Journal of Computers Vol. 31 No. 1, 2020, pp. 232-241
doi:10.3966/199115992020023101021

Generate Football News from Live Webcast Scripts Based on Character-CNN with Five Strokes

Xue-Qiang Lv[1], Xin-Dong You[1*], Wen-Chao Wang[1], Jian-She Zhou[2]

[1] Beijing Key Laboratory of Internet Culture and Digital Dissemination Research,
Beijing Information Science & Technology University, Beijing, China
lxq@bistu.edu.cn, youxindong@bistu.edu.cn, 925203515@qq.com

[2] Beijing Advanced Innovation Center for Imaging Technology, Beijing, China
zhoujianshe@cnu.edu.cn

Received 13 January 2019; Revised 12 March 2019; Accepted 12 March 2019

Abstract. Generating the football news from live webcast scripts is one kind of automatic text summarization in the field of natural language processing (NLP). Two typical characteristics in the live webcast scripts make the traditional methods can't well-generate the football news. Short sentences in live webcast scripts make it difficult to identify all of the keywords. And the sentence category identification is more crucial than the other kinds of automation text summarization. In this paper, the custom dictionary, forward dictionary and reverse dictionary in the football news domain are constructed, which aim to identify all import information from the live webcast. And the character-CNN with five strokes model is designed to classify the sentences category. Finally, the scanning algorithm based on time window is proposed to select the sentences to assemble into football news. Experiments results show that our designed Character-CNN model achieved higher accuracy rate on sentence classification than the state-of-art models. Evaluation conducted on ROUGE-1.5.5 toolkit showed that generating the football news with our proposed scanning algorithm outperforms than some baseline methods.

Keywords: CNN, Live Sports Text, scanning algorithm, Sports News

1 Introduction

Automatic text summarization is one of the standard tasks in the field of natural language processing. Starting from [1], research in automatic summarization has made great progress in many different application domains, which target to solve the different problems. Therefore, several classes of the methods have been explored in automatic text summarization field. Text Analytics Conference (TAC) and Document Understanding Conferences (DUC) have provided the main thrust of research in this area by creating standardized data and evaluation methods.

A large number of recent summarization methods are extractive, which means they select the most relevant sentences from the original document and aggregate them in certain order and then produce the summarization.

Extractive summarization methods assign sentences scores according to their importance firstly. Most early document summarization methods used the word distribution statistics to find the most relevant words, and then pick sentences that contain these words [1-2]. In later work, more sophisticated methods were explored, including using external knowledge [3], supervised machine learning based methods [4], methods based on discourse properties of input text [5, 14] and network based methods [6].

In the field of football news automatic writing, some excellent methods have emerged in recent years. Such as, Rules and templates based method [7], whose results are accurate and stable, but the text is lack of change and the academic value is low. Key sentence selection algorithm based on Restricted

* Corresponding Author

Boltzmann Machines (RBM) [8], and key sentence selection algorithm based on CRF. The above two algorithms select sentences directly from live text, do not deal with redundancy. Key sentence selection algorithm based on Fuse Random Forest (RF) [13] and Determinantal Point Processes (DPP) redundant algorithm for automatic writing [9]. This method well deals with redundant questions, but can't choose short live sentences containing key information, which will be lack of key information.

As a whole, there are many feasible methods to extract key sentences from live text, but only the RF+DPP method removes the live text in advance, and the other methods are very good to preprocess the live text. Due to long sentences, short sentences, redundant sentences and noisy sentences in live text, the final result of writing is less than satisfactory.

We proposed an improved method in this paper, in which the short text classification model based on the character level CNN is used to classify the sentences appearing in the live text, and a time window scanning algorithm is designed to attach the short sentences to the nearby long sentences with the classification information, or to make up a long sentence, at the same time, the noise sentences are deleted. Finally, all live text is standardized as a live sentence with similar structure and complete information. Finally, common sentence selection algorithm is used for keywords selection.

As a whole, in order to generate the well-organized football news from the live webcast scripts, main contributions of this paper are listed as below:

(1) The custom dictionary, forward dictionary and the reverse dictionary are constructed to identify the key information in the live webcast scripts as more as possible.

(2) Character-CNN deep learning model with five strokes method is designed in this paper to classify the sentence category with higher accuracy rate.

(3) Based on the classification results of sentence category, a scanning algorithm based on time window is proposed in this paper to select the key sentence to generate the football news with our customized templates.

