Multilingual Multi-document Summarization with Enhanced hLDA Features

Taiwen Huang, Lei Li, Yazhao Zhang

Beijing University of Posts and Telecommunications, Beijing 100876, China
zhidao2010@bupt.edu.cn, leili@bupt.edu.cn, yazhao@bupt.edu.cn

Abstract. This paper presents the state of art research progress on multilingual multi-document summarization. Our method utilizes hLDA (hierarchical Latent Dirichlet Allocation) algorithm to model the documents firstly. A new feature is proposed from the hLDA modeling results, which can reflect semantic information to some extent. Then it combines this new feature with different other features to perform sentence scoring. According to the results of sentence score, it extracts candidate summary sentences from the documents to generate a summary. We have also attempted to verify the effectiveness and robustness of the new feature through experiments. After the comparison with other summarization methods, our method reveals better performance in some respects.

KeyWords : multilingual multi-document summarization ·sentence scoring ·hLDA modeling

1 Introduction

Facing the explosive expansion of network information, users have to spend vast time to discover what they want. It is a significant challenge to find an effective tool to filter the unnecessary information and provide a precise and concise synopsis. Hence document summarization has become one of the important technologies for research. Researchers have worked on different topics, including single document summarization, multi-document summarization and multilingual multi-document summarization which is just the topic of this paper and also the most difficult one.

The field of document summarization has accumulated a lot of research results after 60 years of research. Summarization methods can be classified into extractive and abstractive ones. Extractive summarization aims to select important sentences from the original document and reorganize these sentences according to their order, while abstractive summarization aims to understand the original text and re-tell it like human. The mainstream is extractive summarization currently because abstractive summarization needs effective and deep natural language processing technologies. For extractive summarization, the representative algorithms involve graph model[1], sentence clustering[2], linear programming[3], topic model[4] and so on.

This paper focuses on the extractive summarization and topic model method. In recent years, we are constantly using and mining hLDA[5] features. The hLDA algorithm is selected as our main algorithm. hLDA has many advantages. On the one hand, it is a language-independent algorithm, so we don't need a lot of professional linguistic knowledge. On the other hand, it is a corpus-oriented algorithm, it can deal with large scale corpus. hLDA can be an excellent multilingual multi-document modeling algorithm. Many hLDA scholars[6][7][8] are just exploring the cluster and document modeling ability of hLDA, yet they haven't made clear the semantic features behind the model. Based on the research of hLDA, this paper proposes a new level distribution feature from it. In combination with other traditional features, we build our new summarization system with a better performance. After comparison with other methods, it shows the superiority of our system in some respects.

2 Background

2.1 hLDA

The hLDA (Blei et al. 2004) can organize latent topics of sentences in a set of documents and their relations into a hierarchical tree. The structure of the tree is determined by the documents and some parameters. A node in this tree represents a latent topic, and words are assigned to a node with some probability. Hence, a topic can be interpreted as probability distribution over words. The following figure 1 shows an example hLDA tree with the depth of 3. Please refer to (Blei et al. 2004) for more details.

