

专 业 学 位 硕 士 学 位 论 文

面向百科数据的军事事件抽取

Military Event Extraction for Encyclopedia Data

作 者 姓 名: 高鸿博

工 程 领 域: 软件工程

学 号: 41817020

指 导 教 师: 刘宇

完 成 日 期:

大连理工大学

Dalian University of Technology

摘 要

知识图谱构建已经成为自然语言处理的一个重要发展方向，也是信息抽取领域一个重要的目标。本文从百科文本中标注军事图谱，并负责其中的事件抽取部分。事件抽取作为信息抽取的关键技术，在信息检索、自动问答等领域被广泛应用。

本文爬取军事相关百科文本，根据文本内容，标注军事图谱数据集，定义了四个事件触发词和七个事件元素实体标签，和实体之间的关系标签和事件之间的关系标签，以及实体、触发词和元素角色之间的约束关系。

事件元素实体抽取和事件触发词抽取是本文事件抽取的关键部分。本文观察数据集中事件元素实体的特征，发现与军事专业相关的实体的后缀取词范围很小，根据这一点设计出了相应的后缀特征，并通过对比实验验证两处输入位置后缀特征重复几次对于模型效果提升最高。在军事事件中，军事专业的事件元素实体和军事事件触发词语义相关，于是本文基于此设计基于事件元素实体的注意力机制，以提高事件触发词的抽取效果。最终经过实验证明，本文的模型与其他经典事件抽取模型的相比取得了较高的效果。

接下来是元素角色抽取。元素角色是指事件触发词和事件元素之间的关系。本文使用双仿射深度文法依赖树算法的关系分类器作为基础，使用词的最后一个字的向量代表整个词，通过掩盖矩阵输入了实体、触发词和元素角色之间的约束关系，提高元素角色抽取效果。最终，通过对比实验得知，掩盖矩阵对模型效果有较大提升。通过对比实验得知，本文的模型与其他经典模型相比取得了优越的效果。

最终本文使用上述的三个模型进行事件抽取，使用模板匹配的方式进行关系抽取，设计了一套事件关系抽取系统。系统通过本文已经验证的神经网络模型进行事件抽取。事件关系抽取模板通过关系连词和正则表达式构建。系统在句子输入之后通过关系连词对句子进行过滤，减少神经网络模型计算量。最后本文列举了系统抽取的部分结果。

关键词：事件抽取；注意力机制；双向 GRU

Military Event Extraction for Encyclopedia Data

Abstract

Knowledge graph building has become an important development direction of natural language processing, and also an important goal in the field of information extraction. In this paper, we annotated the military map from the encyclopedia text and took charge of the event extraction part. As the key technology of information extraction, event extraction is widely used in information retrieval, automatic question answering and other fields.

This paper crawled the military related encyclopedia text, according to the content of the text, annotated the military atlas dataset, defined four event trigger words and seven event element entity tags, the relationship tags between entities and entities, the relationship tags between events and events, and the relationship tags between event triggers and entities.

Event argument entity extraction and event trigger extraction are the key parts of event extraction. In this paper, we observed the characteristics of event argument entities in the dataset, and find that the range of suffixes of entities related to military is very small. According to this, we design the suffix feature and add it to the model. And through comparative experiment, it was found that how many times the suffix feature repeated to make the model best. In military events, the event argument entities about military are related to the military event trigger in semantic information. Therefore, this paper designed an attention mechanism based on event argument entity to improve the effect of event trigger extraction. Finally, compared with other classical models, we have achieved relatively high effect.

Next part is argument role extraction. Argument role is the relationship between event trigger and event argument. This paper was based on the role classifier of biaffine module, used the vector of the last letter to represent the whole word, inputted the constraint relationship among entity, event trigger and argument role through masking matrix, so as to continuously improve the effect of the model. At last, through comparison with other classical models, the model in this paper has achieved superior results.

Finally, this paper uses the above three models for event extraction, and uses template matching method for relationship extraction, and designs a set of event relationship extraction system. The system extracts events through the neural network model which has been verified in this paper. Event relation extraction template is constructed by relation conjunction and regular expression. After the sentences are input, the system filters the sentences through

relational conjunctions to reduce the calculation of neural network model. Finally, some results of system extraction are listed.

Key Words: event extraction; attention mechanism; Bi-GRU

目 录

摘 要	I
Abstract	II
1 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状与分析	2
1.2.1 基于模板匹配的事件抽取	3
1.2.2 基于特征的机器学习方法	3
1.2.3 基于神经网络的事件抽取	4
1.3 论文研究内容	6
1.4 本文的组织结构	7
2 理论基础与相关模型	9
2.1 事件抽取和数据集简介	9
2.1.1 数据集简介	9
2.1.2 事件抽取任务	11
2.2 深度学习相关技术	12
2.2.1 循环神经网络	12
2.2.2 长短期记忆网络	13
2.2.4 注意力机制	15
2.3 词向量	16
2.3.1 word2vec 词向量	17
2.3.2 bert 词向量	18
2.4 性能评价指标	19
2.5 本章小结	20
3 事件触发词和事件元素实体抽取	21
3.1 事件触发词和事件元素实体抽取问题描述	22
3.2 事件元素实体抽取	22
3.2.1 模型总体结构	22
3.2.2 bert 字向量及后缀特征	23
3.2.3 双向门控循环单元	25
3.2.4 自注意力机制	27

3.2.5	CRF 条件随机场	27
3.2.6	损失函数	28
3.3	事件触发词抽取	28
3.3.1	模型总体结构	29
3.3.2	模型输入	29
3.3.3	基于事件元素实体的注意力机制及线性混合	29
3.4	实验结果与分析	30
3.4.1	事件元素实体抽取实验	30
3.4.2	事件触发词抽取实验	32
3.5	本章小结	35
4	元素角色抽取	36
4.1	元素角色抽取问题描述	36
4.1.1	问题描述	36
4.1.2	解决方案	37
4.2	模型总体结构	37
4.2.1	模型输入及 Bi-GRU	38
4.2.2	线性混合及双仿射	39
4.2.3	掩盖矩阵和 softmax	40
4.2.4	损失函数	41
4.3	实验结果与分析	41
4.3.1	掩盖矩阵对比实验	41
4.3.2	不同模型之间的对比实验	42
4.4	本章小结	43
5	中文军事事件关系抽取系统	45
5.1	中文军事事件关系抽取系统框架	45
5.2	数据预处理模块	46
5.3	事件抽取模块	47
5.4	事件关系抽取模块	48
5.5	基于关系连词的文本过滤模块	49
5.6	本章小结	49
结 论	52
参 考 文 献	54

致 谢	58
--------------	----

1 绪论

1.1 研究背景与意义

当今我国的军队建设已经进入信息化,现代化建设时代。信息化军队的主要特征有,武器装备的信息化智能化,同时以计算机和网络为中心,实现武器装备的无缝连接。军队的信息化使得作战不再只关注火力和机动性。无论是在日常作战训练中,还是有关军事的新闻,亦或是不同时期的作战条令作战手册,都会产生大量的文本信息。同时,机器学习和深度学习领域正在飞速发展,以自然语言处理为基础的信息抽取技术逐渐成熟。从海量的信息中,如何快速准确地定位分析出人们真正需要的信息变得十分重要,而这正是信息抽取领域亟待解决的问题^[1,2]。

所谓信息抽取,是从文本中抽取信息创建一个结构化的表示形式,包含了实体识别,关系抽取,事件抽取,事件关系抽取,槽填充(实体属性抽取)等多个领域,可以形成知识图谱,事理图谱等结构化形式,方便于后续的信息检索,预测推荐。将信息抽取技术应用到军事建设中,可以通过实体识别和槽填充,从文本中统计各种装备的性能;通过实体识别和关系抽取,可以构建知识图谱,方便对相关知识进行检索;通过事件抽取和事件关系抽取,可以构建事理图谱,方便对军事事件进行预测。对军事新闻抽取,可以获取各方军队的最新动向,一定程度预知对方接下来的行为;对军事手册进行抽取,可以建立军事行动的电子指南,方便对手册进行检索教学;对过往战争案例进行抽取,可以对以前的作战行动进行概括,方便人们从过往战争中总结经验。

事件的概念起源于认知科学,广泛应用于哲学、语言学、计算机等领域,是信息的重要表现形式之一。在计算机科学的范畴内,自动抽取评测会议(Automatic Content Extraction, ACE)对事件作出如下定义:是发生在特定时间、特定地域范围内的,由一或多动作组成的一件事情^[3-5]。事件识别和抽取研究如何从描述事件相系的文本中识别并抽取出事件信息并以结构化的形式呈现出来,包括其发生的时间、地点、参与角色以及与之相关的动作或者状态的改变。其中定义的主要元素有事件触发词,事件元素,元素角色,事件类型。

