足球赛事战报的自动写作研究

王文超[1] 吕学强[1,†] 张凯[2] 周建设[2]

1. 北京信息科技大学网络文化与数字传播北京市重点实验室，北京 100101；2.首都师范大学，北京 100048；
† 通信作者，E-mail: lxq@bistu.edu.cn

摘要 本文在分析不同类型体育赛事报道的特点后，首次提出了一种以实时数据作为数据源的足球赛事战报的自动写作方法。该法利用历史战报对实时数据进行自动标注，得到训练集，使用卷积神经网络（CNN）对标注后的实时数据进行建模，以自动识别出实时数据中的关键事件，通过模版技术将关键事件中结构化的信息，生成战报风格的自然语言。实验表明，该方法相比其他方法写作效果更好，内容更加详实，可以很方便地扩展到其他赛事的自动写作。

关键词 自动写作;足球;战报;实时数据

中图分类号 TP391

Research on Automatic Writing of Football Game News

WANG Wenchao[1]，LV Xueqiang[1,†] ZHANG Kai[2]，ZHOU Jianshe[2]

1. Beijing Information Science & Technology University, Beijing Key Laboratory of Internet Culture and Digtal Dissemination Research, Beijing 100101; 2. Capital Normal University, Beijing 100048;

†Corresponding Author, E-mail: lxq@bistu.edu.cn

Abstract After analyzing the characteristics of different types of sports events, this paper proposes an automatic writing method for soccer tournament with real-time data as data source for the first time. The real-time data is automatically annotated accoriding to historical news, and the training set is obtained. After annotation the real-time data is modeled by convolution neural network (CNN) to automatically identify the key events in real-time data. Events in the structure of the information, the formation of news style natural language. Experiments show that this method works better than other methods, and the content is more detailed and can be easily extended to the automatic writing of other sports games.

Key words Automatic writing; football; sports news ; real-time data;

足球被称为这个世界上的第一大运动，热爱足球的人们遍布世界的各个角落。我国拥有世界上数量最多的球迷。作为人们了解足球的重要信息来源，足球新闻在体育新闻中占据的比重往往是最大的[1]。因此，针对足球赛事战报的计算机自动写作研究日益成为热点。

自动写作的想法由来已久。近年来随着大数据、自然语言处理以及其他人工智能技术的不断发展，国内外逐渐掀起了用算法自动生成新闻报道的探索和实践[2]。由于中文的复杂性，中文自动写作比英文自动写作更加复杂。2006 年中国科学院计算所 NICA 小组开发了一种叙事与动画智能实验平台 PNAI，可以生成满足用户需求的叙事[3]。微软亚洲研究院分别于 2006 年和 2008 年公开上线微软对联系统第一版和第二版。可以根据用户给出的上联自动生成出若干下联[4]。2015 年腾讯财经开发的写作机器人 Dreamwriter 引用国家统计局公布的 8 月份 CPI 数据和统计分析师的分析预测，第一时间生成一篇财经报道[6]。2016 年 3 月 20 日，由清华大学语音与语言实验中心（CSLT）开发出的作诗机器人“薇薇”通过了“图灵测试”并通过社科院等唐诗专家评定[7]，2016 年 6 月北大计算机所推出的写作机器人 PKUWriter 能够基于电视/文字直播实时生成体育赛事新闻[8]。2016 年 8 月今日头条的 AI 机器人“张小明”，在里约奥运会期间发布了 456 篇奥运简讯[9]。2017 年 1 月南方都市报社写稿机器人“小南”正式上岗，并推出第一篇 300 余字的春运报道[10]。