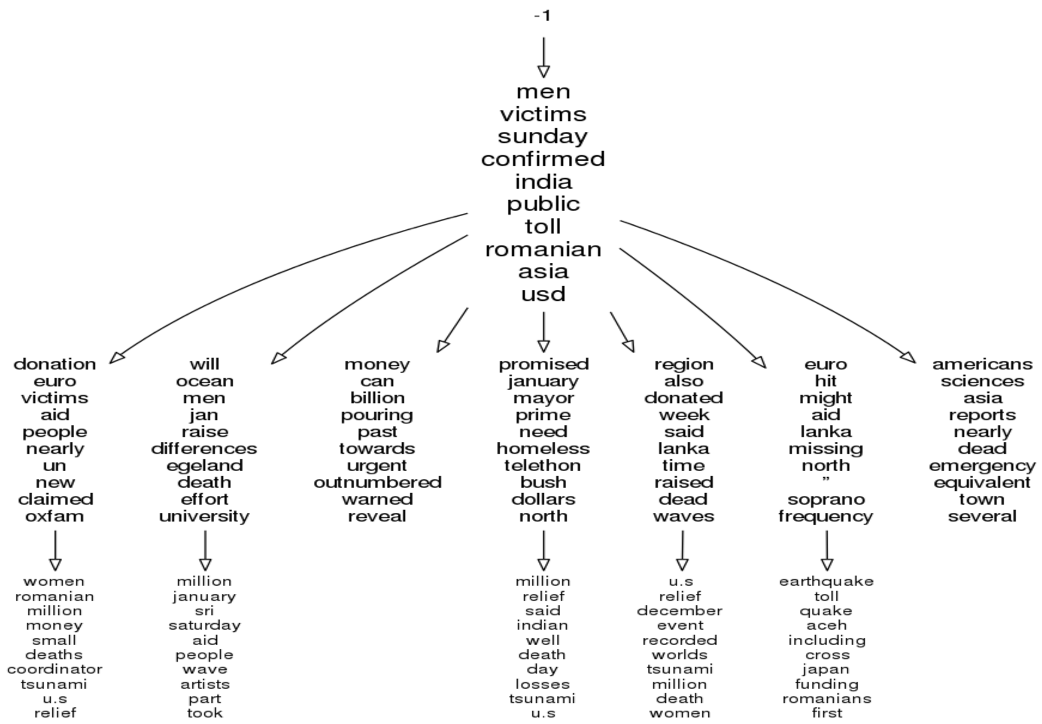

Fig. 1. hLDA modeling result example

Table 1. hLDA parameter setting

Parameter	Setting	Parameter	Setting
ETA	1.2,0.5,0.05	GEM_SCALE	100
GAM	1.0,1.0	SCALING_SHAPE	1
GEM_MEAN	0.5	SCALING_SCALE	0.5

Owing to using the sentence as the input unit of hLDA algorithm, every sentence will be allocated into a path from the root to a leaf in the tree as shown with arrows in figure 1. Sentences having the same path will be considered to be related to the same theme. In light of our previous research experience, we set the depth of the tree to 3. There are 6 parameters in total to control the structure of the tree. The parameter settings in our system can be found in Table 1. More details have been presented in the research of Heng Wei[9].

2.2 Experimental Data

MultiLing (http://multiling.iit.demokritos.gr) is a special session in SIGdial 2015, which holds 4 tasks, i.e., MMS (Multilingual Multi-document Summarization), MSS (Multilingual Single-document Summarization), OnForumS (Online Forum Summarization) and CCCS (Call Centre Conversation Summarization). This multilingual multi-document summarization (MMS) (Giannakopoulos, 2015) task aims to evaluate the application of partially or fully language-independent summarization algorithms. It contains ten languages: Arabic, Chinese, Czech, English, French, Greek, Hebrew, Hindi, Romanian and Spanish.

The multi-document summarization task required participants to generate a fluent and representative summary from the set of documents describing an event sequence. The language of each document set belonged to one of the aforementioned set of languages and all the documents in a set were of the same language. The output summary was expected to be in the same language and between 240 and 250 words. The task corpus is based on a set of WikiNews English news articles comprising 15 topics, each containing ten documents. Each English document was translated into the other nine languages to create sentence-parallel translations. Please refer to [10] for more detailed introduction of MultiLing2015.

The experimental data of this paper is the test data of MMS task. Each language contains 10 to 15 topics and each topic contains 10 documents.