事件触发词:事件描述中最能代表事件发生的词,是决定事件类别的重要特征,在ACE评测中事件触发词一般是名词或动词。

事件元素:事件的参与者,是组成事件的核心部分,与事件触发词共同构成事件的整个框架。事件元素主要由实体、时间和属性值等表达完整语义的细粒度单位组成。

元素角色：事件元素与事件之间的语义关系，也就是事件元素在相应的事件中扮演什么角色。

事件类型：事件元素和触发词决定了事件的类别。很多评测和人物均制定了事件类别和相应模板，方便元素识别及角色判定。

在抽取出上述四种事件组成要素后，就可以结构化描述一个事件。如果像实体一样，抽取出事件之间的关系，可以构建事理图谱。事理图谱具有与知识图谱类似的功能，例如信息检索，推荐系统等，还可以进行事件预测。在军队信息化建设的今天，军事相关的图谱建设可以提高军队在文本方面的信息化应用程度。例如，可以从过往的训练总结或者旧的军事作战详情中构建图谱，可以更清晰地观察到每个作战事件之间的关联，从而对过往的军事行动中的缺点进行总结；也可以将军事作战手册进行图谱构建，方便前方作战人员在缺乏后方信息交流的情况下，有一种相对可靠的作战指导方案，同时还可以从中检索敌方武器的基本性能，为后续作战提供更多的信息指导。

但是，事件抽取和实体识别最大的不同是，事件抽取需要事件触发词和事件元素两者，还要将事件触发词和事件元素连接起来，实体识别只需要实体，所以事件的数据集标注比实体的要复杂困难许多，导致了事件的数据集数量少且质量不高。本文首先从网络上爬取了中文军事百科数据，标注了实体，事件，实体之间的关系，事件之间的关系，形成了事理图谱和知识图谱融合在一起的综合图谱，其中将事件元素作为一个实体存在。本文负责其中的事件抽取部分，包括抽取事件触发词、事件类型、事件元素、元素角色四项，其中事件元素都有对应的实体类型，而且实体类型会对元素角色有约束关系。本文使用中文 bert 字向量抽取，使用 BIO 标签来消除中文事件抽取使用字向量的影响，并且利用了实体类型和元素角色之间的约束关系，使得元素角色的抽取效果获得提升。

1.2 国内外研究现状与分析

事件抽取作为信息抽取的重要组成部分，虽然事件的结构复杂导致抽取难度比实体识别高，但由于其蕴含的信息量更多，从上世纪末就有相关抽取方法出现。最初人们使用的是模板匹配的方案。这种方案适用于特定领域，人工成本高泛化能力差，需要专家构建模板。之后信息抽取会议定义了事件抽取的要素，公布了标准数据集，大大推进了事件抽取的研究工作，事件抽取的评价标准也逐渐统一。尤其是 ace2005 数据集，至今仍是中文事件抽取领域中少有的标准数据集，但是数据集存在数据量少，标签分布不均，有的事件标签数量仅有 10 个左右等问题。但是从此，事件抽取研究迈入正轨。接下来是人工定义特征，然后通过机器学习方法进行抽取的时期。由于使用了机器学习方法，泛化能力得到提高，不再需要专家去定义模板。最后由于使用显卡计算大大推进了神经

网络模型的发展，神经网络的时代到来，深度学习使得事件抽取效果得到进一步提高，进一步减少人工干预，不再需要专家定义模板或者定义特征，但是仍处于缺少高质量的标准数据集的困境中。

1.2.1 基于模板匹配的事件抽取

在早期，信息抽取基本都是通过模板匹配的方式进行抽取。这种方案的问题在于，需要专家来编写模板；模板的泛化能力差，很难应用到另一领域或者格式不同的其他数据源上。Ellen 等^[6]观察数据集规则，例如发现在新闻报告中事件元素第一次出现的地方往往有关事件的最重要的信息，以及事件元素的周围往往会有描述事件元素在事件中的角色的描述，基于这两点 Ellen 等提出了事件模式抽取系统 AutoSlog。这种人工模板的方式需要比较多的人工，于是人们尝试能否从语料中预测出某些模板，或是实现模板的自动获取。例如 Ellen 等^[7]开发的升级版事件模式抽取系统 AutoSlog-TS 就通过预分类语料，使用自然语言处理工具（词性，句法分析等），以及一定的规则和统计学方法，解决标注标准不一致的问题，同时减少了人工。Kim 等^[8]开发的 PALKA 系统，假设了特定领域中高频出现的模式是有限的，通过训练文本以 FP 结构的形式构建模板，在泛化过程中采用了归纳学习机制，实现了模板的自动获取。

1.2.2 基于特征的机器学习方法

基于特征的机器学习方法比基于模板的方法的泛化能力要更强。这些特征虽然需要人工定义，但是只需要标准的数据集，而不再需要专家去定义模板，减少了人工劳动。特征有自然语言处理结果，例如词性，深度文法依赖，或者前后的词，等等；机器学习方法有最大熵，隐马尔可夫模型，支持向量机等。Chieu 等^[9]首次将最大熵模型应用于事件抽取，使用了 unigram, bigram, 命名实体等各种特征。Ahn 等^[10]将事件抽取分成多个小任务，包括触发词识别，事件元素识别，事件属性抽取，事件共指等。然后通过定义特征，包括语义特征（全词，全词小写形式，原词，词性深度文法依赖），wordNet（一个单词词典），左右附近的词，实体特征等，使用最近邻和最大熵方法，抽取事件。在事件触发词抽取中，发现 ace2005 英文数据集中事件触发词的各类型之间数量差别很大，None 类占绝大多数，事件触发词在 95% 下都是一个词，于是将触发词识别做成两阶段的词分类任务，第一阶段判断词是否是一个触发词，第二阶段将词进行分类成各个事件类型，减少事件类别标签不均的影响。Ji 等^[11]通过跨文档信息总结的方式来抽取事件，比如在全文档中该词为事件触发词的概率很低，则即使在该文档中通过周围的信息来判断这个词可能是触发词，也要降低其置信度。而 Hong 等^[12]则从实体的角度出发，经过统计发现实体与事件触发词之间具有关系，例如某种实体作为某类事件触发词的某种元

素角色的概率较高，最后通过 SVM 等方法进行事件抽取。Li 等^[13]则发现抽取事件是应该不仅观察局部特征，还应该观察全局特征，同时联合抽取事件触发词和事件元素，以消除分阶段抽取带来的误差影响。

1.2.3 基于神经网络的事件抽取

随着最近几年显卡作为神经网络计算工具的出现，带动了神经网络技术的发展。神经网络的特点有，不再需要专家来编写模板，也不再需要利用自然语言处理工具来获得特征，只需要合理的神经网络结构，以及足量的标准数据集，预测的效果就可以超过消耗大量人工编写的模板方法和基于特征的机器学习方法。但是抽取一个事件，需要抽取事件触发词，事件元素，已经它们的类别和相互之间的关系。这导致标注一个事件数据集需要标注的东西很多，公开的标准数据集很少，而且数据集的标签比例偏差很大。同时早期的事件抽取方法都是英文的事件抽取，只考虑了事件触发词只是一个词的情况，这样的方法直接应用到中文事件抽取时需要考虑分词带来的影响；而后来出现的 bert 动态词向量，中文的 bert 只有字向量，导致这些方法完全无法使用 bert 抽取中文事件。目前使用 bert 字向量进行中文事件抽取的文献数量很少。

Nguyen 等^[14]使用 CNN 卷积神经网络来进行事件检测。神经网络的输入由三部分组成：静态字向量，位置向量，实体类型的 BIO 信息向量。然后输入经过几个不同尺寸的卷积核，最后是全连接层，得到最终输出。而 Chen 等^[15]提出了 DMCNN 模型，使用 CNN 卷积神经网络，分两个阶段完成了事件抽取。神经网络结构也分成两部分，一部分输入与 Nguyen 等^[14]的类似静态字向量，位置向量，实体类型的 BIO 信息向量，之后经过几个不同尺寸的卷积核，然后将输出依据触发词和事件元素的位置分割成三部分，经过池化连接到一起；另一部分就是将触发词和事件元素的左右范围内的字向量连接；最后将两个分部输出连接到一起通过全连接层输出。在触发词抽取时，需要第一部分卷积神经网络的输出的分割点减少。缺点是，仅适用于触发词和事件元素是一个词的情况，尤其对于主要以字为单位的中文事件抽取。