(4) Experiments are conduct conducted in this paper to verify the advantage of our designed Character-CNN with five strokes method. And the effectiveness of the scanning algorithm to generate the football news is also evaluated in this paper.

2 Data Collection

Live broadcast service is very popular in China. We collect both the live broadcast text and the news text describes the same sports games. Table 1 displays an example of the format of the live texts, containing the main commentary text along with information, current timeline and the score.

Table 1. Ilustration of the live text format

sentence	time	score
梅西助跑，左脚直接打门！！！ (Messi run, the left foot hits the door straight!!!)	上半场 27' (first half 27')	0-0
皮克分球给后场右侧罗贝托 (Peake gives the ball to the back to the right Roberto)	上半场 31' (first half 31')	1-0
罗贝托一脚直塞球给禁区右侧苏亚雷斯！ (Robertot foot straight ball to the right side of the penalty area Suarez!)	上半场 45' (first half 45')	1-0

For every match, two different corresponding sports news reports are collected from Sina Sports Live and 163 Football Matches Live. These news reports are manually written by professional editors. The average number of sentences in the live texts for one match is around 242, containing around 4,590 Chinese characters for that match. The gold-standard news reports contain 1,185 Chinese characters on average, and forming around 32 sentences.

3 Procedures of Generating Football News

In this paper, we focus on Limited Style automatic writing, which is also considered as a special case of single document summarization. The syntactic structures of football news are regular usually: the first and last paragraph introduces the time, address of the game, and historical records of corresponding teams. When coming to body paragraphs, all events are presented base on the timeline. The first paragraph and the last paragraph just use simple method based on rules and template. The most important step is to embed the key sentences from live webcast scripts to generate the body paragraphs.

Current existing work mainly study how to select key sentence from live webcast, but there are redundant information in the live text, and many short sentences containing key information. This paper mainly studies how to deal with live text, and splicing the short sentences containing key information into a long sentence. In this way, the classic sentence selection algorithm can be used to generate the War newspaper. Finally, the effect of automatic writing of football news is improved, and some key information in the writing result can be avoided.

Through reading abundant sports news published by web portals, we have observed the key points in the football sports news are like as: goal in, wonderful shot, flagrant foul, substitution, extra time, penalty. Moreover, High quantity news should contain some information about pass and interception ball when describing the first three key points.

How to connect short sentences into long sentences is also a challenging job. First, we should identify the categories of each sentence accurately, identify the redundant information, and then design a time window scanning algorithm to determine what kind of short sentences should be combined with the adjacent sentences, and what kind of short sentences should be spliced with the adjacent long sentences.

