Overview of the NLPCC-ICCPOL 2016 Shared Task: Sports News Generation from Live Webcast Scripts

Xiaojun Wan, Jianmin Zhang, Jin-ge Yao and Tianming Wang

Institute of Computer Science and Technology, the MOE Key Laboratory of
Computational Linguistics, Peking University, Beijing 100871, China
wanxiaojun@pku.edu.cn

Abstract. Live webcast scripts are valuable resources for describing the process of sports games. This shared task aims to automatically generate sports news articles from live webcast scripts. The task can be considered a special case of single document summarization. In this overview paper, we will introduce the task, the evaluation dataset, the participating teams and the evaluation results. The dataset has been released publicly.

Keywords: Sports news generation; document summarization; shared task; NLPCC-ICCPOL 2016

1   Task

It is an urgent demand to write and publish a sports news article immediately after a sports game ends. Till now, sports news articles are usually written by human experts or journalists. How to automatically generate sports news has always been a very challenging problem.

Many Chinese sports-related web sites (e.g. Sina Sports[1], NetEase Sports[2]) provide live webcast services to users for watching sports games. One popular service is based on live text webcast, and a webcast host frequently writes one or several short sentences (i.e. scripts) to describe the latest progress of a game in real-time. Users will know what happened in the game immediately after reading the latest scripts. The live webcast scripts are valuable resources for sports news generation, which are, however, neglected in the area of sports news generation and summarization. Note that we can also obtain live text scripts from live TV or video by using ASR technologies.

In this shared task, we aim to explore the possibility of generating Chinese sports news from live webcast scripts. The task can be treated as a special document summarization or text-to-text generation task. A very recent work has conducted a pilot study on this task [1]. In this year's task, we focus on football games. We encourage participants to consider the characteristics of live webcast scripts and develop more competitive sports news generation systems.

---

[1] http://sports.sina.com.c
[2] http://sports.163.com/

2 Data

We provided sample (training) data and test data for this shared task. For sample data, we provided a set of live webcast scripts and the corresponding human-written news stories for 30 football games. The live web cast scripts and the news stories were crawled from Sina Sports and NetEase Sports. For each sports game, we provided at least two news stories as the reference news. For test data, we provided a set of live webcast scripts for 30 other football games.

Part of the live webcast scripts for a football game (Chelsea vs Sunderland on 2015-12-19) is shown in Table 1. In each row in the table, the text scripts are provided along with the match time and match score.

Table 1. Part of live webcast scripts for the game of Chelsea vs Sunderland on 2015-12-19

伊万右路45度斜传球到禁区，科茨解围球直接停给了佩德罗，佩德罗将球一拨，左脚抽射，打进！！！	上半场 14'	2-0
威廉前场右路拿球，强突后分边给上来的伊万	上半场 14'	2-0
伊万没机会，回传给到小法	上半场 15'	2-0
小法拿球内切，斜塞禁区找插入的威廉	上半场 15'	2-0
球被回防的范安霍尔特抢先出脚碰出底线，角球	上半场 15'	2-0

In addition to the live webcast scripts, we also provided match statistics and players statistics of the game, as illustrated in Tables 2 and 3. The participants can optionally make use of the match statistics in developing their sports news generation systems.

Table 2. Match statistics for the game of Chelsea vs Sunderland on 2015-12-19

切尔西	项目	桑德兰
17	总射门	11
7	射正球门	3
10	射门偏出	8
0	击中门框	0
3	直塞球	1
0	越位	3
17	抢断	27
12	任意球	12
12	犯规	12
2	角球	2
26	界外球	22
69	超过25码长传	54
84.80%	传球成功率	69.60%
41.20%	传中成功率	33.30%
70.60%	抢断成功率	66.70%
40%	头球成功率	60%
67.50%	控球率	32.50%

