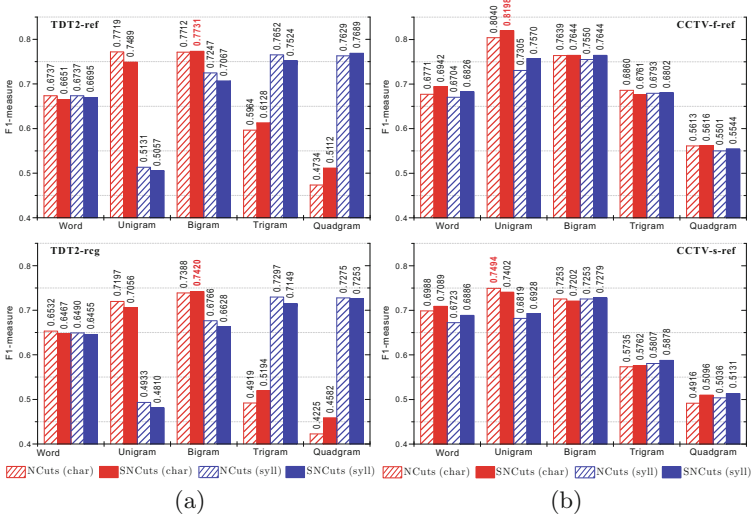

Fig. 4. Comparative segmentation performance of NCuts and SNCuts on (a) TDT2 corpus, (b) CCTV-ref. Best accuracies are shown in red. (Color figure online)

3.3 SNCuts Vs. NCuts

Since our SNCuts is a self-validated extension to the NCuts criterion, we specifically interest in comparing their *best* capabilities in story segmentation on different datasets. For this purpose, we first use TDT2 and CCTV training sets to individually seek the best parameters with the highest average F1-measure, and then compare their accuracies on test sets. For comparison fairness, we conduct automatic parameter-tuning using the Differential Evolution (DE) algorithm [18] with the same (reasonably large enough) number of generations and the same proper parameters ranges.

Accuracy. We first evaluate the segmentation accuracy. As shown in Fig. 4(a), for both TDT2-ref and TDT2-rcg, the best accuracies are achieved by SNCuts using Bigram (char) representation. In some cases (e.g., for n-gram subwords with $n \geq 2$), the NCuts algorithm [12,23] fed by the true story number K may result in worse segmentation than SNCuts does. This is mainly due to an inherent limitation of NCuts criterion that tends to generate false-positive segmentation boundaries and miss correct ones [6]. As validated by our experiments, besides self-validation, SNCuts also helps to amend the inherent limitation of NCuts criterion.

Figures 4(b) and 5 respectively show the detailed best segmentation results of SNCuts and NCuts at every word/subword-level via either 'char' or 'syll' representations on CCTV-rcg datasets with increasing ASR error rates. Similarly, at some particular levels, SNCuts can even outperform NCuts with the correct story number K as an input parameter. And, in most cases, SNCuts can achieve comparable accuracy with NCuts.

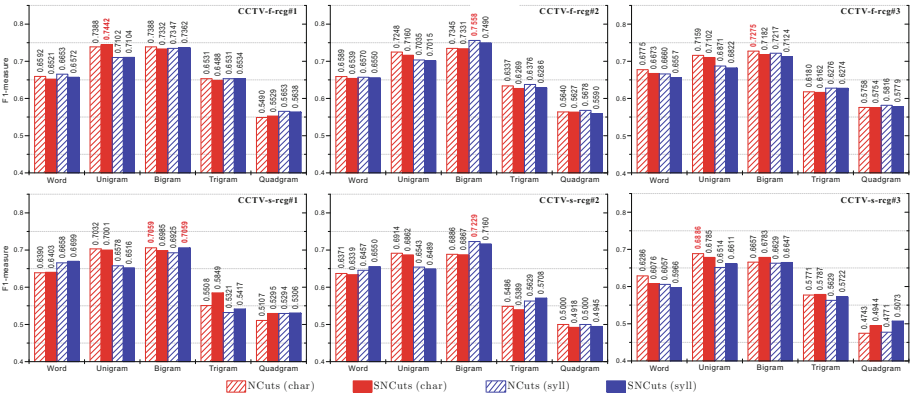

Fig. 5. Comparative segmentation performance of NCuts and SNCuts on CCTV-rcg#1, CCTV-rcg#2, and CCTV-rcg#3. Best accuracies are shown in red. (Color figure online)