由此可见，计算机自动写作应用日趋广泛。不同学者针对文字直播生成赛事新闻也进行了很多研究。Yang 等[11]提出基于隐马尔可夫与规则以及信息抽取相结合的方法，抽取赛事新闻的主要要素。高国洋[12]提出基于条件随机场结合规则，对体育赛事新闻进行命名实体识别，最终抽取赛事新闻中的实体关系。Xu 等[13]利用体育赛事文字直播数据和体育直播视频，对语义时间进行检测。Chen 等[14]使用无监督方法从文字直播中抽取语义时间。利用分层搜索算法协助进行视频标注。陈玉敬等[15]通过提出构建分差函数，基于分差函数的数据分片算法和数据合成算法，自动写作 NBA 赛事新闻。Zhu 等[16]基于 CRFs 结合正向关键词、反向关键词抽取足球直播文本中的关键句，自动写作足球赛事新闻战报。

1 实时数据及战报介绍

1.1 实时数据介绍

在体育赛事的比赛过程中，不同维度的标准化的比赛数据，由经过严格训练的数据公司员工收集并分析。一篇真实的实时数据如表（1）所示。在球场上发生动作后的几秒内，这些数据即被存入数据库并提供给外界。随着互联网产业的日益发展，以往只有专业的体育赛事分析公司才能接触到的实时数据，现在每一个互联网终端都可以便捷实时的得到。

实时数据以秒为最小时间单位，按照明确的标准，详细定义了某一时间点，比赛现场发生的事件。定义$Data$表示某场比赛所有的实时数据，Data_i表示一场比赛实时数据中第 i 个事件，二者之间的关系用(1)式表示。

$$Data_i \in Data(0 \le i \le Len(Data)) \tag{1}$$

完整的事件由 13 个变量组成，这些变量精确地还原了事件发生的前因后果。公式（2）定义了事件与事件变量之间的关系。

$$Data_i = \{GoalY, GoalZ, PassPlayer, PassX, PassY, Event, TeamName, PlayerName, PositionX, PositionY, Second, Minute, ExchangeName\}, \tag{2}$$

$GoalY$表示皮球落点的 Y 坐标，$GoalZ$表示皮球落点的 Z 坐标。如果该事件为射门，则$PassPlayer$表示则为助攻队友，$PassX$为传球的起点 X 坐标，$PassY$为传球的起点 Y 坐标，$Event$表示该事件的类型，$TeamName$表示该事件所描述的球队名称，$PlayerName$表示该事件所描述的球员名字，$PosstionX$表示该事件发生时，皮球所处的球场 X 坐标，$PossitionY$表示该事件发生时，皮球所处的球场 Y 坐标，$Minute, Second$表示该事件发生时的时间，如果该事件为换人，$ExchangeName$表示换上的球员名字。

表 1 实时数据样例
Table 1 Example of live sports data

…	PassPlayer	PassX	PassY	Event	TeamName	PlayerName	PositionX	PositionY	Second	Minute	…
…				射门被后卫挡出	巴塞罗那	苏亚雷斯	83.1	33.3	56	3	…
…	卡斯特罗	84.6	42.1	射偏	皇家贝蒂斯	塞巴略斯	80.1	39.2	36	91	…

1.2 足球战报介绍

战报为比赛结束后新闻编辑撰写的，针对本场比赛的新闻报道的一种题材，是众多体育新闻题材中的一种，一篇由编辑写作的战报节选如下

> 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。贝蒂斯第 21 分钟追平，丹尼斯-苏亚雷斯对古铁雷斯犯规，鲁文-卡斯特罗左侧距门 25 米处主罚任意球挂远角，1 比 1。

足球赛事战报具有简明扼要、篇章固定、通俗易懂、信息量大等特点。越来越多的球迷习惯于在足球比赛结束后的第一时间，通过阅读战报快速地了解整场比赛，通过阅读战报中记录的细节信息，去了解自己喜爱的运动员在绿茵上的风采。为了完成一篇既能满足读者需要，又要保证内容正确的战报，编辑需要认真地观看比赛的每一个细节。战报这项日益增长的用户需求，对撰写战报的编辑们提出了严峻的挑战。