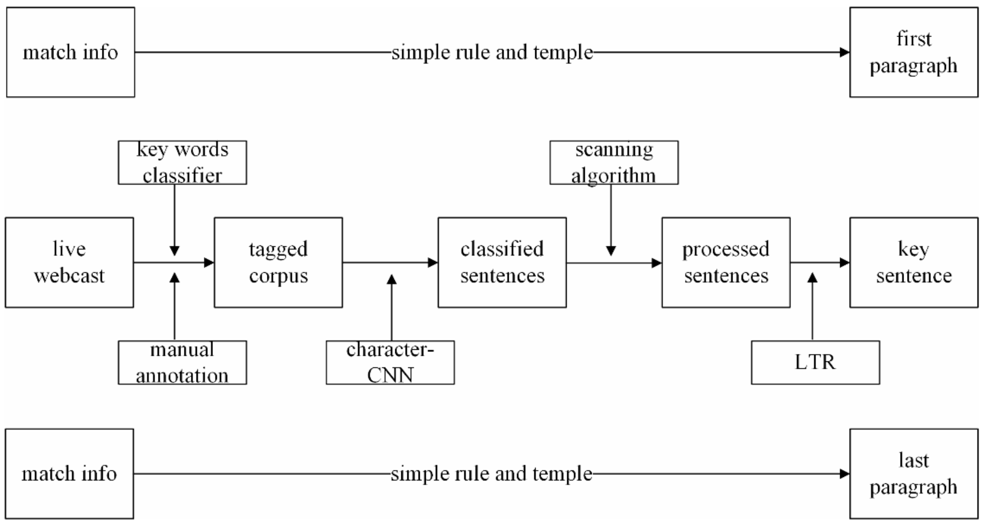

Fig. 1. Flow chart of football news generation

4 Sentence Classifier Based on Keywords

Firstly, all corpuses are processed with word segmentation mentioned before. The frequencies of all words are analyzed by the python code, it is easy to see that some punctuation marks and auxiliary verbs occur in high frequency. Since they are usually meaningless, these words will be removed from the vocabulary. Furthermore, most of low frequency words are member names, places, teams, and other inactive words. And most of middle frequency words are terminology words in football, which well-represent the character of whole sentence and determine the content of this sentence. Thus, these middle frequency words are collected in the vocabulary list and manually divided into six classes, including Substitution, Goal, Foul, Shoot, Dribble, Punish. But it cannot well recognize the type of each sentence only depend on these forward direction key words. Consequently, a reverse direction vocabulary is generated for each class, which improved perfectly as tests going. The following Table 2 shows part of our vocabulary.

Table 2. Part of our vocabulary

class	forward direction	reverse direction
Substitution	换下，换人，换上...	没有，准备...
Goal	打进，破门，入网 ...	进球啦，的进球 ...
Foul	踢翻，掀翻，铲翻 ...	不好意思，看错了…
Shoot	打门，推射，弹射 ...	准备打门...
Dribble	给到，再给，分球 ...	再给一张...
Punish	黄牌，红牌，吹罚 ...	应该...

5 Character level CNN Text Classifier with Five Strokes

Recent research showed that CNN can solve not only computer vision, but also NLP problem. Chinese characters are hieroglyphs, whose components combine in its own set of deep logic. After modifying the input from pinyin to five-strokes, the character-level CNN can learn more about this kind of character. The corresponding equation is shown as: among them, this unit is also referred as the logistic regression model. The neural network model is produced when multiple cells combined with hierarchical structure. The corresponding equation is shown as:

$$h_{W,b}(x) = f(W^T x) = f(\sum_{i=1}^{3} W_i x_i + b). \tag{1}$$

Among them, this unit is also referred as the logistic regression model. The neural network model is produced when multiple cells combined with hierarchical structure.

The corresponding equation is shown as:

$$a_1^{(2)} = f(W_{11}^{(2)} x_1 + W_{12}^{(1)} x_2 + W_{13}^{(2)} x_3 + b_1^{(1)}). \tag{2}$$

$$a_2^{(2)} = f(W_{21}^{(2)} x_1 + W_{22}^{(1)} x_2 + W_{23}^{(2)} x_3 + b_2^{(1)}). \tag{3}$$

$$a_3^{(2)} = f(W_{31}^{(2)} x_1 + W_{32}^{(1)} x_2 + W_{33}^{(2)} x_3 + b_3^{(1)}). \tag{4}$$

$$h_{W,b(x)} = a_1^{(3)} = f(W_{11}^{(2)} a_1^{(2)} + W_{12}^{(2)} a_2^{(2)} + W_{13}^{(2)} a_3^{(2)} + b_1^{(2)}). \tag{5}$$

The training method of neural network is similar to logistic, but due to its multiple layers, the chain derivative method is used for derivation of hidden layer nodes, which is also called back propagation. This paper does not enclose the training algorithm.

The structure of our model is shown in Fig. 2.

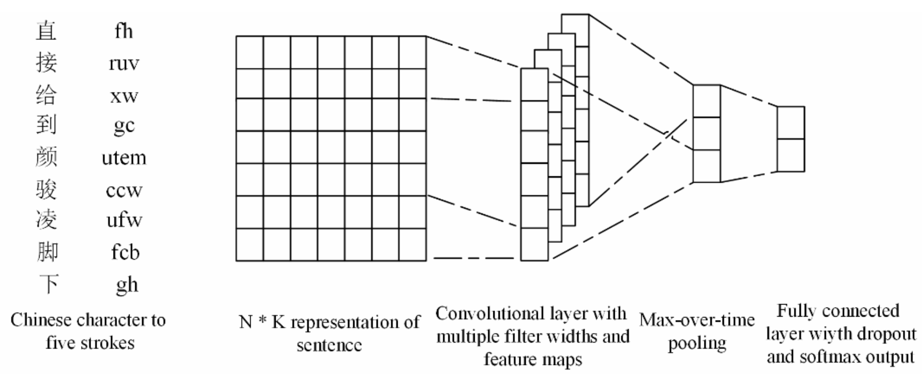

Fig. 2. Structure of our Character-CNN model

As the figure shows, the input level is corpus, whose characters are encoded to matrix in five strokes (from top to bottom). Assume there are n words in a sentence, and the dimension of vector is k, then the matric is n * k.