而在自然语言处理中，RNN 要比 CNN 更常见，之后的算法中，RNN 类的事件抽取算法数量更多。Nguyen 等^[16]提出了 JRNN 算法，使用 word2vec 静态字向量、实体类型向量、深度文法依赖作为输入，经过双向 RNN，然后经过复杂的环节联合抽取事件触发词和事件元素。JRNN 假定了一个句子中同时出现的不同事件触发词和事件元素之间可能会出现某种相互关系，维持了几个已经抽取出的事件触发词和事件元素状态的矩阵，利用这些矩阵和双向 RNN 的输出进行事件抽取。

接下来，出现了一些与众不同的中文事件检测方法。Lin 等^[17]提出了 NPN 事件检测方法，没有采取常用的序列标注的形式。通过观察事件触发词，发现这些词具有某些特

定的格式，而且中文事件触发词的长度十分有限。于是提出一个词分类模型，在候选位置，遍历所有长度小于 3 的字符串，通过 CNN 方法，来进行词的分类。同时算法的输入由字向量和词向量经过多种方式组合而成并进行对比实验。这种分类候选词的方式在神经网络算法中比较罕见。Ding 等^[18]提出的 TLNN 事件检测方法，输入上，使用了静态字向量，以及基于 HowNet 得到的与该句相关性的向量，该向量基于 HowNet 知识库，能够区分多义词；神经网络上，使用了触发词感知的网格 LSTM，可以同时提取字、词、句子的信息；最后使用 CRF 预测序列标注的结果。以上两个模型，一个通过遍历长度小于 3 的候选词并进行分类，一个利用了知识库 HowNet 的信息区分多义词，这些方法源于之前的基于模板的方法和基于特征的机器学习方法，但是到了 bert 的时代，特殊的输入使得这些模型难以接入先进的动态字向量。

在 RNN 和 CNN 都用于自然语言处理之后，人们把目光放在了寻求性能更加强大的动态字向量身上。于 2018 年，谷歌提出了 bert 字向量^[19]，打破了多项记录，同时也改变了自然语言处理的输入。在 TLNN^[18]中，作者使用了外界的知识库来为多义词提供不同字向量表示，而 bert 这类的动态字向量能通过模型轻松分辨不同语言环境下同一个词的不同含义。但是 bert 字向量训练成本很高，所以谷歌等开发者将自己训练的各式动态字向量上传，人们只需要下载，减少重复训练字向量的成本。但同时也默认使用了模型上传者分词方式。目前中文 bert 为字向量，以字为单位，所以之前同时使用词向量和字向量的 NPN 模型无法使用 bert 字向量；TLNN 在输入上需要的特殊的字向量也无法直接转换为 bert。而在 bert 提出之后，事件抽取开始使用 bert 作为输入。

Zhang 等^[20]使用 bert 提出了一种事件元素抽取的方法。首先已经预知了事件触发词的位置，使用 bert 作为模型输入，将事件触发词和句子中其余词对应的向量，放入到双仿射深度依赖树模型的 role 分类器中，判断哪个词是事件元素的结尾以及事件元素的元素角色，再通过一个 MLP 分类器来分别事件元素的起始位置。但是，文中是英文的事件数据集，其中大多数事件触发词为单独的一个词，且 bert 英文为词向量，直接使用事件触发词的 bert 向量代表触发词本身即可，而将这个模型应用到中文上时需要考虑中文与英文之间事件触发词和 bert 字向量上的差异。而 Yang 等^[21]提出了分成两阶段的英文事件抽取模型 PLMEE。事件触发词抽取部分，使用 bert 向量作为输入，然后使用线性全连接层进行分类。事件元素抽取部分，和触发词抽取几乎一样，只是最后的多分类对于每个元素角色分成开始和结尾两种类型，应对同一个事件元素可能有多种元素角色的出现。最后作者为事件元素模型的多分类输出设计了获取事件元素位置的算法。

但是 ace2005 仍是目前公认的标准数据集，数据集的数据量比较少，事件类型的数量比例很不均衡，同时事件的标注难度也比较高，所以人们发展出一些方法适应来自数

据集的困境。Ma 等^[22]人，通过远程监督来提高数据集的数据量；同时假设分辨事件的类型，事件的触发词并不是必须的，只要有关键的事件元素就能敲定事件的类型。Huang 等^[23]却使用另一种方式，通过 zero-shot 的方法，减少模型对于数据集的需求，提高了对于标签数量较少的事件的抽取能力；通过迁移学习方法，提高了模型的泛化能力。

综上所述，事件抽取方法可以简单分为两类：分阶段抽取和联合抽取。分阶段抽取就是先抽取事件触发词，然后抽取事件元素，如 DMCNN 模型；联合抽取就是在一个模型中同时抽取事件触发词和事件元素，如 JRNN 模型。联合抽取可以避免分阶段导致的误差传播，但是联合抽取会导致模型更加复杂，遇到一个句子中出现多个事件触发词或一个事件元素有多种元素角色时很难解决。

1.3 论文研究内容

本文基于一个军事百科文本，创建一个同时包含实体节点和事件节点的图谱，使得图谱能够在检索方面提供更多信息。本文主要工作是图谱的事件抽取部分。根据百科文本内容，数据集中标注了事件触发词，实体，以及实体和事件触发词之间的元素角色，实体之间的关系。同时本文观察军事相关的事件触发词和实体的特征，发掘事件触发词和事件元素实体之间的联系，贴合文本内容需求，建立符合实际的事件抽取模型，并构建事件关系抽取系统。

本文的主要目的是使用神经网络模型抽取事件触发词，事件元素实体，以及事件触发词和事件元素实体之间的元素角色，最终使用这些神经网络模型构建事件关系抽取系统。

首先，本文建立一个军事百科文本内容的图谱数据集。因为目标是军事领域，所以首先爬取军事领域相关的文本。根据百科文本内容及目标军事领域，定义实体和事件等相关标签，进行数据标注。在从标注工具中导出数据之后，又进行了数据清洗，格式变换等处理，使得数据格式易读易懂，能够轻松转换成神经网络模型输入需要的格式。同时，在数据后处理之后，校验了数据的某些信息，例如是否有重复句子，数据集中的冗余信息的正确性，实体等的标注是否完善等。对于校验时发现的问题，进行各种处理：重复句子比较各种标签之后进行合并，文本中回车的问题无法明确将文本舍弃，冗余信息发现问题通过人工进行处理。

其次，本文建立了事件元素实体抽取模型和事件触发词抽取模型。在数据集中，事件元素作为实体存在，而且在最终图谱中作为实体节点存在，不能和其他事件抽取模型一样只获得其位置和元素角色，还要获得其实体类型，所以需要将事件元素作为实体先抽取出来。同时经过观察，军事专业相关的实体的后缀词范围较小容易统计，本文将其

作为后缀特征加入神经网络模型中，提高模型的抽取效果。事件触发词模型则依赖于事件元素实体的抽取结果，因为事件元素实体和事件触发词之间具有一定的语义相关性，所以本文利用事件元素实体信息，通过注意力机制加强事件触发词的抽取效果。

然后，本文建立了一个元素角色抽取模型。在一个事件中，有一个事件触发词和多个事件元素实体，事件触发词和事件元素之间的关系被称为元素角色。一般的关系抽取模型一次输入两个实体的位置信息，抽取两个实体之间的关系。但是这样对于本文来说效率低下。本文尝试通过借鉴双仿射深度文法依赖树的构建方法，一次输入一个事件触发词和句子中所有的事件元素实体，抽取它们之间的元素角色信息，这样能够大大提升训练和预测效率。同时考虑数据集在标注的时候，定义了事件元素实体类别、事件触发词类别和元素角色之间具有严格的约束关系，将这种约束关系加入到神经网络模型中，提高模型的抽取效果。

最后，本文基于上述的模型和算法，以及通过事件关系连词构建模板，构建了一套军事事件关系抽取系统。最终，上面的各个模型是可以应用于军事图谱构建的。本文构建的军事事件关系抽取系统，应用上述已经实验证明有效的模型和算法，使用抽取结果准确率高的模板方案抽取事件关系，最终能够得到高质量的军事事件关系数据。

1.4 本文的组织结构

本文标注了军事百科图谱数据集，并实现了图谱构建中的事件抽取部分，并进行了模型效果评价。本文的论文组织结构如下：

第一章：首先介绍了事件抽取的研究背景和意义。接下来，总结了国内外对于事件抽取的研究现状及分析，按照技术发展介绍了三类事件抽取方法：基于模板匹配的方法，基于特征的机器学习方法，基于神经网络的方法。之后简单介绍本文的研究内容。

第二章：首先介绍了本文所用的数据集的详细情况，以及本文的事件抽取任务的细节。然后详细地介绍了事件抽取方法中涉及到的相关模型和理论，包括了循环神经网络，注意力机制，词向量等。最后介绍了事件抽取任务的性能评价指标。