2 足球战报的自动写作

2.1 整体流程

互联网上有针对足球赛事的多种类型的报道。包括每一场赛事的直播文本、实时数据、以及战报。本文的核心思路是：首先，分析处理已有的战报，找到实时数据中，编辑们认为最为关键的事件。间接地完成对实时数据进行标注，得到训练集。然后，训练一个基于卷积神经网络的模型，自动识别实时数据中的关键事件。由于实时数据中定义的事件，能够精确的反映出事件发生的时间、位置、动作目标、甚至是助攻队友。最后，提取出关键事件后，仅需针对不同类型的事件制作出少量的模板句，再将这些模板句填入模板库，一篇生动翔实的战报便呈现在读者面前。整体流程图如图 1 所示

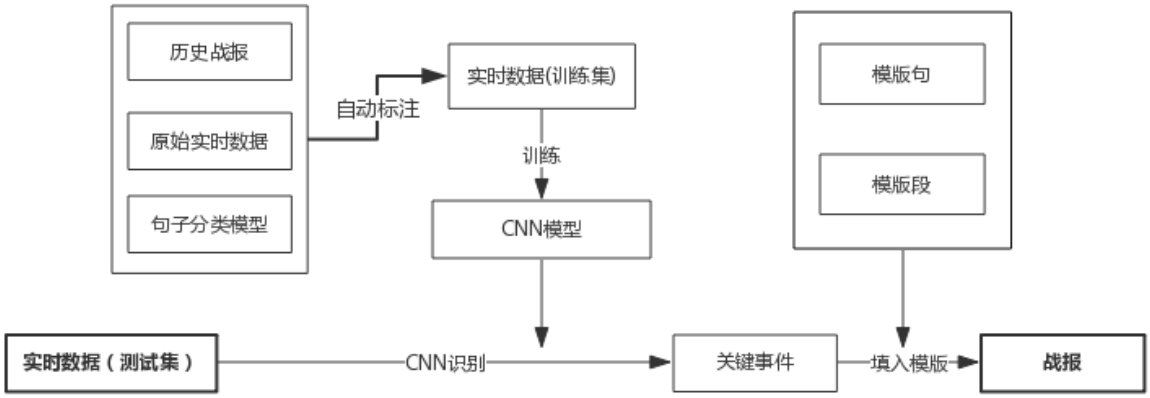

图 1 足球赛事战报流程图
Fig.1 Flow chart of football news

2.2 利用历史战报对实时数据进行标注

现实中，战报中的时间，实时数据中的时间，以及直播文本中的时间并不是一一对应的，为了更好的使用历史战报对实时数据进行标注，避免人工标注的同时，提高训练集的质量，本文提出了一个基于关键词的战报句子分类模型，将战报句子进行分类，辅以时间线信息，将二者的对应关系找到。

2.2.1 识别战报句子的类型

标准的事件由 13 个变量组成，Event是其中一个变量，它代表着该事件所属的类型。针对战报这一文体，通过关键词很容易分辨出一句话属于进球事件还是红牌事件。如表（2）所示。本文先是手动构建了一部分关键词表，然后基于爬取的全量历史战报训练一个 Word2vec 模型对此关键词表进行扩展，取得了很好的分类效果。

表 2 进球事件的关键词示例
Table 2 Example of key words in event goals

时间	战报句子	正向关键词	反向关键词
6	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	破门，禁区，推射，入网，门前，	阻止，解围，
30	梅西第 30 分钟错过进球，他与阿尔巴打出踢墙配合，阿尔巴突入禁区左侧回传，梅西在门前 11 米处左脚推射被挡后右脚半凌空抽射击中横梁！	配合，禁区，回传，推射，左脚，抽射，击中，横梁	被挡

Word2vec 是一个谷歌公司在 2013 年发布的一个开源程序。它使用深度学习算法能够有效地将词映射到向量空间里。本文采用了基于 Skip-gram 的预测模型，把对文本的处理简化为向量空间中的向量运算，通过计算向量空间上的相似度来表示文本语义上的相似度，实现相关词的扩展。

定义w_1为一个词，w_2为第二个词，定义$distance(w_1, w_2)$为两个词的距离，v_{w1}, v_{w2}为两个词的词向量表示

$$distance(w_1, w_2) = v_{w1} \cdot v_{w2}\,, \tag{3}$$

通过 Word2vec 扩展的关联词示例见表（4）。本文手工构造了 2430 个关键词，通过 Word2vec 选取与之相对应的相似度最高的 10 个关键词，最终得到了 6233 个关键词。

