Generating Sports News from Live Commentary: A Chinese Dataset for Sports Game Summarization

Kuan-Hao Huang[1]   Chen Li[2]   Kai-Wei Chang[1]

[1]University of California, Los Angeles

[2]Tencent

AACL 2020
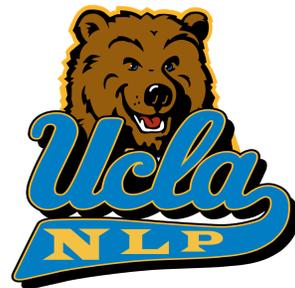
Motivation

- There are a lot of sports games playing everyday
- Manually writing sports news summaries is labor-intensive
- Automated sports game summarization

Butler's big night helps Heat cut Lakers' Finals lead to 2-1

Associated Press   Oct 4, 2020

LAKE BUENA VISTA, Fla. -- Biggest moment of his career. Best game of his life.

Jimmy Butler is clearly not ready to go home quite yet. With a triple-double, he joined NBA Finals lore -- and the short-handed Miami Heat might have made this title matchup a series after all.

Butler finished with 40 points, 11 rebounds and 13 assists, and the Heat beat the Los Angeles Lakers 115-104 on Sunday night to get within 2-1 -- doing so with starters Bam Adebayo and Goran Dragic still unable to play because of injury.

"I tell Coach all the time, ``I'm ready for this," Butler said. "The biggest stage, whatever you ask me to do, I can do."

sina 新浪竞技风暴   体育 > 国际足球-德国 > 正文

德甲-里贝里2助攻克罗斯传射 拜仁3-0汉堡领先7分

2012年11月04日03:22 新浪体育 微博   A⁻ A⁺
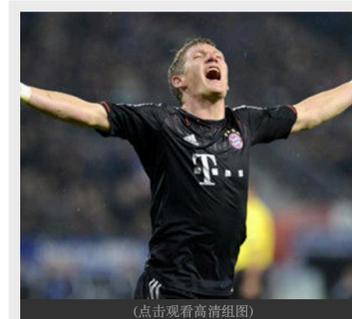
新浪体育讯　北京时间11月4日1时30分(德国当地时间3日18时30分)，德甲[微博]第10轮在英泰竞技场开始1场焦点战，拜仁[微博][微博]客场3比0完胜汉堡。克罗斯助攻施魏因斯泰格首开纪录。下半时，里贝里助攻穆勒和克罗斯破门。沙尔克客场2比3负于霍芬海姆，此前1天，法兰克福主场1比1平菲尔特，拜仁因而将领先优势扩大到7分。多特蒙德主场0比0平斯图加特；沃尔夫斯堡客场0比1负纽伦堡；汉诺威主场2比0胜奥格斯

(点击观看高清组图)

Generating Sports Summary from Live Commentary

- Live commentary
  - Play-by-play records
  - Transcripts of commentators
- Generate sports news from live commentary

TIME	TEAM	PLAY	SCORE
12:00		Dwight Howard vs. Meyers Leonard (LeBron James gains possession)	0 - 0
11:52		**Dwight Howard makes 1-foot alley oop dunk shot (LeBron James assists)**	**2 - 0**
11:28		**Tyler Herro makes two point shot (Duncan Robinson assists)**	**2 - 2**
11:09		Dwight Howard lost ball turnover (Duncan Robinson steals)	2 - 2
10:53		Tyler Herro bad pass (Dwight Howard steals)	2 - 2
10:51		**LeBron James makes two point shot**	**4 - 2**
10:51		Meyers Leonard shooting foul	4 - 2
10:51		LeBron James misses free throw 1 of 1	4 - 2

TIME	SCORE	Commentary
0'	0 - 0	随着主裁判一声哨响,上半场比赛开始.
3'	0 - 0	汉堡球员范德法特拼抢犯规,对手获得控球权.
3'	0 - 0	阿拉巴为拜仁慕尼黑赢得一个任意球.
4'	0 - 0	巴德尔吉为汉堡赢得一个任意球.
4'	0 - 0	拜仁慕尼黑球员古斯塔沃拼抢犯规,对手获得控球权.
4'	0 - 0	孙兴敏为汉堡在对方半场,赢得一个任意球.
4'	0 - 0	拜仁慕尼黑球员古斯塔沃拼抢犯规,对手获得控球权.
5'	0 - 0	曼西恩尼为汉堡在右路,赢得一个任意球.