第三章：首先介绍了事件触发词抽取和事件元素实体抽取的任务定义与内容。接下来描述了事件元素实体抽取的模型总体结构，详细描述了事件元素实体的后缀特征的统计学意义，后缀特征向量的生成方式，以及将向量输入到模型的方法。然后介绍了事件触发词抽取的模型总体结构，介绍了基于事件元素实体的注意力机制。最后通过多个对比实验证明了本文的事件触发词抽取模型和事件元素实体抽取模型的效果。

第四章：首先介绍了元素角色抽取的任务定义与内容。然后详细描述了元素角色抽取的模型，包括双仿射深度文法依赖树抽取的角色分类器的使用，掩盖矩阵的生成以及

在模型中的使用。最后通过对比实验证实了掩盖矩阵对于模型的提升，以及通过与事件抽取中标志性的模型进行对比，证明本文事件抽取模型的性能。

第五章：通过第三章和第四章中的模型和算法抽取事件，和基于关系连词的事件关系抽取模板，构建事件关系抽取系统。首先简单介绍了事件关系抽取系统的流程设计思路。其中使用事件关系连词对句子进行过滤，降低后续模块计算压力。然后逐个模块进行剖析，讲解事件关系抽取系统。

2 理论基础与相关模型

2.1 事件抽取和数据集简介

2.1.1 数据集简介

本文首先从头条百科中爬取军事装备相关词条，将词条内的文本经过整理后，按照词条内容标注了一个数据集。首先本文从头条百科中爬取军事武器相关的词条，然后将原始语料经过清洗和分句之后，整理出 3573 条句子。其中节点包括实体和事件触发词，同时标注了相应的实体类型和事件类型；边包括实体与实体之间的关系及其类型，以及事件触发词和实体（事件元素）之间的关系（元素角色）。

下面介绍一下数据集中与事件抽取相关的节点和边的标签类型，以及之间的相互约束关系。首先数据集标注了四类事件触发词，包括参战事件，作战事件，移动事件，以及部署事件；以及七种实体类型，装备实体，时间实体，地点实体，数值实体，任务实体，单位实体，战争名称实体。其中装备实体指各种武器装备，包括飞机舰船，导弹鱼雷等；时间实体指各种时间，包括时间点或时间段；地点实体与传统的地点实体一致；数值实体指的是各种数量以及后面的单位；任务实体指各种作战任务；单位实体指的是各种作战单位，例如陆海空军，或者某连某排；战争名称实体则指的是例如中途岛海战，第二次世界大战等类似的实体。其中参战事件指参加战斗，具体的触发词包括参加，参战，参与等；作战事件指各种作战行动，包括侦察，轰炸，击沉等；移动事件指各种作战单位的地点转移，包括重返，起飞，到达等；部署事件指命令作战单位到指定地点，包括部署，派遣等。

基于上文标注的各种结点，本文标注了各种关系。本文在事件之间标注了顺承和组成关系。不像实体之间的关系多种多样，不同种类的实体对之间有不同的关系，相同种类的实体对之间也可能有几种关系，事件之间的关系种类比较固定，例如顺承，组成，因果，条件等。本文同样标注了实体之间的关系。作战单位和装备之间有配置关系，装备之间有装配关系。最后，事件触发词和实体之间具有元素角色一样的关系。所有事件触发词都和单位实体，装备实体，任务实体，地点实体，事件实体有关系，但是具体的关系可能有不同。例如地点实体与参战事件，作战事件和部署事件之间就是简单的事件发生地的地点关系，而对于移动事件，有出发地点和抵达地点两种关系；作战事件的目标关系，目标可以是装备实体、作战单位实体、地点实体。而且在定义数据集标签时，规定了事件元素实体类型和元素角色之间具有约束，具体十分复杂，表格如下：

表 2.1 doWar（作战）事件元素角色和实体类型之间关系

Tab. 2.1 Relationship between doWar event argument and entity type

元素角色 实体类型	目标	使用	参加	时间	地点	完成
装备	√	√	×	×	×	×
时间	×	×	×	√	×	×
地点	√	×	×	×	√	×
数值	×	×	×	×	×	×
任务	×	×	×	×	×	√
单位	√	√	√	×	×	×
战争名称	×	×	×	×	×	×

表 2.2 arrange（部署）事件元素角色和实体类型之间关系

Tab. 2.2 Relationship between arrange event argument and entity type

元素角色 实体类型	地点	使用	参加	时间	完成
装备	×	√	×	×	×
时间	×	×	×	√	×
地点	√	×	×	×	×
数值	×	×	×	×	×
任务	×	×	×	×	√
单位	×	×	√	×	×
战争名称	×	×	×	×	×

表 2.3 move（移动）事件元素角色和实体类型之间关系

Tab. 2.3 Relationship between move event argument and entity type

元素角色 实体类型	离开	使用	到达	时间	完成	参加
装备	×	√	×	×	×	×
时间	×	×	×	√	×	×
地点	√	×	√	×	×	×
数值	×	×	×	×	×	×
任务	×	×	×	×	√	×
单位	×	×	×	×	×	√
战争名称	×	×	×	×	×	×

表 2.4 joinWar（参战）事件元素角色和实体类型之间关系
Tab. 2.4 Relationship between joinWar event argument and entity type

元素角色 实体类型	时间	使用	参加	战争名称	完成	地点
装备	×	√	×	×	×	×
时间	√	×	×	×	×	×
地点	×	×	×	×	×	√
数值	×	×	×	×	×	×
任务	×	×	×	×	√	×
单位	×	×	√	×	×	×
战争名称	×	×	×	√	×	×

以上即数据集标签的所有情况。数据集中与事件相关的有 4 种事件触发词，作为事件元素的 7 种实体，以及对应每种事件触发词的若干种元素角色，都是本文需要抽取的对象。而且，同一类事件元素实体，可能在同一类的事件触发词下，可能具有不同的元素角色。复杂的映射关系，需要模型具有一定的兼容性；同时这种复杂的关系也可以作为提升模型效果的重要信息。

2.1.2 事件抽取任务

事件抽取，即采用某些算法，将非结构化的自然语言数据中的事件信息抽取出来，以结构化的形式展现出来。事件抽取任务包含事件触发词抽取和事件元素抽取。ACE 会议给出事件抽取中的相关术语如下：

事件描述（Event Mention）：指描述事件发生的句子或短语，包含事件触发词和事件元素。

事件类型：指事件对应的类别，ace2005 数据集中定义了 33 中事件类型，本文的数据集中定义了 4 种事件类型。

事件触发词：事件描述中最能代表事件发生的词，多为动词或名词，是一个事件的核心，一个事件只有一个事件触发词。

事件元素：事件的参与者，是组成事件的核心部分，与事件触发词一起构成了事件的整个框架。事件元素主要由实体、时间和属性值等表达完整语义的细粒度单位组成。

元素角色：事件元素与事件之间的语义关系，也就是事件元素在相应的事件中扮演了什么角色。

事件抽取任务，就是从非结构化的文本中，抽取上述的事件触发词，事件类型，事件元素，元素角色 4 项，而这 4 项共同组成了事件描述。在本文的军事图谱数据集中，

事件元素在句子中作为实体存在，且标注了相应的实体类型。而本文的任务，即抽取相应的实体，以及事件触发词，事件类型，事件元素，元素角色。

2.2 深度学习相关技术

自从 GPU 作为深度学习的计算工具，解决了深度学习计算量的问题之后，深度学习得到极大发展。深度学习神经网络技术，是从人脑的神经元结构得到启发，模拟神经元的输入，数据处理，输出，使用简单的数据处理和复杂的模型结构作为机器学习模型，获得远超普通的机器学习算法的性能。其中，通过张量来保存需要输入的数据，通过简单的加减乘除模拟数据处理，通过激活函数来使神经网络获得更多变化，通过优化器和误差反向传播算法减小算法的损失函数值来优化神经网络模型，最终结构合理的神经网络会通过训练获得很好的性能。

经过发展，神经网络出现了各种结构，例如简单的全连接层，多应用于图像的卷积神经网络 CNN，应用于序列的循环神经网络 RNN，以及注意力机制层。新的结构不断出现，旧的结构也出现不同的变种，因为虽然理论证明只要模型够复杂就能够拟合各种函数，但只有更符合实际问题的神经网络模型结构，才会有更高的泛化能力。

2.2.1 循环神经网络

在最开始的时候，神经网络主要是全连接层，即层与层之间是全连接的，层之间的节点是无连接的。但是如果遇到了例如与序列相关的问题，比如预测一个句子中当前词的词性，其结果一般与之前的词相关，因为句子中的词不是独立的，就会遇到问题。循环神经网络就是为了使句子之间的结点能够互相传递信息而出现的。