表 3 与“进球”相关的词
Table 3 The related words of “goals”

关联词	距离
破门	0.7012
击中	0.6984
入网	0.6891
推射	0.6732
直射	0.6725
攻门	0.6658

定义$j \in$ {所有事件类型}，$Evidence_{ij}$表示第 i个句子类型为j的证据值，P_{ij}表示第i个句子类型为j的概率值，wZ, wF, wP分别代表正向关键词、反向关键词、以及不同事件类型的参数矩阵，则有

$$Evidence_{ij} = wZ_{ij} * countZ_j/totalZ_j - wF_{ij} * countF_j/TotalF_j \ , \tag{4}$$

$$P_{ij} = wP_{ij} * Evidence_{ij}/\sum wP_{ij} * Evidence_{ij}, \tag{5}$$

使用该分类模型对战报中出现的每一句话进行计算，得到的结果如表（4）所示。

表 4　分类概率实例
Table 4 Example probabilities of classification

时间	战报句子	进球概率	射门被扑概率	…
6	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	0.84	0.11	…
30	梅西第 30 分钟错过进球，他与阿尔巴打出踢墙配合，阿尔巴突入禁区左侧回传，梅西在门前 11 米处左脚推射被挡后右脚半凌空抽射击中横梁！	0.22	0.79	…

2.2.2 结合分类模型对实时数据进行标注

为了更加准确地找到战报中报道的句子与实时数据中出现的事件对应起来，本文提出了一个基于时间窗口的扫描算法。为了计算战报句子与实时数据中事件相对应的概率值，该算法计算战报句子中出现的词，在该事件对应的事件类型，所对应的关键词表中的命中率，得到句子与事件相对应的概率

定义$count_i$为一篇战报中第i句所有词的总数，定义$count_j$表示这些词在事件j所属类型的正向关键词表中命中个数。P_{ij}表示第i句战报与第j个事件相关联的概率。$time$表示事件或者句子发生的时间，则有如下公式。如果$Right_i$所得值不唯一，说明算法没有准确匹配到，为保证训练集质量，丢弃该句

$$P_{ij} = \begin{cases} \dfrac{count_j}{count_i} \ , \left|time_i - time_j\right| \le 3 \\ 0 \quad , \left|time_i - time_j\right| > 3 \end{cases} , \tag{6}$$

$$Right_i = \text{argmax}(P_{ij}) \ , \tag{7}$$

至此通过构建正向关键词反向关键词表，并使用 word2vec 对词表进行扩展，本文得到了一个基于关键词表的句子分类模型。根据战报中每一句的类型，通过事件窗口算法，找到了每一句战报在实时数据中相应的事件。完成了实时数据的标注。

2.3 利用已标注的实时数据训练模型

2.3.1 数据预处理

标准 11 人足球场地长 105m 宽 68m，在标准的实时数据中，以客场球队所处的球场边的上方，角球区为原点，以球门线为 x 轴，以边线为 y 轴，地平面的垂线为 z 轴建立直角坐标系。一个描述进球事件的可还原成如图（2）所示。

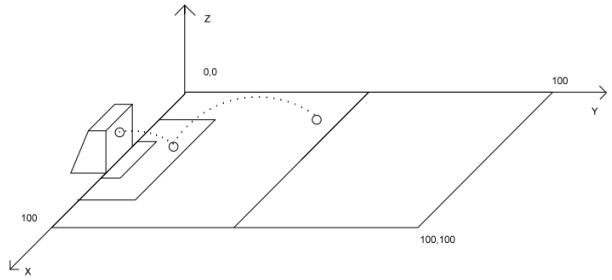

图 2　模拟进球事件
Fig.2 Simulated goal event

通过实时数据中的$PassX, PassY, PositionX, PositionY, GoalY$能够精确的知道皮球从球场的哪个点由谁传向哪个点，以及皮球最后射门时候相对于球门的空间位置。也能够方便的转化为读者所熟知的禁区、大禁区、左路、中路、边角等概念。