There are three layers on our Character-CNN model:

First layer is convolutional layer, in which Several Feature Maps are obtained by input level convolution, and the size of convolution window is h * k, which represents the number of vertical words, k represents the dimension of words. Several Feature Maps with one column are obtained through such a large convolution window.

Second layer is pooling layer, in which the method called Max-over-time pooling is used in pooling layer, which simply extract the max value (represent the most significant signal) from the previous layer of Feature Map. This pooling method can solve the problem of convertible length of sentence. Eventually, the outputs of pooling layer are the max values of each Feature Map, which is a one dimensional vector.

Third layer is all connection + Soft-max layer, which is the output of pooling layer, in which a one dimensional vector is connected to Soft-max layer with all connections. Soft-max layer is set according to the needs of the tasks.

Dropout technology is used for the penultimate part of all connection, which means weight the parameter of all layer connection with L2 regularization restriction. The advantage is to prevent the hidden layer unit adaptive, thus reducing the degree of fitting.

6 Core Sentence Splicing

We described the classification of short text by using the character-CNN model. After classifying the category of each short sentence in the live text, we designed a number of short text spliced into a battlefield algorithm. Finally, according to the time sequence of sentences like line, together. A complete report is formed. For example, the type of dealing with a short sentence is a goal. As we all know, goals are a very important message for readers. We should not only reflect in the reports of this goal event, but also the goal event antecedents and consequences. Therefore, a scanning algorithm based on time window is proposed in this paper.

We define that sentence with more than or equal 30 Chinese characters is a long core sentence. Sentence with less than 30 Chinese characters are defined as short sentences. If the category of a short sentence is Goal or Punish or Shoot, the sentence is defined as the short core sentence, and the sentence outside the basic category is defined as the noise sentence. Our goal is to match short sentences which are close to the core sentence. Xiaobian usually introduces substitution information in one sentence, so Substitution type sentences do not splice any short sentences before and after. It is found that noise sentences usually occur in the absence of important events on the court, so it is possible to use the Substitution sentence and the noise sentence to divide the live text into different intervals. Sentences in different intervals do not perform stitching. The proposed time window based scanning algorithm is described in Algorithm 1.

Algorithm 1. Time window based scanning algorithm

```
Input：webscript live text
Output:  joint core senetence
Begin
   for sentences_happened_in_same_minute in live_webcast
        joint_core_sentence();
   End for
   for sentences_happened_in_same_minute in updated_live_webcast
      for sentence in sentences_happened_in_same_minute:
        if  sentence is not core_sentence
          if next_sentence_is_core_sentence and
             previous_sentence_is_core_sentence
            chose_shorter_core_sentence_to_joint();
          End if
        else if next_sentence_is_core_sentence
              joint_with_next_sentence();
```

```
        End else
        else if previous_sentence_is_core_sentence
             joint_with_previous_sentence();
        End else
      End if
     End for
   End for
   for core_sentence_less_than_50_characters in core_sentences
         split_core_sentence_by_punctuation();
         first_class,last_class =
     get_classname_of_first_short_sentence_and_last_short_sentence();
      for sentence in sentence_in_front_of_the_core_sentence
          if class_of_sentence != first_class
             put_this_sententc_before_core_sentence()
          End if
      End for
      for sentence in sentence_after_of_the_core_sentence
         if class_of_sentence != first_class
              put_this_sententc_after_core_sentence();
         End if
       End for
   End for
   joint_last_short_sentences()'
End
```

And the sketch map of core sentence splicing is demonstrated in Fig. 3.