因为在很多问题中，在当前节点需要之前节点的信息，循环神经网络（Recurrent Neural Network, RNN）出现了。在循环神经网络中，一个节点会有两项输入，即之前节点的隐藏层信息，以及当前节点之前的层输出；输出也有两项，即当前节点的隐藏层信息，以及当前节点当前的层的输出。循环神经网络也可以视为按照序列顺序链式链接的递归神经网络，能够将之前节点的隐藏层信息与当前时刻输入词的信息进行结合，影响后续节点的输出，从而形成长期依赖。这样当前节点利用之前节点的信息，形成记忆性，对于序列处理具有一定优势，例如文本生成，机器翻译，尤其在自然语言处理中，通过之前的字或词，能够更容易预测当前的字或词的信息，例如词性等，也被用于各种时间序列的预测。

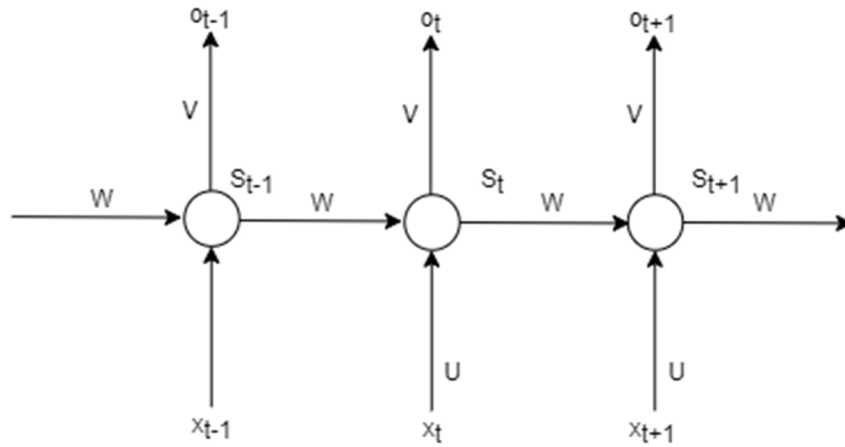


图 2.1 循环神经网络

Fig. 2.1 The RNN network

循环神经网络的详细结构如图所示。整体可以看成三组数据之间的运算，即输入 x ，隐藏层状态 s ，输出 o 。计算公式如下：

$$s_t = f(U * x_t + W * s_{t-1} + b) \quad (2.1)$$

$$o_t = \sigma(V * s_t) \quad (2.2)$$

其中 s_t 为当前时刻隐藏层状态， s_{t-1} 为上一时刻隐藏层状态， x_t 为当前时刻的输入， o_t 为当前时刻的输出， f 、 σ 为激活函数，其余为神经网络的参数。可见当前时刻隐藏层状态 s_t 和当前时刻输出 o_t 由上一时刻隐藏层状态 s_{t-1} 以及当前时刻的输入 x_t 共同决定。其中隐藏层状态传递之前节点的信息，当前时刻的输入传递当前节点的信息，使得 RNN 有能够按照序列顺序进行预测的能力。

2.2.2 长短期记忆网络

上节讲述的 RNN 已经能够获取之前时刻的信息，但是还是有一定的不足之处。当神经网络在计算误差反向传播的时候，循环神经网络中误差会沿着隐藏层状态反向传播；而在不同时刻中，RNN 使用的参数都是相同的，而且运算时都是乘法，这也是导致梯度消失和梯度爆炸的原因。而梯度消失和梯度爆炸会导致神经网络无法捕捉长距离依赖。这也是 LSTM 所解决的难点^[24]。长短期记忆网络（LSTM，Long Short-Term Memory）是 RNN 的一种变体，主要有两点改进：一是使用细胞状态来记忆存储，二是改进了隐藏状态的计算方法，避免了长期依赖问题。

LSTM 的结构如图。整个 LSTM 的结构围绕着细胞状态的更新，其中遗忘门控制使上一时刻的细胞状态丢弃某些信息，输入门控制本时刻的输入中哪些数据是需要更新的，输出门控制哪些信息是需要输出的。

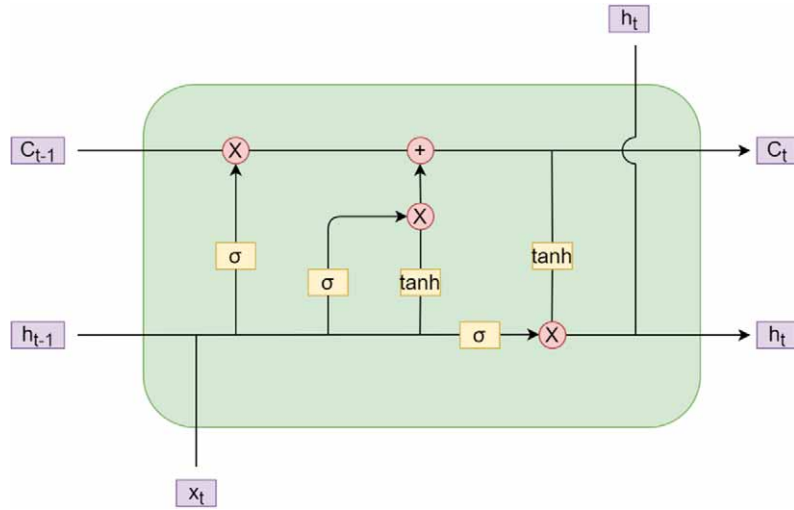


图 2.2 长短期记忆网络

Fig. 2.2 The LSTM network

LSTM 的第一步是通过遗忘门来决定需要从细胞状态中丢弃什么信息。此处根据本时刻的输入 x_t 以及上一时刻的隐藏状态 h_{t-1} ，经过线性运算之后，再经过 sigmoid 激活函数得到一个 0 到 1 的数值，用来控制上一时刻细胞状态的保留，0 代表全部舍弃，1 代表全部保留。

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (2.3)$$

LSTM 的第二步是通过输入门来得出本时刻需要向细胞状态中输入的候选值 \tilde{C}_t ，以及控制这个输入的值的比例 i_t 。这个比例的计算方法与遗忘门的类似，不过里面的参数换了一套而已，同样是一个 0 到 1 的数值。而需要向细胞状态输入的候选值 \tilde{C}_t 也是类似的计算方法，只是激活函数换成了 tanh。

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (2.4)$$

$$\tilde{C}_t = \tanh(W_C * [h_{t-1}, x_t] + b_C) \quad (2.5)$$

然后就是更新当前时刻的细胞状态。通过之前遗忘门的输出 f_t ，用来控制上一时刻的细胞状态保留 C_{t-1} 的信息量，输入门的 i_t 用来控制新输入的 \tilde{C}_t 的输入比例，而细胞状态的更新通过加法运算，避免了由于乘法运算导致的长期依赖导致的梯度消失和梯度爆炸问题。

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (2.6)$$

最后，通过输出门来得到需要输出的信息。其中当前时刻的隐藏状态更新与当前的输出值和细胞状态有关。

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (2.7)$$

$$h_t = o_t * \tanh(C_t) \quad (2.8)$$

LSTM 通过遗忘门,输入门,输出门三个门控单元来控制核心信息细胞状态的丢弃,更新和输出,并同过加法而不是乘法完成上述步骤,来避免由于长期依赖导致的梯度爆炸或梯度消失问题。但与 RNN 相比计算步骤增多,计算量提高,时间成本变高,神经网络的参数量提高,不过模型效果相对于 RNN 有了很大提升。

2.2.3 注意力机制

注意力机制的灵感来源于人类的视觉。人们在通过眼睛观察东西的时候,会根据需求注意特定的某处,忽视其他地方。所以人们根据以上原理,发明了神经网络中的注意力机制。注意力机制的核心在于,通过计算得出每个分量的权重,通过加和得到最终的输出,而其中的权重,就是对于各处注意力分配的多少。这也是注意力机制缓解神经网络模型复杂度的方法,即不需要将所有的信息都输入到后续过程中进行计算,只需要从输入中选择一些与任务相关的信息输入到后续过程中即可。

注意力机制不仅应用在视觉图像领域^[26],也在许多其他领域得到广泛应用,例如自然语言处理^[27],推荐系统^[28],语音识别^[29]等。在自然语言处理中,bert 词向量的出现使得注意力机制获得了更多的关注。同时经过多年的发展,在注意力机制也因为各种各样的问题变换出各种各样的形式。