实时数据的射门事件中，有一个$GoalZ$变量，它记录了皮球在飞跃球门框一瞬间，所处的空间高度，以分米为单位。下图显示了一场比赛中所有射门时候皮球的落点。

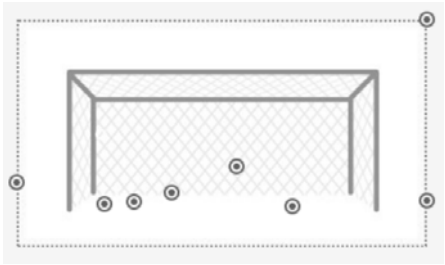

图 3 足球落点坐标

Fig.3 Placement coordinates of football

为了更好的训练 CNN 模型，本文按照如下公式将实时数据中的坐标信息离散化。

$$PositionX = \begin{cases} 0, & PositionX < 5.5 \text{ 或 PositionX} > 94.5 \\ 1, & 5.5 \le PositionX < 16 \text{ 或 } 84 \le \text{PositionX} < 94.5 \\ 2, & 16 \le PositionX \le 84 \end{cases}, \tag{8}$$

$$\text{PositionY} = \begin{cases} 0, & PositionY < 20 \text{ 或 PositionY} > 80 \\ 1, & 20 \le PositionY < 36.5 \text{ 或 } 63.5 \le \text{PositionY} \le 80 \\ 2, & 36.5 \le PositionY < 63.5 \end{cases}, \tag{9}$$

$$\text{GoalZ} = \begin{cases} 0, & GoalZ < 20 \\ 1, & 20 \le \text{GoalZ} < 35 \\ 2, & GoalZ \ge 35 \end{cases}, \tag{10}$$

其中PassX，GoalX与PositionX变换规则一致，PassY与PositionY变换规则一致。

TeamName, PlayerName, ExangeName, PassPlayer, Event等变量均可以通过新浪体育频道的公开接口，获取到与其对应的唯一 ID 值。查询过程如果出现了缺失值、错误值均用 0 代替。

将 10 万篇实时数据按照上述规则进行预处理后得到列数为 12 行数不定的 10 万个矩阵。至此，数据预处理完成。

2.3.2 设计卷积神经网络模型

卷积神经网络是近年发展起来，并引起广泛重视的一种高效识别方法。现在，卷积神经网络已经成为众多科学领域的研究热点之一。近年的研究表明，卷积神经网络同样适用于自然语言处理领域，在文本分类中的任务中表现优异。

本文设计了一个浅层的卷积神经网络模型。虽然仅仅是浅层的模型，却取得了很不错的分类效果。其模型大体结构如图（4）所示。

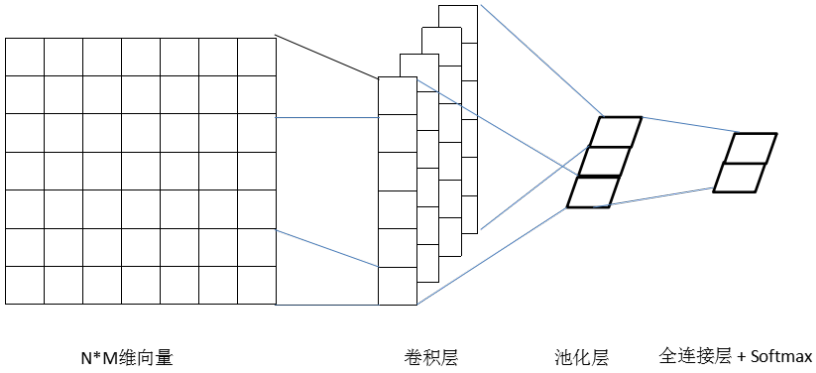

图 4 神经网络结构图

Fig.4 Architecture of neural network

假定n为一场赛事的实时数据中事件的个数。则输入层宽 12 列，高max(n)行。由于输入维度为 12，故设定卷积核的大小为$8 * 12$。如此一个卷积操作可以得到一个$\max(n) - 8 + 1$的特征图。多个卷积操作