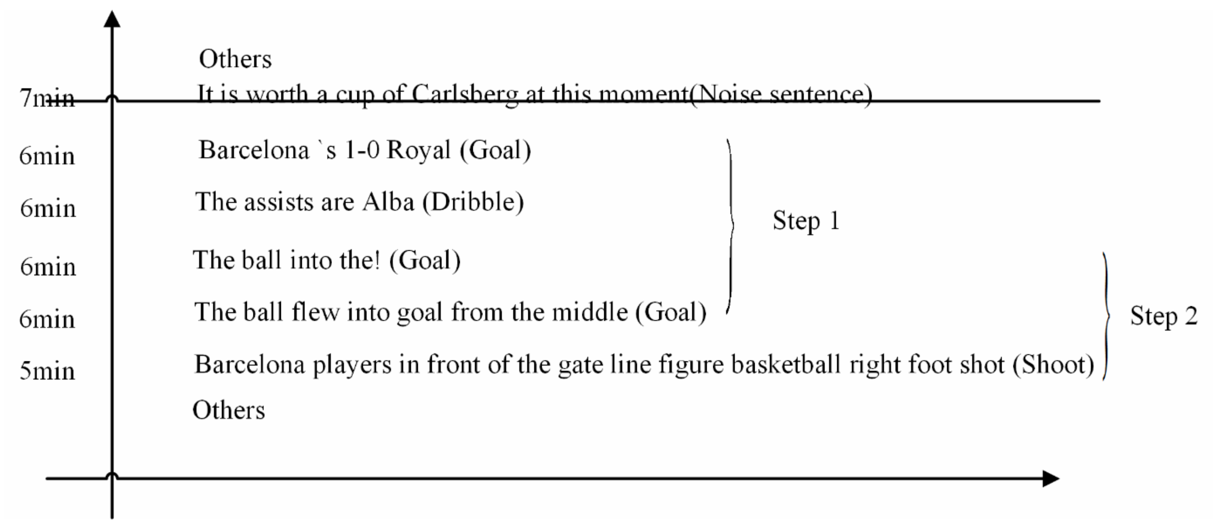

Fig. 3 Sketch map of core sentence splicing

7 Experimental and the Results

The experiment data chose the sport news with manually annotation which has introduced by the second chapter. The training data used in this experiment are 900 live texts that crawled from web, the test data are the 30 sample files.

Our method is compared with several baselines, which are typically traditional summarization approaches:

HeadTail: Using head and tail sentences only. This baseline resembles the baseline of leading sentences for traditional summarization.

Centroid: In centroid-based summarization [10].

LexRank: LexRank [6] computes the importance of sentence based on the concept of eigenvector centrality in a graph representation of sentences.

ILP: Integer linear programming (ILP) approaches [11] cast document summarization as combinatorial optimization.

RF+DPP: it is a probabilistic approach based on determinantal point processes [12]. This approach can naturally integrate the predicted scores from the LTR model while trying to avoid certain redundancy by producing more diverse extractions.

For fair comparisons the length of each constructed news report is limited to be no more than 1,000 Chinese characters, roughly the same with the average length of the gold-standard news.

The task mainly focuses on evaluating document summarization techniques for producing Chinese sports news articles from live webcast scripts. Due to the analytical characteristics of input data set and output news articles. We found that it is crucial to precisely classify the type of each sentence of the live webcast. Therefore, we conduct the experiment on the sentence classification firstly.

We conducted a series of experiments with both traditional and deep learning methods. We tried our best to choose models that can provide comparable and competitive results, and the results are as follow Table 3.

Table 3. The experiment results among different models

Precision model	Dribble	Foul	Goals	Punish	Shooting	Substitution
Keywords	0.73	0.74	0.75	0.74	0.84	0.82
SVM	0.92	0.88	0.84	0.86	0.89	0.88
BOW	0.91	0.87	0.86	0.85	0.90	0.86
n-grams	0.88	0.86	0.83	0.86	0.91	0.87
CNN	0.93	0.90	0.91	0.88	0.91	0.90
Char-CNN-Chinese	0.86	0.75	0.83	0.82	0.85	0.84
Char-CNN-pinyin	0.92	0.91	0.90	0.89	0.91	0.89
Char-CNN-five-stoke	0.95	0.91	0.89	0.90	0.92	0.89

The effect of the keyword method based classification is obviously worst, but the original intention of designing this method is to assist us in tagging the corpus. When dealing with Chinese, we have to split the text into words first. Although we collected all the players ahead of the team's name made a custom dictionary, the results of SVM and BOW and N-Grams and CNN method are closely. Before using the word segmentation algorithm, the segmentation effect has significantly increased. But from the final classification results, the result is still less than satisfactory. One possible reason is that simply segmenting Chinese text, and then processing the corresponding vectors of these words, cannot capture deep information between words and words. The other reason may be the live text contains a large number of short sentences, and a small number of unrelated interference information.