Current Public Datasets

- Only a few public datasets for sports game summarization[1]

- Most of them are very small (100~200 games)

[1] Jianmin Zhang, Jin-ge Yao, and Xiaojun Wan. Towards constructing sports news from live text commentary. ACL 2016.

SportsSum: Sports Game Summarization Dataset

- A Chinese dataset
- Crawled from Sina Sports Live[1]
- 5,428 soccer games from 7 leagues
- For each game
  - Live commentary
  - Sports news article
- Data cleaning
  - Remove HTML tags
  - Remove unrelated news sentences
- https://github.com/ej0cl6/SportsSum

Live Commentary

Time	Scores	Commentary Sentence
66'	0-0	多特蒙德球员格策拼抢犯规,对手获得控球权. Dortmund's player Götze fouled, and the opponent got the possession of the ball.
66'	0-0	施魏因斯泰格为拜仁慕尼黑赢得一个任意球. Schweinsteiger got a free kick for Bayern Munich.
67'	1-0	进球啦！！！拜仁慕尼黑球员克罗斯大禁区外左脚射门,球从右下角飞进球门,球进了!助攻的是穆勒. 拜仁慕尼黑1-0多特蒙德. Goal!!! Bayern Munich's player Kroos shot with his left foot from the outside of the penalty area. The ball flew into the goal through the lower right corner. The ball went in! Muller gave the assist. Bayern Munich 1-0 Dortmund.
71'	1-0	拜仁慕尼黑球员里贝里大禁区左侧尝试右脚射门,可惜皮球高出球门.给他传球的是拉姆. Bayern Munich's player Ribery tried to shoot with his right foot from the penalty area's left side, but the ball was higher than the crossbar. Lahm passed the ball to him.

Sports News Ariticle

开场3分钟，克罗斯左侧任意球被顶到后点，里贝里禁区边缘抽射偏出近门柱。第8分钟，穆勒右路与曼朱基奇打出踢墙配合，在门前12米处推射被苏博蒂奇铲出底线。第13分钟，里贝里右路塞球，克罗斯在门前27米处抽射偏出近门柱。(...)
In the 3rd minutes, Kroos's free kick on the left was tipped to the back, and Ribery's shot from the penalty area missed. In the 8th minute, Muller and Mandzukic had teamwork, and Muller's shot from the 12 meters ahead the goal line was touched out by Subotić. In the 13th minute, Ribery passed the ball from the right, and Kroos's shot near the 27 meters ahead the goal line missed. (...)

[1] https://match.sports.sina.com.cn

Data Format for SportsSum

- Live commentary $C = \{(t_1, s_1, c_1), (t_i, s_i, c_i), \ldots, (t_m, s_m, c_m)\}$
  - t_i: timeline information
  - s_i: current scores
  - c_i: commentary sentence
- Sports news article $R = \{r_1, r_2, \ldots, r_n\}$
  - r_j: news sentence
- Metadata
  - Rosters
  - Starting lineups
  - Player positions

Live Commentary

Time	Scores	Commentary Sentence
66'	0-0	多特蒙德球员格策拼抢犯规,对手获得控球权. Dortmund's player Götze fouled, and the opponent got the possession of the ball.
66'	0-0	施魏因斯泰格为拜仁慕尼黑赢得一个任意球. Schweinsteiger got a free kick for Bayern Munich.
67'	1-0	进球啦！！！拜仁慕尼黑球员克罗斯大禁区外左脚射门,球从右下角飞进球门,球进了!助攻的是穆勒. 拜仁慕尼黑1-0多特蒙德. Goal!!! Bayern Munich's player Kroos shot with his left foot from the outside of the penalty area. The ball flew into the goal through the lower right corner. The ball went in! Muller gave the assist. Bayern Munich 1-0 Dortmund.
71'	1-0	拜仁慕尼黑球员里贝里大禁区左侧尝试右脚射门,可惜皮球高出球门.给他传球的是拉姆. Bayern Munich's player Ribery tried to shoot with his right foot from the penalty area's left side, but the ball was higher than the crossbar. Lahm passed the ball to him.