最开始注意力机制在图像领域中应用。原始的注意力机制中权重值为 0 或 1,被称为 Hard attention^[30]。后来人们将权重值调节为 0 到 1 的概率分布,通过反向传播进行参数优化,被称为 Soft attention。

Scaled dot-product attention,缩放点乘注意力机制,属于比较简单而基础的注意力机制形式^[31]。首先对应序列中每一个节点,都有一个对应的 q、k、v 向量,所有节点的向量一起分别组成了 Q、K、V 矩阵,其中 V 矩阵就是我们的注意力需要投射的对象。通过 QK^T ,基本得出了对于输出序列中每个节点对应输入的 V 中每个节点的注意力权重。最后经过除以 $\sqrt{d_k}$ 去除向量的尺寸因素,通过 softmax 进行向量归一化,最后得到注意力机制对应 V 中每个节点的注意力权重。最后将权重与 V 矩阵相乘,得出每个节点位置的输出。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.9)$$

缩放点乘注意力机制,简洁地表达了注意力机制的主旨,但是其中缺少可以训练的参数。虽然相对于 RNN 减少了参数量,但是无法更进一步加快计算速度。而更深的神

神经网络能够以更大的参数量，使模型获得更好的效果；但是更深的模型需要更快的计算速度，减少训练成本。

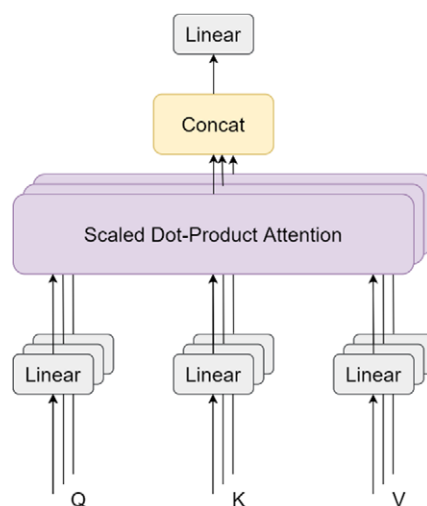


图 2.3 多头注意力机制

Fig. 2.3 The multi-head attention

Multi-head attention，多头注意力机制^[32]，比 Scaled dot-product attention 更进一步。首先，将原本的 Q 、 K 、 V 通过线性变换得到多组不同的结果，同时输入原本的 Scaled dot-product attention 中，将最后得到的结果连接到一起，通过线性变换调整输出维度。好处在于可以同时计算多组注意力机制，加快计算速度方便并行计算，获取更充分的信息；同时相对循环神经网络及其变体而言，参数量更少，训练速度更快。

上面只是比较典型的注意力机制的形式。对于具体的问题，还可以发展出更多的使用方法。比如将 Q 矩阵的来源更换，可以获得与来源相关的注意力机制；可以控制注意力的来源范围，得到局部注意力或者全局注意力。最核心的是通过权重来分配注意力。

2.3 词向量

最开始基于模板抽取的时候，不会有使用向量代替字或词的需求。后来基于特征的机器学习的时代，输入特征时不用考虑向模型中输入字或词。而当模型需要输入字或词的时候，最开始使用了独热码的形式。但是独热码的形式下，两个向量之间缺少关联，向量中特征很稀疏；由于一个语言中的词的数量以万计数，得到的向量的长度很长。为了使用向量能够更好表达字或词的含义，人们发明了 word2vec 词向量、bert 词向量等词向量。

2.3.1 word2vec 词向量

使用向量代替一个词输入到模型中，也就是将词放入一个多维空间中，也可以称之为词嵌入。但是如何获得一个好的词向量？这时有人提出来，可以使用一个词周围的词来定义一个词的含义。

word2vec 就是在这种思想下产生的。于 2013 年，Google 开源推出了 word2vec 词向量^[33]。它分为两种不同的模型，一种是 CBOW (Continuous Bag-of-Words Model) 连续词袋模型，通过周围的词来预测中间的词，一种是 Skip-gram (Continuous Skip-gram Model) 跳词模型，给定一个词来预测周围的词。在训练的时候，需要维持一个矩阵，给对应每一个词随机初始化一个向量。然后确定一个窗口，确定中间的词和周围的词。如果是 CBOW 模型，则将周围的词对应的词向量加和平均，然后通过优化器使中间的词的词向量更接近周围词的词向量的平均值；如果是 Skip-gram 模型，则通过优化器使周围的词的词向量更接近中间的词的词向量。通过优化方法，使得在句子中，一个词的词向量与其周围的词的词向量接近，包含了通过一个词周围的词定义一个词的思想。

当然，在 2013 年，计算资源还不如现在充足，于是人们发明了许多用于加速 word2vec 训练的方法，比如负采样^[34]，层序 softmax 等方法。word2vec 仍是当今自然语言处理中重要的词嵌入技术，其有占用空间小，运行速度快等优点。

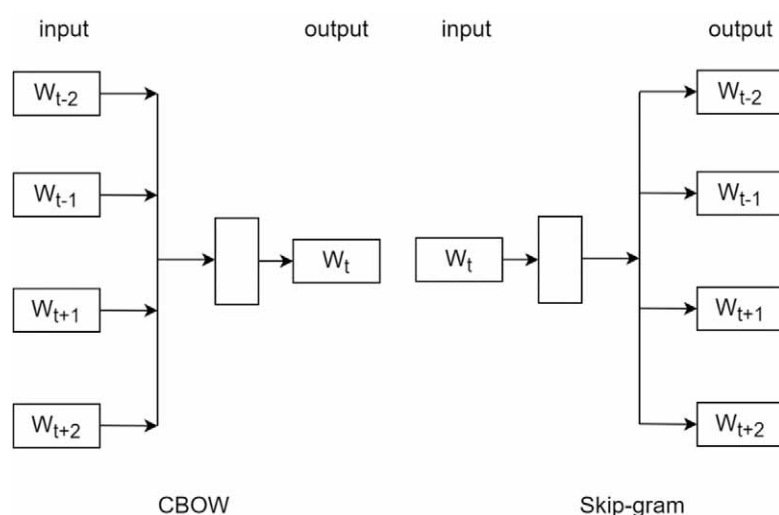


图 2.4 word2vec 模型结构图

Fig. 2.4 The structure of word2vec

word2vec 开辟了崭新的词输入机器学习模型的方式之后，人们开始观察这种带来模型效果提升的词嵌入方法。比如，人们发现，词向量之间具有某种可计算的关系，比如“皇帝”向量减去“皇后”向量约等于“男人”向量减去“女人”向量；相似的词向

量通常聚集到一起，比如代表数字的词。同时，人们还发现，多义词的词向量是多种含义代表的向量的加和。于是，人们又将其中一些发现作为训练一个高质量的 word2vec 词向量或者其他静态词向量的标志^[35]。

2.3.2 bert 词向量

word2vec 词向量开启了词嵌入的大门，但是根据人们对于词向量的不断探索观察，还是发现了一些不足之处。后来人们提出了许多不同的词嵌入方式。其中人们把一个词对应一个固定向量的成为静态词向量，把获取词向量时需要输入整个句子然后通过模型动态生成的成为动态词向量。动态词向量能够根据词的上下文语境生成词向量，能够消除多义词的影响。

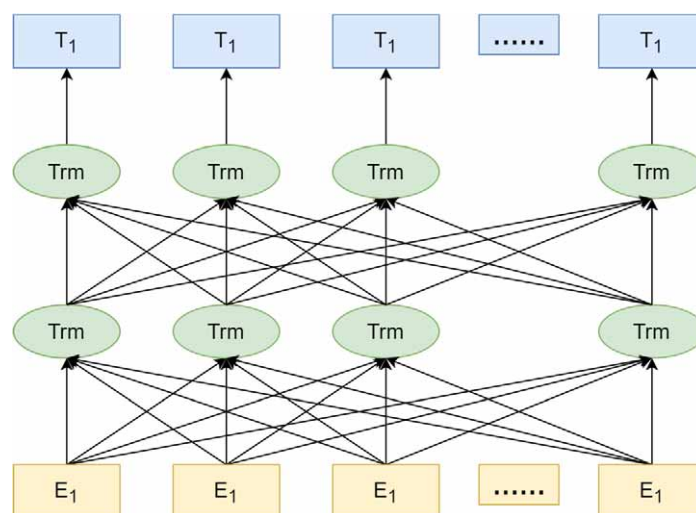


图 2.5 bert 模型结构

Fig. 2.5 The structure of bert

动态词向量提出之前，人们又提出了 glove^[36]等各种静态字向量。动态词向量提出之后，人们开始使用不同的神经网络模型来实现效果更好的词向量模型。例如 ELMO (Embeddings from Language Models)^[37]使用 3 层双向 LSTM 模型生成动态词向量。

bert 词向量^[38]比这些动态词向量更进一步。它使用了多层基于自注意力机制的 transformer 来构建一个很深的神经网络，选择完形填空和句对预测来作为预训练任务。在训练数据上，随机选取一些词，80%的概率 mask 掉它们，10%用其他词随机替换，10%保留原词。通过神经网络上的创新和工程上的改变成功训练出一个很深的自然语言模型并取得比较好的效果。