3 System Architecture

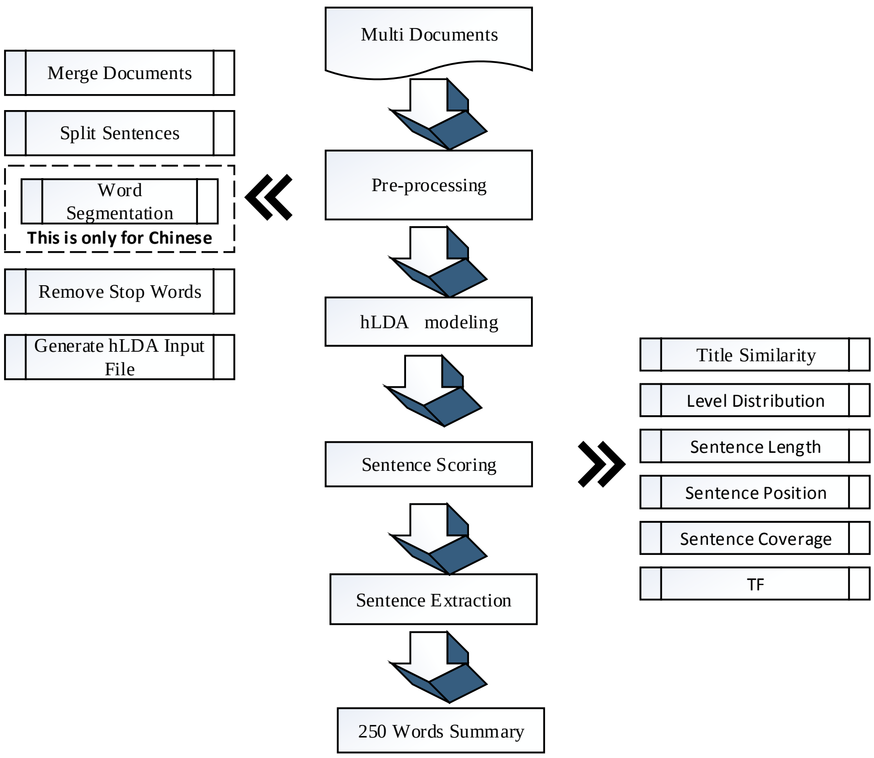

Fig. 2. Multilingual Multi-document summarization system framework

The framework for our system is shown in Figure 2. In particular, we only treat Chinese with word segmentation. The kernel module is constructing an hLDA model.

3.1 Pre-processing

1. Merge Documents
   For every topic of each language, we merge all the 10 documents into one big document. This big document is named as merged document. We will mainly extract summary from this merged document. The following pre-processing steps are all done upon the merged document, too.
2. Split Sentences
   Sentence is the processing unit. There are two special lines of title and date ending with no punctuation marks in the corpus. We add a full stop to avoid them being connected with the first following sentence. We then split sentences according to the ending punctuation marks like full stop, question mark and apostrophe.
3. Word Segmentation
   Since Chinese hanzi is not separated by spaces as other languages, we need to do word segmentation firstly.
4. Remove Stop Words
   We construct stop lists for all languages. For English and Chinese, the stop list contains punctuation marks and some functional words, while for other languages, it contains punctuation marks, which could unify the whole process easily. In fact, in the experiment, we compare the performance of the system of removing stop words with system of not removing stop words. Specific differences will be described in the experimental section.
5. Generate hLDA input File
   We build a dictionary for the remaining words, which are sorted in descending order according to their frequencies. This is a mapping from a word to a number varying from 1 to dictionary size. Finally we generate an input file for hLDA, in which each line represents a sentence:
   [number of words in the sentence] [wordNumberA]:[localfrequencyA] [wordNumberB]:[local frequencyB]...

3.2 hLDA Modeling

The hLDA algorithm we used in this paper is originally developed by Blei. (http://www.cs.princeton.edu/~blei/topicmodeling.html). Through hLDA modeling, we can get the final hLDA modeling tree information. The main three files are 'mode' file, 'mode.assign' file, and 'mode.levels' file. The information in these files is independent with language. Our level distribution feature is extracted from them. The only purpose that we utilize hLDA in this paper is to get the value of level distribution feature for sentence scoring. The other features for sentence scoring are independent with hLDA modeling process. In the next section, we will elaborate on introduction to all features we used in this paper.

4 Sentence Scoring

Our system selects 6 features to score sentences. In these 6 kinds of features, there are 5 kinds of features (except Level Distribution) that are used in single document summarization widely. The validity of these features has been demonstrated by many experiments and researchers. For each of these 6 features, we have also performed extensive experiments to illustrate its effectiveness in multilingual multi-document summarization.

(1)Level Distribution: this feature is just the achievement of our latest research. In this feature, we think that after hLDA modeling, if a word appears in more levels or more nodes, then it has a higher possibility to be involved into summary. Formula 4 is our scoring method.

Blei, the author of hLDA, has proposed that the lower level nodes in the hLDA modeling results are more abstract. Thus in previous years, we considered that the abstract word is more suitable to appear in summary. In light of this idea, our previous level scoring formula is:

$$S_{abstract} = a \times \frac{num(w_0)}{|s|} + b \times \frac{num(w_1)}{|s|} + c \times \frac{num(w_2)}{|s|} \qquad (1)$$