One of the most obvious facts observed from Table 3 is that the Char-CNN model performs best almost in every case. Compared with traditional models, character-level Conv-Nets could work for text classification without the need for words. This is a strong indication that language could also be thought of as a signal no different from any other kind. Conv-Nets may work well for user-generated data. User-generated data vary in the degree of how well the texts are curated. For example, in datasets, Sport live texts tend to be raw user inputs.

When Char-CNN is dealing with Chinese, we can have three different ways. The first is to enter the Chinese character directly. The second is to translate Chinese characters into phonetic alphabet, and then deal with them as English. The last one is to transform each Chinese character into the corresponding five strokes encoding, and it can also be processed in English. Experiments show that the classification effect of Chinese character input is the worst. After converting Chinese characters to Pinyin, the effect has been significantly improved. This may be due to the fact that the total number of Chinese characters is hundreds of times the total number of English characters. The model has to face huge sparse when dealing with Chinese directly. When the five-stroke code is used, the effect is slightly improved than the phonetic alphabet. It may be that the Chinese character is a pictograph, and the five strokes code is more closely connected with the Chinese character. The Char-CNN model can capture more information.

After classifying the category of each short sentence in the live text, the sentence selection model is introduced in the fifth section to produce a smooth sentence. Then we can easily splice them into football news.

ROUGE metrics—Recall and F-scores in ROUGE-1 ROUGE-2 and ROUGE-SU4 is utilized to evaluate to our proposed algorithm comprehensively. The average performance of different methods is displayed in Table 4.

Table 4. Comparison results of different methods

Method	ROUGE-1	ROUGE-2	ROUGE-SU4
HeadTail	0.30147	0.07779	0.10336
Centroid	0.32508	0.08113	0.11245
LexRank	0.31284	0.06159	0.09376
ILP	0.32552	0.07285	0.10378
RF+DPP	0.39391	0.11986	0.15097
Our	0.39409	0.11991	0.15207

The ILP model, which is believed to be suitable for multi-document summarization, did not perform well in our settings. Head and tail sentences are informative but merely using their lacks specific descriptions for procedural events, therefore not providing competitive results either.

The comparison shows the effectiveness of our sentence selection strategy. However, the increase is very close to RF+DPP.

As seen in Table 4, the effect of HeadTail or LexRank and other common text summarization algorithm is not so good in those models. Each sentence in the corpus is considered as an independent individual, which can't effectively utilize the complex features. However, there is a certain correlation between sentences in the corpus. And the effect of RF+DPP is not so obvious also, although it can solve some problems.

Common text summarization algorithm will suffer from redundancy in commentary, since those scores are predicted independently for each sentence and e key event. Therefore, without special care in sentence selection, the result will be bad. RF+DPP tried to avoid certain redundancy by producing more diverse extractions. But we use information of types of sentences combine with other information, called scanning algorithm based on time window to avoid this situation. Thus RF+DPP and method described by this paper obtain a better result.

The above comparing experiments showed that the proposed method, which based on character level CNN model along with time window based scanning algorithm, achieved a good result on the summary sentences extraction in the field of live sport text.

Because RF+DPP doesn't have high value on short sentences, the final writing results may miss some important information. The method described in this paper splits the important short sentences to avoid this phenomenon. The football news generated by RF+DPP method and our method are shown in Table 5, along with the news written by Editors, which verify the effectiveness of our proposed algorithm. The most important information has been identified, and the sentences in the news are smooth and readable.

Table 5. Generation result contrast

Version	Sentence
RF+DPP	开场 6 分钟, 巴塞罗那球员图兰门前前右脚射门，球从中路飞进球门，巴塞罗那 1-0 皇家贝蒂斯。 In the opening 6 minutes, Barcelona players shot the front right foot, and the ball flew from the middle to the goal, and the 1-0 Royal Bettis of Barcelona.
Our algorithm	开场 6 分钟，巴塞罗那球员图兰球门线跟前右脚射门，球从中路飞进球门球进了！助攻的是阿尔巴，巴塞罗那 1-0 皇家贝蒂斯。 In the opening 6 minutes, Barcelona players in front of the gate line figure basketball right foot shot, the ball flew into the goal from the middle, the ball into the! The assists are Alba, Barcelona's 1-0 Royal Bettis.
Editors	巴萨开场 6 分钟率先破门，梅西右路转移，阿尔巴禁区左侧回传，图兰在门前 6 米处推射，中卫曼迪，门线解围，但还是无法阻止球入网，1 比 0。 Barcelona opened the first 6 minutes to break the door, Messi right transfer, Alba restricted area left back, Tuan in front of the 6 meters in the door push, Mandy, the gate line, but still can not stop the ball into the net, 1 to 0.