Sports News Ariticle

开场3分钟，克罗斯左侧任意球被顶到后点，里贝里禁区边缘抽射偏出近门柱。第8分钟，穆勒右路与曼朱基奇打出踢墙配合，在门前12米处推射被苏博蒂奇铲出底线。第13分钟，里贝里右路塞球，克罗斯在门前27米处抽射偏出近门柱。(...)
In the 3rd minutes, Kroos's free kick on the left was tipped to the back, and Ribery's shot from the penalty area missed. In the 8th minute, Muller and Mandzukic had teamwork, and Muller's shot from the 12 meters ahead the goal line was touched out by Subotić. In the 13th minute, Ribery passed the ball from the right, and Kroos's shot near the 27 meters ahead the goal line missed. (...)

Sports Game Summarization

- Input: live commentary $C = \{(t_1, s_1, c_1), (t_i, s_i, c_i), \ldots, (t_m, s_m, c_m)\}$
- Output: sports news article $\tilde{R} = \{\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_n\}$
- Goal:
  - $\tilde{R}$ covers most of important events in the sports game
  - $\tilde{R}$ is fluent and factually correct

Two-Step Summarization Model

- Consist of a selector and a rewriter
- Selector
  - Select important events from commentary
  - Classification problem
- Rewriter
  - Convert selected events to a news article
  - A sentence-level seq2seq model

C

$$(t_1, s_1, c_1)$$
$$(t_2, s_2, c_2)$$
$$...$$
$$(t_m, s_m, c_m)$$

Selector

Important Events

$$\{\tilde{c}_1, \tilde{c}_2, ..., \tilde{c}_n\}$$

Rewriter

New Article $\tilde{R}$
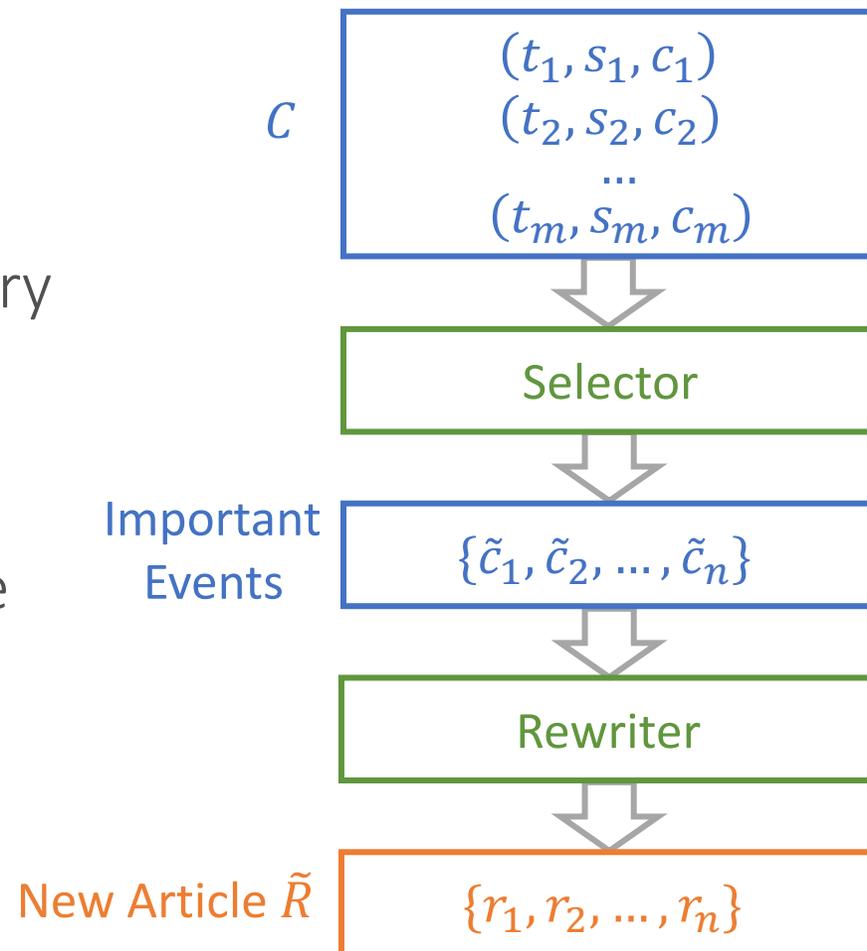
$$\{r_1, r_2, ..., r_n\}$$