In formula 1, $num(w_0)$,$num(w_1)$ and $num(w_2)$ refer to the amount of words in a sentence assigned to each level respectively. Usually we set the three weights according to our own experience, which is a as 0.75, b as 0.25, and c as 0.1. We have utilized this feature for sentence scoring to implement several summarization systems. For example, TAC-2014 multi-document summarization system, multiLing-2015 summarization system, NLPCC-2015 text summarization system. Through these experiences, we find out that the effect of this feature and the performance is not satisfactory. Usually, it will reduce the performance of summarization system. After participating in the NLPCC2015 Shared Task4, we use NLPCC2015's corpus to perform more deeply exploration in features contained in hLDA modeling results. We find a more reasonable understanding of the abstract nature of the level. In hLDA modeling tree, the abstract node possibly means that the words appearing in the node either appear in a lot of contexts or their frequency is very low. In the high level of nodes, the words appearing in these nodes usually have some very specific context and even only one context. In our work, we have found out that most of the words in root node are low frequency words, while in the leaf nodes, although some words are high frequency ones, they always appear in a fixed context. Under the guidance of this understanding, a new law has been found in the hLDA modeling that if a word appears in more nodes and more levels, then the word has a higher probability to be selected into the summary. After more exploration of this rule, we propose a new formula of level feature for sentence scoring called level distribution. Formula 2 shows the result of our research.

$$\mathrm{S}_{Levels} = \sum_{i=0}^{N}(W_i T_{i-Distribution} + T_{i-NodeFrequency} + T_{i-Keywords}) \qquad (2)$$

In formula 2, N is the total number of words in a sentence. W_i is the weight of the level in the hLDA modeling tree where the word T_i is assigned to. $\mathrm{T}_{i-Distribution}$ is the distribution score of T_i. The distribution score is based on the law above. Every word will be mapped to a single score according to their distribution information in hLDA modeling tree. Mapping rule is defined through many controlled experiments. $T_{i-Keywords}$ is the keywords score of T_i. If T_i belongs to keywords, $T_{i-Keywords}$ will be 0.5, else 0. $\mathrm{T}_{i-NodeFrequency}$ is the node frequency score of T_i , it represents the frequency of T_i in the node where T_i is assigned to. Formula 3 reveals its calculation method.

$$T_{i-NodeFrequency} = \frac{counts(T_i)}{\sum_{i=0}^{V} counts(T_i)} \qquad (3)$$

Note that NLPCC 2015 summarization task is only for single Chinese document. We will demonstrate the validity of this feature through another paper. In this paper, we mainly attempt to find the effectiveness and robustness for multilingual multi-document summarization (MMS). Hence, we select

MultiLing2015's MMS test data as our experimental data. The most important point is that we abandoned the last item in formula 2. This is mainly because we cannot get the keywords of other languages except Chinese. So the final level distribution formula in this paper is as formula 4.

$$S_{Levels} = \sum_{i=0}^{N}(W_i T_{i-Distribution} + T_{i-NodeFrequency}) \quad (4)$$

We will specifically discuss the performance of level distribution feature in later section for experiments. In short, this feature greatly improves the performance of Chinese language. For some other languages, the effect is not obvious. The reason may be that this feature is originally obtained from NLPCC 2015 Chinese summarization task. Through some experimental exploration, we acquire the parameter settings in formula 4. Table 2 shows the final settings in this paper.

Table 2. Parameters of formula 4

W_0	W_1	W_2	Word score (0-1-2)	Word score (0-1)
0.5	0.1	0.4	4	1
Word score (0-2)	Word score (1-2)	Word score (0)	Word score (1)	Word score (2)
3	1	-1	0	0.5

In table 2, W_0 is the weight of the first level of hLDA modeling tree. W_1 is the weight of the second level of hLDA modeling tree. W_2 is the weight of the last level of hLDA modeling tree. For every word, we assign a global fixed score according to their distribution in hLDA tree. For instance, if a word appear in three layers, (0,1,2), then the word score will be set to 4. If a word only appears in the first layer (0), then the word score will be -1.

It must be noted that we do not elaborate on how we find the law of Level Distribution and how we adjust parameters in this paper. Limited to the length of this paper, we will discuss the law of Level Distribution and verify the assumption in another paper in detail. In fact, the discovery of Level distribution law is based on the statistical analysis of the human summary. Through performing hLDA modeling on human summary and analyzing the law of these hLDA modeling results, we acquire the law of Level distribution. In this paper, we just want to demonstrate that the Level distribution is effective in multilingual multi-document summarization.

(2)Title Similarity: For news, the title sentence can be very good at revealing the central theme. So we calculate the tf-idf similarity between each sentence with the title sentence. In every topic, there are 10 documents. We choose the first sentence of each article as the title sentence. The similarity score of each sentence is calculated as the similarity between the sentence and the title sentence. Formula 5 reveals the calculation method.

$$S_{TS} = \frac{\sum_{i=1}^{n}(\boldsymbol{Sen}_i \times \boldsymbol{Title}_i)}{\sqrt{\sum_{i=1}^{n}(\boldsymbol{Sen}_i)^2} \times \sqrt{\sum_{i=1}^{n}(\boldsymbol{Title}_i)^2}} \quad (5)$$

Sen is the document vector representation of a sentence. *i* is the *i*th dimension of the vector. ***Title*** is the document vector representation of the title sentence.