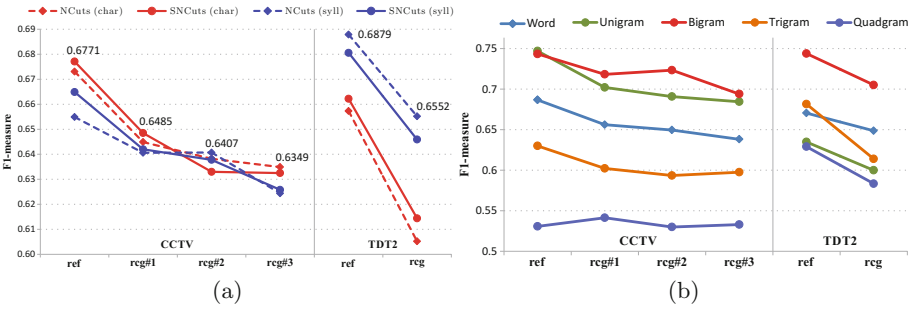

Fig. 6. Influence of ASR errors on story segmentation performance. (a) Mean segmentation accuracy of NCuts and SNCuts on CCTV and TDT2 corpora. (b) Average story segmentation accuracy for different word/subword-levels on CCTV and TDT2 corpora.

Influence of ASR Errors. Figure 6(a) compares the average segmentation accuracy of NCuts and SNCuts for benchmark datasets with increasing ASR error rates. Averagely speaking, both mean segmentation accuracies and the degradation ratios of NCuts and SNCuts are comparable. Specifically, on CCTV corpus, SNCuts (char) obtained the best accuracy for low ASR errors, and SNCuts (syll) performed the best for higher ASR errors. On TDT2 corpus, NCuts (syll) achieved the highest mean accuracy; and for 'char' representation, SNCuts performed better than NCuts. We then evaluate the robustness of word/subword-levels to ASR errors in Fig. 6(b). The degradation effect of ASR error is also evident. Among all word/subword-levels, bigram and unigram performed the best for CCTV corpus; while for TDT2 corpus, bigram evidently outmatched the other levels. On both corpora, we can clearly see the robustness of subword representations to ASR errors.

Table 1. Comparison of best segmentation accuracy (Acc.) and relative degradation (Degrad.) ratio of different methods on CCTV corpus. For each dataset, the best accuracy is in red font.

Approach	CCTV-ref	CCTV-rcg#1		CCTV-rcg#2		CCTV-rcg#3	
	Acc.	Acc.	Degrad. ratio	Acc.	Degrad. ratio	Acc.	Degrad. ratio
TT [8]	0.6231	0.5526	11.31%	-	-	-	-
LE-TT [21]	0.6775	0.6509	3.93%	-	-	-	-
SC [21]	0.7283	0.6925	4.92%	-	-	-	-
LE-DP [21]	0.7549	0.7260	3.83%	-	-	-	-
NCuts	0.7767	0.7224	6.99%	0.7393	4.82%	0.7023	9.85%
SNCuts	0.7800	0.7222	7.41%	0.7325	6.09%	0.6983	10.47%

4 Conclusions

In this paper, we have proposed a simple yet effective approach, namely n-gram subword SNCuts, to accurately segmenting Chinese broadcast news via inaccurate lexical cues. Our approach can automatically determine the story number, and can properly take care of inter- and intra-story similarity. Extensive experiments have validated that our approach can achieve comparable or better accuracy to state-of-the-art non-self-validated methods on benchmark corpora.