Table 3. Players statistics for the game of Chelsea vs Sunderland on 2015-12-19

号码	位置	球员名	出场	时间	进球	助攻	威胁球	射门	射正	射正率	犯规	被犯	扑救
13	门将	库尔图瓦	首发	90'	0	0	0	0	0	0	0	0	2
28	后卫	阿斯皮利奎塔	首发	90'	0	0	0	0	0	0	3	0	0
2	后卫	伊万诺维奇	首发	90'	1	0	0	2	1	50%	2	0	0
26	后卫	特里	首发	90'	0	0	0	0	0	0	0	1	0
5	后卫	祖马	首发	90'	0	0	0	0	0	0	1	1	0
22	中场	威廉	首发	90'	0	1	4	3	2	66.70%	0	2	0
8	中场	奥斯卡	首发	82'	1	0	2	5	2	40%	1	1	0
21	中场	马蒂奇	首发	90'	0	0	2	0	0	0	2	1	0
4	中场	法布雷加斯	首发	71'	0	0	1	0	0	0	2	1	0
17	中场	佩德罗	首发	90'	1	0	3	3	1	33.30%	0	4	0
19	前锋	迭戈-科斯塔	首发	76'	0	0	0	2	0	0%	1	0	0
1	替补	贝戈维奇	替补	0'	0	0	0	0	0	0	0	0	0
24	替补	加里-卡希尔	替补	0'	0	0	0	0	0	0	0	0	0
12	替补	米克尔	替补	19'	0	0	0	0	0	0	0	1	0
18	替补	雷米	替补	14'	0	0	0	2	1	50%	0	0	0
7	替补	拉米雷斯	替补	8'	0	0	0	0	0	0	0	1	0
14	替补	特劳雷	替补	0'	0	0	0	0	0	0	0	0	0
6	替补	拉赫曼-巴巴	替补	0'	0	0	0	0	0	0	0	0	0

3 Participants

Each team is allowed to submit at most two runs of results. The length of each sports news is limited to 1000 Chinese characters and longer news articles will be truncated. The participants are allowed to use any NLP resources or toolkits, but it is NOT allowed to crawl and use any news articles related to the given games on the Web.

There are 7 teams participating in this shared task and they submitted a total of 10 valid runs of results. The participating teams are shown in Table 4. Various techniques have been used by the participating teams and several teams have crawled additional live scripts and sports news articles for other sports games to enhance the training data. For example, IACAS_Human_HCI mainly relies on key sentence extraction by considering various factors, including keywords, time, and so on. The team also designs and uses a few patterns to improve the readability of generated sports news. ICDD_SportsNews uses different strategies for generating different parts of a sports news. The team generates the first paragraph of a sports news with patterns to briefly overview the sports game, and then applies supervised learning methods to extract key sentences to describe the process of the game, and finally provides statistics of players in the last paragraph. BIT Coder generates sports news with slot filling techniques.

Table 4. Participating teams.

Team ID	Organization Name
IACAS_Human_HCI	Chinese Academy of Sciences; Guilin University of Electronic Technology; Hebei University of Technology
ICDD_SportsNews	Beijing Information Science and Technology University
RDNH	Chongqing University of Technology
BIT Coder	Beijing Institute of Technology
CQUT_AC996	Chongqing University of Technology
CCNU2016NLP	Central China Normal University
BIT Hunter	Beijing Institute of Technology

4   Results

The generated sports news articles were evaluated both automatically and manually.

4.1   Automatic Evaluation

We used the ROUGE-1.5.5 toolkit [2] for automatic evaluation[3].In order to make the ROUGE toolkit work well for evaluating Chinese news, the code related to text encoding in ROUGE-1.5.5.pl should be modified. The recommended options for the toolkit were -c 95 -2 4 -U -r 1000 -n 4 -w 1.2 –a –l 1000. Note that the length of each sports news was limited to 1000 Chinese characters, so we used –l 1000 for truncating longer news articles.