(3)Sentence Length: We believe that the length of the sentence in the news is subjected to Gaussian distribution. If a sentence is closer to the average length, then it is more likely to be selected into summary.

$$S_{senLen} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(L_i-\mu)}{2\sigma^2}} \tag{6}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(L_i-\mu)^2}{n} \tag{7}$$

μ is the average value of sentence length in all documents of a topic. L_i is the length of a sentence at position *i* of the merged documents.

(4)Sentence Position: Here we use the monotonous sentence position scoring method. Usually, if the position of a sentence is closer to the starting of news, the sentence is more likely to be selected into summary. Note that our sentence position score is not referring to the sentence position in the merged document. The sentence position score is just calculated as the sentence position score in the original single document the sentence belongs to. Hence, there are possibly many sentences with the same sentence position score. If two documents have the same amount of sentences, the sentence in the same position will have the same sentence position score.

$$S_{senPos} = \frac{n-i+1}{n} \tag{8}$$

n is the total amount of sentences in the single document the sentence belongs to. *i* represents the sentence position in the single document the sentence belongs to.

(5)Sentence Coverage: If a word appears in many sentences, then the possibility of the sentence containing the word being selected into summary will be higher.

$$S_{\text{coverage}} = \frac{\sum_{i=1}^{|s|} \frac{num_S(t_i)}{n}}{|s|} \tag{9}$$

(6)TF: this feature is referred from [11].The system won the first prize in the NLPCC 2015 Shared Task 4.

$$\text{TF}(t_i) = \frac{m_i}{\sum_{j=1}^{N} m_j} \tag{10}$$

$$S_{TF} = \sum_{t\in Sen} TF(t) \tag{11}$$

Where *N* represents the number of different words that are included in the article. m_i is the times the word t_i appears in the article. t_i is the *i*th word in the article.

At last, we combine all features' scores mentioned above. Formula 12 is the final score of one sentence. $para_{1\sim6}$ are the weights of each feature score. In the experimental section, we will discuss their different influences in the performance of summarization.

$$\text{S} = para_1 * S_{TS} + para_2 * S_{Levels} + para_3 * S_{senLen} + para_4 * S_{senPos} + para_5 * S_{coverage} + para_6 * S_{TF} \tag{12}$$

5 Sentence Extraction

The following algorithm describes the sentence extraction strategy in our system.

```
Input: Merge documents sentence set mergeSentenceSet
Output: a summary Summary
begin
First we should Determine para1~6
for each sentence Sen in mergeSentenceSet:
        calculate the final score of every sentence
end for;
for each sentence Sen in mergeSentenceSet:
        ranking(Sentence,SentenceScore);//Descending order
end for
for each sentence Sen in mergeSentenceSet:
        select the top 1 sentence topSen from mergeSentenceSet
        totalLength += topSenLength
        if(totalLength > 250):
                cut the topSen until the totalLength can equal to 250 when adding topSen
                to Summary
                break;
        else:
                add topSen into Summary
                remove topSen from mergeSentenceSet
end for
for each sentence Sen in Summary:
                ordering(Sentence,SentencePos);//Ascending order
end for
Output Summary
end
```

Fig. 3. Sentence extraction algorithm

Our algorithm firstly selects a set of parameters to calculate the final score of every sentence. Then we rank the sentences of the merged document by function ranking() according to the ultimate significance score of each sentence calculated by formula (12). We control the final total length of the summary to the limit of 250 words using the parameter totalLength and order the selected sentences in the generated summary with function ordering() according to the sentence position.

6 Experiments

We use the Rouge package[12][13] to evaluate effectiveness of our approach vs. other summarization methods. The results of other summarization methods are published by MultiLing2015's organizer. The experimental data is MultiLing2015 MMS test data. We evaluate the results of 63 different feature combinations of parameters (each para in $para_{1\sim6}$ is either 0 or 1). Through these experiments, we demonstrate the effect of level distribution feature, especially for Chinese summarization. For all ten languages, we evaluate their Rouge scores. Table 3 shows the Rouge scores of summaries generated using only one feature.