当然，模型越深，层数越多，参数量就越多，训练起来就越消耗时间，存储起来也更消耗空间。bert 标准版本有约 1 亿的参数量，大型版本有 3 亿多参数量，google 训练时使用了 16 个 TPU 集群训练几天的时间。这种成本毫无疑问不是普通的自然语言处理研究者所能承受的，于是 google 将 bert 模型放到网络上供人们自由下载。

2.4 性能评价指标

事件抽取任务主要分为两部分：事件触发词抽取和事件元素抽取。因此，对于任意一个事件单元，本文从事件触发词识别，事件触发词分类，事件元素识别，事件元素分类四个方面对事件抽取任务进行综合评价：

事件触发词识别 (Trigger Identification)：如果识别出来的事件触发词位置和数据中标注的位置相同，则事件触发词识别正确，属于二分类任务。

事件触发词分类 (Trigger Classification)：如果抽取的触发词的位置和触发词的事件类型和数据集中标注的位置和类型都相同，则该事件触发词分类正确，对应的事件分类正确，是事件的多分类任务。

事件元素识别 (Argument Identification)：在已经抽取完事件触发词的条件下，对事件元素的位置进行识别。如果识别出事件元素的位置和数据集中标注的事件元素的位置相同，则该事件元素识别正确，属于二分类任务。

事件元素分类 (Argument Classification)：在已经抽取完事件触发词的条件下，对事件元素的位置和类别进行识别。如果识别出的事件元素的位置和类别与数据集中标注的事件元素的位置和类别相同，则该事件元素分类正确，属于多分类任务。

在自然语言处理任务中，序列标注类的任务的评价指标主要有精确率 (Precision, P)，召回率 (Recall, R) 和 F 值 (F-score, F) 来进行评价，事件抽取的四个任务同样采用这三项指标进行评价。

其中精确率是指，在事件抽取结果中，正确的事件触发词和事件元素占样本总数的比例；召回率是指，在事件抽取结果中，正确的事件触发词和事件元素占数据集中所有正确的事件触发词和事件元素的比例。而 F 值则是综合考虑精确率和召回率，常用的 F-1 值是精确率和召回率的调和平均数，其他的 F 值是不同的加权下的精确率和召回率的调和平均数。F 值是一种对于模型的总体评价。精确率，召回率和 F 值的计算公式如下：

$$P = \frac{TP}{TP+FP} \quad (2.10)$$

$$R = \frac{TP}{TP+FN} \quad (2.11)$$

$$F = \frac{2 \times P \times R}{P+R} \quad (2.12)$$

其中，TP 代表真正例（True Positive），代表模型预测结果中被正确预测的正例样本个数；FP 代表假正例（False Positive），代表模型预测结果中被正确预测的假例样本个数；FN 代表假反例（False Negative），代表模型预测结果中被错误预测的正例样本个数。

以上只是对于二分类任务的计算方式。对于多分类任务，为了评价模型在数据集上所有种类标签的性能，有宏平均和微平均两种评价指标。微平均是首先建立全聚德混淆矩阵，然后计算总体的精确率和召回率，然后计算 F-值，而不是计算每一类的精确率和召回率。宏平均则是先计算每一类的精确率和召回率，然后求得精确率和召回率的平均值，通过这两个平均值再计算 F-值。本文采用微平均的评价方法，其公式如下，其中 n 表示多分类中的类别总数：

$$Micro_P = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FP_i} \quad (2.13)$$

$$Micro_R = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FN_i} \quad (2.14)$$

$$Micro_F = \frac{2 * Micro_P * Micro_R}{Micro_P + Micro_R} \quad (2.15)$$

2.5 本章小结

本章介绍了事件抽取使用的模型、方法和性能指标。首先，介绍了本文所使用的数据集的情况，包括文本来源，标签种类，标注时的某些约定。接下来，本文介绍了事件抽取任务的详细定义。事件抽取即抽取事件触发词，事件元素，事件类型，元素角色四者。然后介绍了事件抽取主要使用的模型，包括循环神经网络，长短期注意力机制 LSTM，门控循环单元 GRU，以及注意力机制模型。其中循环神经网络，长短期注意力机制 LSTM，门控循环单元 GRU 三者常用于自然语言处理中，能够从前文中获得信息来对后文进行预测。最后的注意力机制模型，则能够获取序列中远处的信息，能够适应模型的不同需求设计出各种不同的形式。然后，本文介绍了在自然语言处理中将文本输入到神经网络模型的重要方法，词向量。本文首先介绍了静态词向量中的典型模型 word2vec，包括其出现背景，核心思想和训练方法。接下来本文介绍了 bert 词向量。作为动态词向量，bert 词向量能够为多义词按照语境提供不同含义的词向量；使用了基于注意力机制的 transformer 模型，使得 bert 能够有效利用远处的信息。作为典型的动态词向量，bert 至今仍被大量使用。最后本章介绍了事件抽取的性能指标。事件抽取的性能指标包括四个任务，包括事件触发词识别，事件触发词分类，事件元素识别，事件元素分类；每个任务都有三项指标，包括准确率，召回率和 F-值，作为整个事件抽取的评价指标。

3 事件触发词和事件元素实体抽取

事件触发词和事件元素是事件的重要组成部分。本文数据集标注了装备实体，任务实体，单位实体，战争名称实体等七种实体，以及作战事件，参战事件，移动事件，部署事件四种事件触发词，以及这些实体和事件触发词之间的关系（元素角色），标注内容如图。任何事件的发生，一定有空间和时间的坐标，在数据集中体现为地点实体和时间实体。其余四类实体装备实体，任务实体，单位实体，战争名称实体均与军事密切相关，专业性比较强。所以我们可以先抽取事件元素实体，然后利用这四种实体作为注意力机制的输入，来提高事件触发词的抽取效果。同时，我们观察得知，这四种实体的后缀部分取词范围比较小，可以做成相应的词库，然后将这种后缀信息输入神经网络模型中，可以提高这些实体的抽取效果。