8 Conclusion and Future Work

Automatically generated Football News is one of refined field in document summary and natural language generation. The biggest challenge in this field is how to extract all the key information at the maximum extent. When humans writing news, the first step is to grasp the overall context of the whole game, then locking key information base on their experience and intuition, after that they are able to organize an overview of all key information in appropriate form, appear a wonderful game of football news. The system can't identify important information point game intuitively as a professional editor at the present stage. In this paper, we build a machine learning-based algorithm to learn the relationship and difference between critical and non-critical information through reading the manual tagging corpus, and then try to identify the key sentences from all inputs. And the char-CNN with five strokes model is designed in this paper, which can enhance the accuracy rate of sentences category, and receive a better performance in generating football news in result.

During identifying key words, the human is not simply based on a sentence, but the comprehensive context, and even the whole football game. Therefore, we believe that if we can make the system understand the ins and outs of the whole game, we will be able to establish appropriate model then make the system lock key information from the global perspective. Therefore, our future work is aiming to design a more advanced key words recognition system and manually tagging more corpuses, which can identify category of all phrases from the live webcast. Moreover, a recognition model based on the time line from the beginning to the end of the game will be established, with which the progress of the whole game will be understand, and then our system will has the ability to grasp the overall situation like the human editors.

Acknowledgements

This work is supported by National Natural Science Foundation of China under Grants No. 61671070, National Science Key Lab Fund project 6142006190301, National Language Committee of China under Grants ZDI135-53, and Project of Three Dimension Energy Consumption Saving Strategies in Cloud Storage System in Promoting the Developing University Intension–Disciplinary Cluster No. 5211910940, Qin Xin Talents Cultivation Program, Beijing Information Science & Technology University No. QXTCP B201908.

References

[1] H.P. Luhn, The automatic creation of literature abstracts, IBM Journal of Research and Development 2(2)(1958) 159-165.

[2] P.B. Baxendale, Machine-made index for technical literature: an experiment, IBM Journal of Research and Development 2(4)(1958) 354-361.

[3] R. Barzilay, M. Elhadad, Using lexical chains for text summarization, in: Proc. 1997 the ACL Workshop on Intelligent Scalable Text Summarization, 1997.

[4] J. Kupiec, A trainable document summarizer, in: Proc. 1995 the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1995.

[5] I. Mani, D. House, G. Klein, The tipster summac text summarization evaluation, in: Proc. 1999 Conference on European Chapter of the Association for Computational Linguistics, 1999.

[6] G. Erkan, D.R. Radev, LexRank: graph-based lexical centrality as salience in text summarization, Journal of Artificial Intelligence Research 2(1)(2004) 457-479.

[7] M.-F. Liu, Q.-S. Qi, H.-J. Hu, Sports news generation from live webcast scripts based on rules and templates, in: Proc. 2016 International Conference on Computer Processing of Oriental Languages National CCF Conference on Natural Language Processing and Chinese Computing, 2016.

[8] R.-J. Tang, K. Zhang, Football news generation from Chinese live webcast script, in: Proc. 2016 Fifth Conference on Natural Language Processing and Chinese Computing & The Twenty Fourth International Conference on Computer Processing of Oriental Languages, 2016.

[9] J.-M. Zhang, J.-G. Yao, X.-J. Wan, Towards constructing sports news from live text commentary, in: Proc. 2016 Meeting of the Association for Computational Linguistics, 2016.

[10] D.R. Radev, H.-Y. Jing, M. Budzikowska, Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies, Information Processing & Management 40(6)(2004) 919-938.

[11] D. Gillick, B. Favre, The ICSI summarization system at TAC 2008, in: Proc. 2008 the Text Analysis Conference Workshop, 2008.

[12] A. Kulesza, B. Taskar, Determinantal point processes for machine learning, Foundations and Trends in Machine Learning 5(2-3)(2012) 123-286.

[13] L. Breiman, Random forests, Machine Learning 45(1)(2001) 5-32.

[14] D. Marcu, G. Hirst, An implemented formalism for computing linguistic presuppositions and existential commitments, in: Proc. 1994 the International Workshop on Computational Semantics, 1994.