From table 3, a bad fact we can find is that almost all Rouge values in Levels column are worst. That's to say, the single Level Distribution feature seems to have no reason to exist. Nevertheless, when we try to combine this feature with other features, a very surprising result appears. This surprising and good information is that this feature can improve the comprehensive result when combined with others. Furthermore, it has positive effect with other features in most situations. There is no same phenomenon on other features. Table 4 reveals part of this phenomenon.

Table 3. single feature summarization result (Owing to the limit of this paper space, all rouge scores in this paper except table 6 are ROUGE-1 F, Ts = Title Similarity, Levels = Level Distribution,SenLen = Sentence Length,SenPos = Sentence- Position,SenCov = Sentence Coverage)

	Ts	Levels	SenLen	SenPos	SenCov	TF
All Languages	0.36303	**0.30325**	0.31687	0.35115	0.34012	0.31642
Arabic	0.23775	**0.13752**	0.14194	0.16408	0.21231	0.19236
Chinese	0.49661	**0.30538**	0.34594	0.53596	0.31484	0.36193
Czech	0.44861	**0.43309**	0.45452	0.45027	0.44519	0.43403
English	0.42502	**0.39756**	0.39852	0.39844	0.41834	0.40326
French	0.47388	**0.43186**	0.44607	0.46911	0.46888	0.44210
Greek	0.25339	**0.18067**	0.20507	0.22718	0.29537	0.15405
Hebrew	0.25719	**0.19954**	0.20060	0.29113	0.21739	0.24119
Hindi	0.07933	**0.05629**	0.10094	0.04426	0.06948	0.04846
Romanian	0.43059	**0.38766**	0.39540	0.41449	0.42047	0.40081
Spanish	0.52443	**0.45659**	0.46978	0.50906	0.47676	0.45796

Table 4. Level Distribution feature combined with other features (All mixture ratios are 1)

	Ts + Levels	SenLen+ Levels	SenPos+ Levels	SenCov+ Levels	TF+ Levels
All Languages	0.35481	0.32714	0.36608	0.35647	0.33786
Arabic	0.24917	0.19710	0.23659	0.23261	0.21748
Chinese	0.47326	**0.40162**	**0.61449**	**0.54868**	**0.52351**
Czech	0.44943	0.44195	0.47419	0.45349	0.43256
English	0.41500	0.41063	0.43187	0.42530	0.41703
French	0.47811	0.43080	0.47870	0.45036	0.44211
Greek	0.24394	0.21578	0.17089	0.21187	0.15830
Hebrew	0.23027	0.19130	0.24472	0.27597	0.26993
Hindi	0.08899	0.07778	0.09419	0.07273	0.03997
Romanian	0.41866	0.40842	0.43144	0.41035	0.40280
Spanish	0.50755	0.47206	0.50846	0.47696	0.46728

The results in table 4 show that the All Language Rouge score will increase when Level Distribution is combined with other features except Title Similarity. Note that this phenomenon of performance increasing is not unilateral. For example, when Sentence Length is combined with Level Distribution, the final result not only surpasses the result using single Sentence Length, but also surpasses the result using single Level Distribution. For other features, the phenomenon of performance increasing is usually unilateral. What’s more, an important fact in table 4 is that this feature improves the performance of Chinese summarization greatly. This result has just verified our expectation. Because Level Distribution feature is origined from Chinese single document summarization. We find this feature through performing experiments on NLPCC2015 Shared Task 4 data.

Table 3 and Table 4 are the experimental results based on the corpus without removing stop words. Actually, stop words have a great impact on the performance of the summarization system. Especially for Level Distribution feature, the effect will be weaken if we do not remove stop words. In light of our research results, if we use the corpus with stop words to perform hLDA modeling, the stop words will have the similar distribution characteristics with those important words which Level Distribution feature prefers. Table 5 shows the summarization results of single feature based on removal of stop words.

Table 5. single feature summarization based on removal of stop words

	Ts	Levels	SenLen	SenPos	SenCov	TF
All Languages	0.36466	0.33745	0.31729	0.35115	0.35440	0.33688
Arabic	0.23775	0.18475	0.16756	0.16408	0.22202	0.19385
Chinese	0.52479	**0.53915**	0.33620	0.53596	0.49557	0.51904
Czech	0.44869	0.43867	0.44678	0.45027	0.44514	0.44318
English	0.42470	0.41071	0.39425	0.39844	0.41544	0.41382
French	0.47291	0.41608	0.44536	0.46911	0.45765	0.45388
Greek	0.25339	0.20009	0.20507	0.22718	0.29537	0.15405
Hebrew	0.25719	0.24113	0.21421	0.29113	0.25934	0.26475
Hindi	0.07933	0.03997	0.08794	0.04426	0.07704	0.04999
Romanian	0.42776	0.41170	0.39564	0.41449	0.40101	0.41623
Spanish	0.52388	0.47904	0.46444	0.50906	0.46055	0.46548