可以得到多个这样的特征图。在输出层使用全连接层加 Softmax，得到一个$\max(n)*1$的概率分布。与标注的结果计算损失值。在模型训练上采用 ADAM 优化器来做梯度最速下降。

由于本文在之前只对每一个事件进行了 0 1 标注，（0 代表不重要，1 代表重要）在此处需进行一定的变换。定义V_i表示一篇实时数据里第 i 个事件的重要值度量，NewV_i表示转变后的新的重要值度量，则有

$$\mathrm{NewV}_i = V_i / \sum V_i \ , \tag{11}$$

本文选用了交叉熵作为损失函数。交叉熵产生于信息论里面的信息压缩编码技术，后来演变成为从博弈论到机器学习等其他领域里的重要技术手段。它的定义如下：

$$H_{y'}(y) = -\sum_i y'_i \log(y_i) \tag{12}$$

其中，Y 是要预测的分布，y' 为实际的分布。经过公式（11）转变后得到了实时数据的真实概率分布，按照公式（12）计算交叉熵后得到损失值，损失值在训练过程中前向传播，最终得到一个能够识别实时数据中关键事件的神经网络模型。

2.4 构建模版库，生成战报。

分析关键事件的各个变量可以精确的了解到整个事件发生的细节。比如根据一条进球类型的事件，可以清楚的知道，皮球从哪个坐标被哪个助攻队友传到了哪个坐标，而这个球员接到球后，把球踢进球门的哪个相对位置。本文将事件中球员的 x 轴坐标转化为禁区、大禁区、将 y 坐标转化为中路、左路、右路、边角。为了更加生动形象地描述事件，本文构建了一个自适应模版算法，最大化地将事件细节使用生动的语言描述出来。模版句设计思想如图 5 所示。将模板句代入由简单的连接词构成的模板库中，即完成了赛事战报的自动写作。

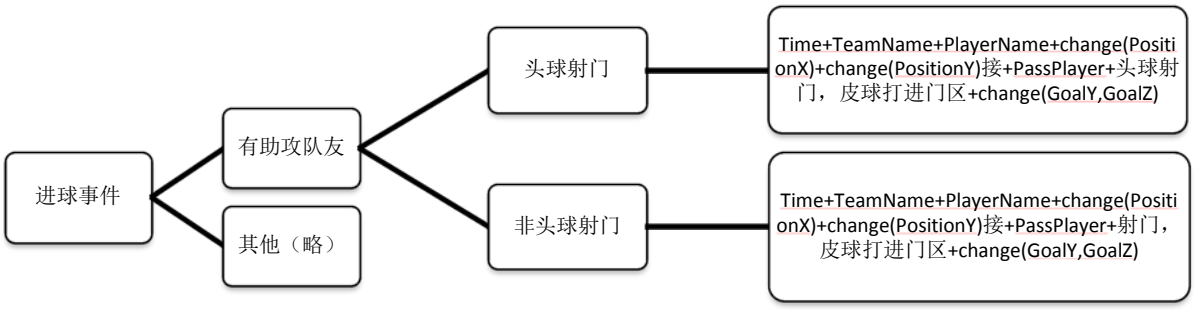

图 5　模版设计思想
Fig.5 Stencil design idea

3 实验结果与分析

评价一篇体育赛事战报的指标因人而异。从读者角度出发，本文提出了三个方面的评价指标，分别是：关键事件的覆盖率（Critical incident coverage）、细节的还原率（Detail reduction rate）、语言表达正确率（Correct expression rate），组织了三名同实验室的足球爱好者在不知道战报来源的情况下，对其进行评分，最后将得到的评分简单平均，以期能够量化读者对战报的评价。从机器角度出发，本文将采用公开的评价算法 ROUGE 进行对比评价。在实验设置上，本文将进行纵向对比来说明优化问题，横向对比来说明优势问题。本文所使用的全部实验数据均从新浪体育国际足球频道爬取。共 10 万条历史比赛数据。其中，8 万条作为模型训练数据，另外 2 万条作为测试数据。