Table 5. Automatic evaluation results

Team	Run	ROUGE-1		ROUGE-2		ROUGE-SU4	
		Recall	F-measure	Recall	F-measure	Recall	F-measure
IACAS_Human_HCI	run1	0.57782	0.59846	0.24998	0.26293	0.25464	0.26652
	run2	0.55643	0.60331	0.24448	0.26092	0.24777	0.26581
ICDD_SportsNews	run1	0.56515	0.59261	0.25235	0.26444	0.25404	0.26613
	run2	0.56768	0.59179	0.25059	0.26119	0.25438	0.26497
RDNH	run1	0.55235	0.5865	0.25527	0.27081	0.25333	0.26863
BIT_Coder	run1	0.49728	0.55851	0.22524	0.25333	0.22484	0.25263
CQUT_AC996	run2	0.5222	0.55728	0.22182	0.23688	0.22689	0.2422
CCNU2016NLP	run1	0.46105	0.52478	0.19486	0.22128	0.19322	0.21947
	run2	0.4948	0.52425	0.20894	0.22123	0.21102	0.22325
BIT_Hunter	run2	0.36532	0.47758	0.16072	0.2106	0.16504	0.21589

---

[3] The ROUGE toolkit can be downloaded from http://www.berouge.com.

Table 6. Manual evaluation results

Team	Run	Factor	Average Score
IACAS_Human_HCI	run1	Read.	3.84444
		Cont.	3.54444
		Overall	3.63333
	run2	Read.	3.88889
		Cont.	3.64444
		Overall	3.73333
ICDD_SportsNews	run1	Read.	3.34444
		Cont.	3.32222
		Overall	3.24444
	run2	Read.	3.46667
		Cont.	3.32222
		Overall	3.28889
RDNH	run1	Read.	2.92222
		Cont.	1.85556
		Overall	2.17778
BIT Coder	run1	Read.	2.55556
		Cont.	2.74444
		Overall	2.45556
CQUT_AC996	run2	Read.	2.61111
		Cont.	2.21111
		Overall	2.30000
CCNU2016NLP	run1	Read.	2.06667
		Cont.	2.02222
		Overall	1.93333
	run2	Read.	2.24444
		Cont.	2.67778
		Overall	2.21111
BIT Hunter	run2	Read.	1.68889
		Cont.	1.94444
		Overall	1.60000

The recommended ROUGE metrics are Recall and F-measure scores of Character-based ROUGE-1, ROUGE-2 and ROUGE-SU4. Character-based evaluation means that we do not need to perform Chinese word segmentation when running the ROUGE toolkit. Instead, we only need to separate each Chinese character by using a blank space.

Table 5 gives the automatic evaluation results. We can see that the three teams of IACAS_Human_HCI, ICDD_SportsNews and RDNH perform better than the other four teams.

4.2 Manual Evaluation

We further performed manual evaluation, which is more reliable than the automatic evaluation for this special task. Three graduate students who are fluent in Chinese were asked to perform manual ratings for each news article with respect to three factors: readability (Read.), content coverage (Cont.) and overall score. The ratings were in the range of 1-5, with higher scores denoting better quality. Finally the scores were averaged across the 30 news articles, and then averaged across the three judges.

The results are shown in Table 6. We can see that the team of IACAS_Human_HCI performs the best over all the three factors, and the ICDD_SportsNews team also performs very well over the three factors. The two teams considered various factors for key sentence extraction, and they used both machine learning and pattern-based techniques for sports news generation.

5   Conclusions

We proposed a new shared task of sports news generation from live webcast scripts for NLPCC-ICCPOL 2016, and 7 teams participated in this shared task. The evaluation dataset has been released publicly[4]. We expect more advanced text summarization and generation methods will be proposed for this special Chinese sports news generation task.

Acknowledgments

References

1. Zhang, Jianmin, Jin-ge Yao and Xiaojun Wan. "Toward Constructing Sports News from Live Text Commentary", *Proceedings of ACL*, 2016.
2. Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." *Text summarization branches out: Proceedings of the ACL-04 workshop*. Vol. 8. 2004.

---

[4] http://www.icst.pku.edu.cn/lcwm/wanxj/files/NLPCC2016Eval-Task5-AllData.rar