Table 5 shows that almost all results are better than the corresponding ones in table 3 when we compare the same feature. This is a good evidence of that stop words can reduce the Rouge score of summarization system. In these features, we can find out that the Levels feature can benefit most from removing stop words especially for Chinese summarization. When we use only one feature, Level Distribution is top 1 in Chinese summarization. For other languages, the result of Level Distribution is not that bad like that of table 3. That’s to say, the effect of Level Distribution are strongly related to whether we remove stop words or not.

Finally, we compare our best experimental result with several other summarization methods. There are 9 systems in MultiLing2015 MMS task involving a human summatization (This summarization is written by human). Not all systems have participated in all languages. Table 6 reveals the comparison result. In table 6, we represent a system not participating the language by the symbol "--".

Through the comparison in table 6, we can find out that our system has absolute advantage in Chinese summarization. The results are very similar to the human results. In Hebrew, the result of our system is also just second to the result of human summarization. However, in Arabic,Greek, Hindi, the performance of our system is extremely bad. Especially for Hindi, our results are far less than others. We need more experiments later to explore the reasons for this situation. But on the whole, our system is still competitive. The performances on Chinese, Czech and Hebrew are all just ranked behind human results. One thing we need to emphasize is that our best Chinese summarization results have used Level Distribution feature. Although the best summarization results of other languages is not always involving Level Distribution, it can indicate that Level Distribution feature is most helpful in Chinese summarization. In light of our experience on single document text summarization with Level Distribution, the parameter setting in this paper awaits to be improved in the future. Generally speaking, the proportion of other features and Level Distribution feature is at least 3:1. It maybe because this reason that our Level Distribution feature cannot improve the best summarization results of other languages.

Table 6. All system results(every cell in table have three line data, the first line is the result of ROUGE-1 F, the second line is the result of ROUGE-2 F, the last line is the result of ROUGE-SU4 F)

	cist	esi	giau	human	mms3	muse	occams	poly	wbu	Ours
Arabic	0.45918 0.13729 0.19228	0.48999 0.18379 0.23314	0.50778 0.19055 0.23987	0.69513 0.49773 0.51931	0.49923 0.19645 0.24189	0.50861 0.21669 0.25620	0.52084 0.21181 0.25615	0.52959 0.19408 0.23778	0.53161 0.23783 0.27340	0.28914 0.12625 0.14151
Chinese	0.25456 0.04935 0.08051	0.36556 0.13445 0.15410	--	0.70972 0.55843 0.55575	0.42077 0.20097 0.20824	--	0.29237 0.06799 0.09382	--	0.36919 0.15947 0.16971	**0.62952** **0.44720** **0.46876**
Czech	0.40045 0.12686 0.15917	0.41095 0.14192 0.17150	0.45677 0.19723 0.21365	0.66770 0.46779 0.48111	0.44405 0.17239 0.19291	--	0.47445 0.18494 0.20427	--	0.46226 0.20409 0.22001	**0.48183** **0.20737** **0.23114**
English	0.42646 0.10529 0.16669	0.47204 0.16764 0.21223	0.45919 0.14924 0.20676	0.67636 0.46165 0.49444	0.47266 0.16849 0.21526	0.48578 0.18175 0.23075	0.50270 0.17961 0.22746	0.47144 0.13919 0.19324	0.48978 0.19215 0.23035	0.45734 0.16773 0.20778
French	0.43953 0.12148 0.18731	0.51377 0.19442 0.24795	0.48887 0.17381 0.23468	0.71406 0.49190 0.52968	0.53222 0.21797 0.27466	--	0.53814 0.20702 0.26068	--	0.55427 0.24840 0.30160	0.50024 0.19164 0.19164
Greek	0.44377 0.12060 0.18938	0.44595 0.12624 0.19524	0.47945 0.17550 0.23715	0.66741 0.44346 0.48424	0.46582 0.14829 0.21342	--	0.48539 0.15989 0.21849	--	0.48237 0.16623 0.22627	0.29537 0.07758 0.11127
Hebrew	0.20209 0.04968 0.05963	0.25679 0.09124 0.09395	0.26528 0.09653 0.09807	0.50295 0.37887 0.38633	0.25297 0.08192 0.08533	0.26662 0.08359 0.08903	0.27524 0.08099 0.08595	0.24657 0.07432 0.07760	0.26755 0.09337 0.09723	**0.30942** 0.08141 **0.11863**
Hindi	0.64779 0.24051 0.37544	0.44918 0.18841 0.27516	0.64819 0.25874 0.37937	0.82748 0.62603 0.67671	0.66257 0.27152 0.40093	--	0.67586 0.25877 0.38042	--	0.66408 0.28622 0.40507	0.11090 0.00838 0.03848
Roma- nian	0.40667 0.09476 0.14593	0.42843 0.12710 0.17294	--	0.81966 0.69859 0.70505	0.44079 0.14403 0.18482	--	0.47011 0.16013 0.19549	--	0.47630 0.18904 0.21783	0.46142 0.18256 0.21317
Spanish	0.41554 0.13581 0.19776	0.53046 0.22059 0.28798	--	0.72331 0.52141 0.55985	0.55622 0.24963 0.31482	--	0.56984 0.24440 0.30120	--	0.57402 0.28102 0.33020	0.53113 0.24323 0.27850