为了最大程度提高算法的效果，本文进行了以下优化：

1) 提出使用 Word2vec 对关键词表扩充，提升句子分类算法的正确率，
2) 提出一个算法计算句子与指定事件相对应的概率值，在设定的时间窗口内选择概率最大的事件，以提高实时数据的标注质量。
3) 在数据预处理部分，将实时数据中的坐标转化为具有真实意义的离散值。降低了模型的训练难度提升了识别性能。
4) 由于每场比赛中事件是否属于关键事件都是相对的。同为带球过人事件，球星被普通球员过，能算作一个重要事件，普通球员被球星过，便是稀松平常。实验结果表明采用卷积神经网络模型相对于 SVM 模型有助于提升关键事件的识别率。

为了构造纵向对比实验，以说明上述优化策略的有效性，本文实现了含有不同优化项目的代码版本，优化结果如表（5）所示。从实验结果来看，这些优化取得了理想的效果，并且在未来的工作中还有进一步优化提升的空间。

表 5 优化效果对比
Table 5 Optimization effect comparison

优化	ROUGE-1		ROUGE-2		ROUGE-SU4		CIC	DRR	CER
	Recall	F-value	Recall	F-value	Recall	F-value			
1	0.369	0.361	0.114	0.159	0.189	0.192	0.73	0.69	0.98
1,2	0.474	0.552	0.189	0.199	0.201	0.195	0.77	0.73	0.98
1,2,3	0.573	0.661	0.239	0.258	0.241	0.263	0.89	0.88	0.98
1,2,3,4	0.579	0.668	0.250	0.261	0.243	0.269	0.92	0.91	0.98

在横向对比实验中，为了对比此方法与同类方法的优缺点，本文实现了 Zhu 等基于 CRF 信息抽取进行自动写作的程序，Chen 等基于分差函数数据分片以及数据合成算法的自动写作程序。二者描述的方法在 NLPCC-ICCPOL2016 评测任务 5：足球新闻自动生成中都取得了优异的成绩。在输入相同的测试数据后，在上述的同一评价体系下得到了下表（6）的数据。

表 6 横向对比
Table 6 Horizontal contrast

方法	ROUGE-1		ROUGE-2		ROUGE-SU4		CIC	DRR	CER
	Recall	F-value	Recall	F-value	Recall	F-value			
Zhu 等	0.555	0.602	0.246	0.253	0.262	0.242	0.79	0.81	0.90
Chen 等	0.369	0.361	0.114	0.159	0.189	0.192	0.81	0.79	0.91
本文	0.579	0.668	0.250	0.261	0.243	0.269	0.92	0.91	0.98

以 2016-2017 赛季西甲第一轮巴塞罗那对皇家贝蒂斯为例。表（7）节选了包括小编战报在内的四段内容。

表 7 生成结果对比
Table 7 Generation result contrast

版本	内容
Zhu 等	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，助攻的是阿尔巴，巴塞罗那 1-0 皇家贝蒂斯
Chen 等	进球啦！！！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
本文	上半场第 5 分钟巴塞罗那 7 号图兰小禁区中路接 18 号阿尔巴传球右脚射门，皮球打进门区正下方。1 比 0.
编辑	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。

从内容上看，Zhu 等通过对直播文本进行类型识别，去除了一些导致上下文语句不连贯的感叹词或者将一些短语变换表达方式。语言整体比较流畅。Chen 等通过分差函数等方法正确的识别到这一关键事件，由于缺乏进一步的处理，语句整体风格跳跃性比较大，取决于直播文本编辑的语言文字风格。二者与小编战报一样，语言都比较灵活。本文方法在细节描述上仅次于小编战报，优于其他二者，但现阶段

语言表达上有些许死板，缺乏一些灵活轻松的元素。值得强调的是，由于本文使用算法分析实时数据，数据相关的内容不会出错，而小编战报中涉及数据的内容却不一定能经得起推敲，因而，在数据信息准确性上本文可能是要优于小编战报的。

