NLPCC-ICCPOL 2016 Shared Task Guideline:
Sports News Generation from Live Webcast Scripts

1. Task Description

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, NetEase Sports) 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 the game in real-time. Users will know what happened in the game immediately after reading the latest scripts. The live webcast scripts are useful resources for sports news generation, which are, however, neglected in the area of sports news generation and summarization.

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. 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.

2. Data

We will provide sample (training) data and test data later. For sample data, we will provide a set of live webcast scripts and the corresponding news stories. For each sports game, we will provide at least two news stories as the reference news. For test data, we will only provide a set of live webcast scripts.

Part of the live webcast scripts for a football game (Chelsea vs Sunderland on 2015-12-19) is shown in Table 1.

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

Table 1: Part of live webcast scripts for the game of Chelsea vs

Sunderland on 2015-12-19

In addition to the live webcast scripts, we will also provide 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.

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 2: Match 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

Table 3: Players statistics for the game of Chelsea vs Sunderland on

2015-12-19

However, the participants are NOT allowed to crawl and use any news articles related to the given games on the Web.

3. Evaluation

The generated sports news articles will be evaluated both automatically and manually.

For automatic evaluation, we will use the ROUGE toolkit. ROUGE-N F-measure will be used as evaluation metrics.

For manual evaluation, we will use the pyramid method for evaluating the content quality. We will also manually evaluate the readability and overall quality of the whole articles.