7 Conclusion

This paper introduces an improvement of multilingual multi-document summarization system based on Multi2015 MMS data. We propose a new feature of Level Distribution, combine it with multiple other features and demonstrate the superiority of this new summarization system through experiments in some aspects. Specifically, Chinese summarization results have been improved greatly. Experimental results have also revealed that our new feature is valuable in multilingual multi-document summarization system. In the future, we still need more experiments and research to improve the application of Level Distribution feature. We will perform more experiments to verify the robustness and effectiveness of

Level Distribution feature in more languages. More features will be introduced into our system. For level distribution law, we will explore more rational and scientific explanations applications.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant 91546121, 61202247, 71231002 and 61472046; EU FP7 IRSES MobileCloud Project (Grant No. 612212); the 111 Project of China under Grant B08004; Engineering Research Center of Information Networks, Ministry of Education; Beijing Institute of Science and Technology Information; CapInfo Company Limited.

Reference

1. Ferreira,R. et.al. A Four Dimension Graph Model for Automatic Text Summarization[C]. Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 17-20 Nov. 2013: Pages(389 – 396).
2. Kalyani Bhagat, M.D.Ingle. Multi Document summarization using EM Clustering. IOSR Journal of Engineering (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719 Vol. 04, Issue 05 (May. 2014), ||V6|| PP 45-50.
3. Marina Litvak, Natalia Vanetik. Multilingual Multi-Document Summarization with POLY2[C]. Proceedings of the MultiLing 2013 Workshop on Multilingual Multi-document Summarization, pages 45–49, 2013.
4. Asli Celikyilmaz. Dilek Hakkani-Tur. A hybrid hierarchical model for multi-document summarization[C]. 48th Annual Meeting of the Association for Computational Linguistics, 815–824, Uppsala, Sweden, 2010
5. Blei D. M., Griffiths T. L., Jordan M. I., Tenenbaum J B. Hierarchical topic models and the nested Chinese restaurant process. Advances in Neural Information Processing Systems 16.Cambridge MA: MIT Press 2004.
6. Liu Pingan. Chinese multi document summarization based on HLDA model [D]. Beijing University of Posts and Telecommunications, 2013
7. Liu Hongyan. Multi document summarization based on hLDA hierarchical topic model [D]. Beijing University of Posts and Telecommunications, 2012
8. Liu Yu. Multi document summarization based on topic model and semantic analysis [D]. Beijing University of Posts and Telecommunications, 2015
9. Heng Wei, Yu Jia, Li Lei, Liu Yongbin. Reasearch on key factors of multi document topic modeling using hLDA [J]. Chinese Journal of information science and Technology (06) 2013
10. G.Giannakopoulos. 2015. MMS MultiLing2015 Task. http://multiling.iit.demokritos.gr/pages/view/1540/task-mms-multi-documentsummarization-data-and-information.[Online; accessed 19-July-2015].
11. Liu, Maofu, L. Wang, and L. Nie. Weibo-Oriented Chinese News Summarization via Multi-feature Combination. Natural Language Processing and Chinese Computing. Springer International Publishing, 2015:581-589.
12. Lin,Chin-Yew.2004. ROUGE: a Package for Automatic Evaluation of Summaries. In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), Barcelona,Spain, July 25 - 26, 2004.
13. http://www.berouge.com/Pages/default.aspx