Sentence Labeling and Mapping

- Labeling and mapping by timeline information
- Live commentary $C = \{(t_1, s_1, c_1), (t_i, s_i, c_i), \ldots, (t_m, s_m, c_m)\}$
  - t_i : timeline information
  - s_i : current scores
  - c_i : commentary sentence
- Sports news article $R = \{r_1, r_2, \ldots, r_n\}$
  - r_j : news sentence
  - Some news sentences start with "in the n-th minute"
- If a commentary sentence c_i and a news sentence r_j are mapped
  - For selector: $c_i \rightarrow$ positive label
  - For rewriter: $c_i \rightarrow r_j$

Name Mismatch Problem for Rewriter

- Simple seq2seq rewriter has the name mismatch problem
  - Generate high-frequency player names rather than the correct names

Live Commentary Sentence	里贝里禁区左侧尝试右脚射门,皮球高出球门.给他传球的是拉姆. **Ribery** tried to shoot with his right foot from the left side of the penalty area, but the ball was higher than the crossbar. **Lahm** passed the ball to him.
Gound Truth News Sentence	拉姆转移到左侧，里贝里突入禁区左侧距门12米处抽射高出。 **Lahm** passed the ball to the left, and **Ribery** cut in the left penalty area and shot from 12 meters ahead the goal line. The shot was too high.
News Sentence Generated by Seq2seq Model	里贝里传球，曼朱基奇禁区左侧射门偏出远门柱。 **Ribery** passed the ball and **Mandzukic**'s shot from the left side of the penalty area was out of the goalpost.

Tem2tem Rewriter

- Train the rewriter in a template-to-template (tem2tem) way
  - Replace player names with special tokens

	Input Sentence	Output Sentence
Seq2seq	射门!!!里贝里球门线跟前右脚射门, 被阿德勒横身扑出. 给他传球的是拉姆. Shoot!!! Ribery's right foot shot in front of the goal line was saved by Adler. Lahm passed the ball to him.	拉姆右路低传，里贝里前点铲射被阿德勒封出。 Lahm made a low pass on the right and Ribery's shot from the front was blocked by Adler.
Tem2tem	射 门!!![**player1**] 球 门 线 跟 前 右 脚 射 门, 被[**player2**] 横身扑出. 给他传球的是[**player3**]. Shoot!!! [**player1**]'s right foot shot in front of the goal line was saved by [**player2**]. [**player3**] passed the ball to him.	[**player3**] 右 路 低 传，[**player1**]前 点 铲 射 被[**player2**]封出。 [**player3**] made a low pass on the right and [**player1**]'s shot from the front was blocked by [**player2**].

Evaluation

- Similarity between ground truth news R and generated news $\tilde{R}$
  - ROUGE scores
- Two new scores for factual correctness
  - Name matching score (NMS): factual correctness of player names

$$\text{NMS}(R, \tilde{R}) = \text{F-score}(N_g, N_p)$$

  - Event matching score (EMS): factual correctness of events

$$\text{EMS}(R, \tilde{R}) = \text{F-score}(E_g, E_p)$$

Experimental Results

Method	Model	ROUGE-1	ROUGE-2	ROUGE-L	NMS	EMS
Extractive Models	RawSent	26.52	7.64	25.42	57.33	36.17
	LTR	24.44	6.39	23.19	51.63	29.03
Abstractive Models	Abs-LSTM	30.54	10.16	29.78	10.87	14.03
	Abs-PGNet	34.02	11.09	33.13	17.87	19.76
Selector + Rewriter (Seq2seq)	LSTM	41.39	16.99	40.53	28.48	25.19
	Transformer	41.71	18.10	40.96	35.63	30.94
	PGNet	**43.17**	**18.66**	**42.27**	48.18	36.94
Selector + Rewriter (Tem2tem)	LSTM	41.71	17.08	40.82	**59.54**	40.34
	Transformer	41.47	17.18	40.54	58.26	39.33
	PGNet	41.95	17.09	41.01	59.35	**40.46**

Our Two-Step Model   Other Baselines

Conclusion

- SportsSum
  - A Chinese sports game summarization dataset
- Two-step summarization model for SportsSum
  - Consist of a selector and a rewriter
- Two new scores for evaluating factual correctness
  - Name matching score
  - Event matching score

Thank You