由于 Zhu 等以及 Chen 等方法的本质，是从直播文本中选择关键句，然后对关键句进行处理，形成一篇战报，而本方法是从精确的结构化实时数据中选取关键事件，基于数据生成合适的语言文字，形成一篇战报，在细节描述，以及语言通顺上具有得天独厚的优势。同时，由于设计上的优势，本方法仅需要手工收集整理几千个对应事件的关键词，编辑少量的模版句。避免了费时费力的人工标注过程。

通过针对 NBA 赛事新闻，重新构建事件类型的关键词表以及少量的模版，成功写作出 NBA 新闻稿。本文提出的战报写作方法，同样适用于于 NBA 赛事新闻写作，因此该方法在赛事新闻自动写作领域具有很强的通用性。

4 总结

本文通过对历史战报的分析，对实时数据进行标注，得到训练集，训练了一个卷积神经网络模型来识别实时数据中的关键事件，依据这些关键事件中的结构化数据，生成自然语言对数据进行描述，最终完成足球赛事新闻的自动写作。整个流程只有提取关键词以及构建模版句方面需要专业人士参与。实验结果表明该方法可以方便地扩展到其他的体育赛事项目中。不足之处为模版库数量过少。导致生成的结果语句上略显古板。在未来的工作中可以考虑使用结合外部数据的循环神经网络自动生成文字。

参考文献

[1] 邵常恩. 汉英足球新闻特点及翻译[D].中国海洋大学,2012.

[2] Manyika J, Chui M, Brown B, et al. Big data: The next frontier for innovation, competition, and productivity[J]. 2011.

[3] 曹存根, 岳小莉, 眭跃飞. PNAI:一种新型的叙事与动画智能实验平台[J]信息技术快报，2006.

[4] Jiang L, Zhou M. Generating Chinese couplets using a statistical MT approach[C]//Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, 2008: 377-384.

[5] Zhou M, Jiang L, He J. Generating Chinese Couplets and Quatrain Using a Statistical Approach[C]//PACLIC. 2009: 43-52.

[6] 腾讯财经开发自动化新闻写作机器人 Dreamwriter[EB/OL].(2015-09-11)[2017-06-01].http://www.cac.gov.cn/2015-09/11/c_1116532821.htm

[7] Wang Q, Luo T, Wang D, et al. Chinese song iambics generation with neural attention-based model[J]. arXiv preprint arXiv:1604.06274, 2016.

[8] Zhang J, Yao J, Wan X. Toward constructing sports news from live text commentary[C]//Proceedings of ACL. 2016.

[9] AI 机器人张小明[EB/OL].(2016-08-29)[2017-06-01].http://www.sohu.com/a/112543973_372481

[10] 写稿机器人“小南”上岗[EB/OL].(2017-01-18)[2017-06-01].http://epaper.oeeee.com/epaper/A/html/2017-01/18/content_4285.htm

[11] Yang Y, Li L. Research on sports game news information extraction[C]//Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. International Conference on. IEEE, 2007: 96-101.

[12] 高国洋. 体育领域信息抽取系统的研究 [D][D]. , 2009.

[13] Xu C, Zhang Y F, Zhu G, et al. Using webcast text for semantic event detection in broadcast sports video[J]. IEEE Transactions on Multimedia, 2008, 10(7): 1342-1355.

[14] Chen C M, Chen L H. A novel approach for semantic event extraction from sports webcast text[J]. Multimedia tools and applications, 2014, 71(3): 1937-1952.

[15] 陈玉敬, 吕学强, 周建设, 等. NBA 赛事新闻的自动写作研究[J]. 北京大学学报 (自然科学版), 2017, 53(2): 211-218.

[16] Zhu L, Wang W, Chen Y, et al. Research on Summary Sentences Extraction Oriented to Live Sports Text[C]//International Conference on Computer Processing of Oriental Languages. Springer International Publishing, 2016: 798-807.