Besides accuracy and efficiency, self-validation is also an important requirement in segmentation, especially in the era of Big Data, to automatically handle huge number of media data. At last, we believe properly encoding soft similarity measurements in the classical cosine similarity may further improve the segmentation performance.

References

1. CCTV Corpus: Story segmentation and topic detection of CCTV Mandarin broadcast news (2010)
2. Chan, S.K., Xie, L., Meng, H.M.L.: Modeling the statistical behavior of lexical chains to capture word cohesiveness for automatic story segmentation. In: INTERSPEECH, pp. 2408–2411 (2007)
3. Choi, F.: Advances in domain independent linear text segmentation. In: NAACL, pp. 26–33 (2000)
4. Choi, F., Wiemer-Hastings, P., Moore, J.: Latent semantic analysis for story segmentation. In: EMNLP (2001)
5. Feng, W., Huang, W., Ren, J.: Class imbalance ensemble learning based on the margin theory. Appl. Sci. **8**(5), 815 (2018)
6. Feng, W., Jia, J., Liu, Z.Q.: Self-validated labeling of Markov random fields for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **32**(10), 1871–1887 (2010)
7. Guo, Q., Sun, S., Ren, X., Dong, F., Gao, B.Z., Feng, W.: Freqeuncy-tuned active contour model. Neurocomputing **275**(31), 2307–2316 (2018)

8. Hearst, M.: TextTiling: segmentation text into multi-paragraph subtopic passages. Comput. Linguist. **23**(1), 33–64 (1997)
9. Kyoto University: Multipurpose large vocabulary continuous speech recognition engine - Julius (rev 3.2) (2001)
10. Lee, L.S., Chen, B.: Spoken document understanding and organization. IEEE Signal Process. Mag. **22**(5), 42–60 (2005)
11. Liu, Z., Xie, L., Feng, W.: Maximum lexical cohesion for fine-grained news story segmentation. In: INTERSPEECH (2010)
12. Malioutov, I., Barzilay, R.: Minimum cut model for spoken lecture segmentation. In: ACL, pp. 25–32 (2006)
13. Nie, X., Feng, W., Wan, L., Xie, L.: Measuring similarity by contextual word connections in Chinese news story segmentation. In: ICASSP (2013)
14. NIST: The topic detection and tracking phase 2 (TDT2) evaluation plan, version 35 (1998)
15. Ren, J., Jiang, J.: Hierarchical modeling and adaptive clustering for real-time summarization of rush videos. IEEE Trans. Multimed. **11**(5), 906–917 (2009)
16. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000)
17. Stokes, N., Carthy, J., Smeaton, A.: SeLeCT: a lexical cohesion based news story segmentation system. J. AI Commun. **17**(1), 3–12 (2004)
18. Storn, R., Price, K.: Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. **11**, 341–359 (1997)
19. TDT2 Corpus: Topic detection and tracking phase 2, July 2000. http://projects.ldc.upenn.edu/TDT2/
20. Wang, X., Xie, L., Ma, B., Chng, E.S., Li, H.: Modeling broadcast news prosody using conditional random fields for story segmentation. In: APSIPA ASC (2010)
21. Xie, L., Zheng, L., Liu, Z., Zhang, Y.: Laplacian Eigenmaps for automatic story segmentation of broadcast news. IEEE Trans Audio Speech Lang. Process. **20**(1), 264–277 (2012)
22. Yang, Y., Xie, L.: Subword latent semantic analysis for texttiling-based automatic story segmentation of Chinese broadcast news. In: ISCSLP, pp. 358–361 (2008)
23. Zhang, J., Xie, L., Feng, W., Zhang, Y.: A subword normalized cut approach to automatic story segmentation of Chinese broadcast news. In: Lee, G.G., Song, D., Lin, C.-Y., Aizawa, A., Kuriyama, K., Yoshioka, M., Sakai, T. (eds.) AIRS 2009. LNCS, vol. 5839, pp. 136–148. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04769-5_12