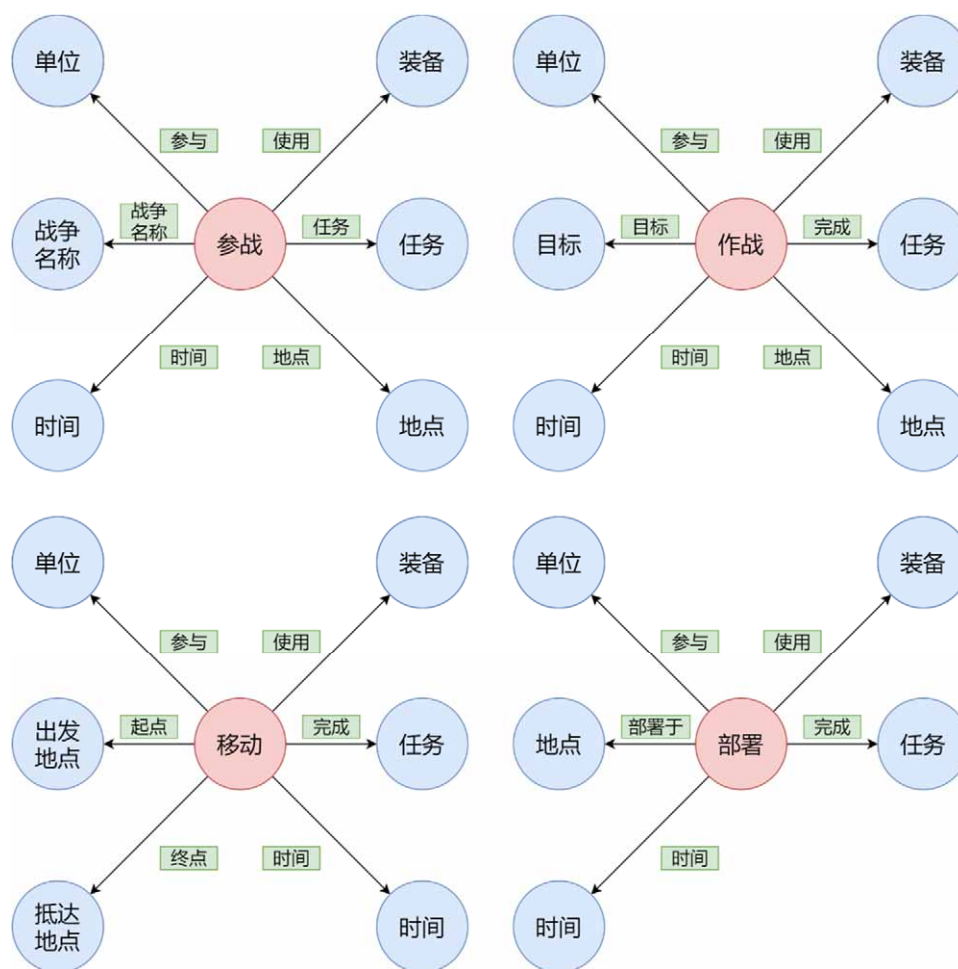


图 3.1 数据集标签示意图

Fig. 3.1 diagram of dataset label

3.1 事件触发词和事件元素实体抽取问题描述

事件触发词和事件元素是组成一个事件的基本单位。如前文所述，数据集中标注了四类事件触发词，以及七类事件元素实体。例如，在“9月16日晚舰队终于出海训练”中，标注了参战事件“出海”，时间实体“9月16日”，单位实体“舰队”，任务实体“训练”。本章的事件触发词和事件元素实体抽取就是将这些“出海”、“舰队”和“训练”抽取出来，并识别它们是事件触发词还是事件元素实体，以及对应的类型。

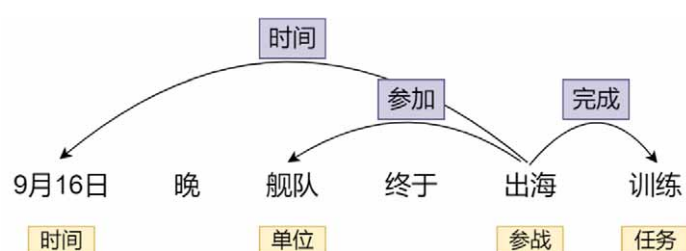


图 3.2 数据标注示例

Fig. 3.2 An example of dataset

而通过上节的描述可知，本文将事件触发词和事件元素实体抽取分成了两部分，首先抽取事件元素实体，然后再使用事件元素实体及注意力机制抽取余下的事件触发词。

3.2 事件元素实体抽取

经过上节的描述，本部分抽取的是七种事件元素实体。观察数据集可发现，装备实体，任务实体，单位实体，战争名称实体四种实体的结尾部分词范围比较小，可以做成相应的词库；大部分实体的长度在 5 个字左右，可以通过注意力机制来更好地获取远端信息；同时实体识别经常作为序列标注任务，可以利用经典的 Bi-GRU-CRF（Bi-GRU: Bi-Gated Recurrent Unit，双向门控循环单元；CRF: conditional random field，条件随机场）来作为基础框架，通过上面的几点需求来改造模型。

3.2.1 模型总体结构

模型总体以 Bi-GRU-CRF 为基础，通过序列标注的方式来抽取七种实体。对于字向量，选用 bert 字向量。装备实体，任务实体，单位实体，战争名称实体四种实体的结尾部分词范围做出来的对应的四个词库，可以生成后缀特征。Bert 字向量和后缀特征连接在一起作为模型的输入。接下来数据经过 Bi-GRU 模型，获取前后文的相关信息。同时观察实体的长度比较长，大多数在 5 个字左右，接下来通过自注意力机制能够有效接受

远端的信息。然后再与重复多次的后缀特征进行连接。最后使用 CRF，能够学习到序列标注的 BIO 标签之间的隐含关系，提高最终的模型效果。

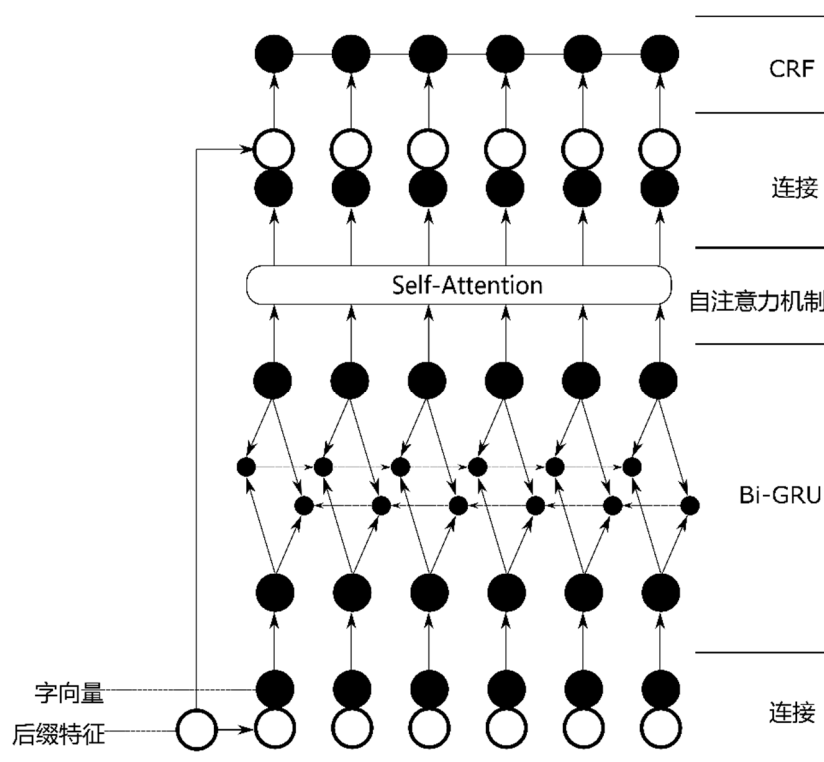


图 3.3 事件元素实体抽取模型结构图

Fig. 3.3 The structure of event argument entity extraction model

3.2.2 bert 字向量及后缀特征

本模型的输入由两个部分组成，bert 字向量以及后缀特征。bert 字向量已经提出多年，在 google 之后人们又在 bert 的基础上不断改进，而哈工大提出的全词掩盖，在掩盖的时候不是随机掩盖字，而是掩盖整个词，而中文的含义是以词为单位的，所以得出的 bert 字向量有更高的性能。

军事专业相关的词汇，大部分都是修饰词+后缀组成的，例如哈曼号+驱逐舰，20 毫米+火炮，第 14 舰载机+联队等，这是因为专业性的实体词汇一般使用常见的后缀表述实体的大类，然后使用实体的某些属性特征或特定称号作为修饰词。同时由于这些后缀词数量有限，容易统计，在对应的军事特定领域内具有一定的泛用性。这是由于专业词汇受众面小，做出的让初学者或外行也能易懂的中文构词方法。而本文标注的军事专业相关的实体，如装备实体，任务实体，单位实体，战争名称实体四种实体，就具有这种特征。

为了将这种后缀特征输入到神经网络模型中，首先需要建立对应的后缀词库。本文对于装备实体，任务实体，单位实体，战争名称实体四种实体，每种建立了一个后缀词库。同时，在使用后缀特征之前，还统计了一下这些词库与实体之间的契合程度，例如被词库匹配上的字有多大概率是对应实体的一部分，以及有多少实体的结尾能够被词库匹配。只有这两个统计指标足够高，这些信息输入模型之后才可能对效果有提升。

经统计得知，被词库匹配上的字有 92.4% 的概率是对应实体的一部分。而实体的结尾能够被词库匹配的概率如下表。

表 3.1 实体后缀被词库命中概率
Tab. 3.1 The probability of entity word hit by corpus

实体类型	equip (装备)	task (任务)	unit (单位)	warName (战争名称)
词库命中概率	0.607	0.717	0.871	0.886

可见，如果被词库匹配上有极大概率是对应实体的一部分，而也有较大部分的实体的结尾被词库命中，所以后缀特征是一种在统计学上有利于事件元素实体识别的信息。

在建立了后缀词库，验证了词库信息的有效性之后，下一步是将这种信息转化成能够输入到神经网络模型的形式。本文尝试将后缀信息转换成向量的形式。对于每一个词库，首先输入的句子匹配词库中的词，然后匹配中的字的位置为 1，其余位置为 0，则得到了每个字对应一个数字，四个词库就是长度为 4 的向量。这就是输入到神经网络模型的后缀特征向量。

表 3.2 后缀词库举例
Tab. 3.2 Some examples of suffix corpus

实体类型	装备实体	任务实体	单位实体	战争名称实体
后缀举例	战斗机	演习	联队	行动
	坦克	护航	民兵	演习
	舰只	作战	苏军	作战
	船只	侦查	船队	空战

如图 3.4，例如对于“美国号和珊瑚号航母战斗重返锡德拉湾，参加黄金峡谷行动”这句话。其中“航母”在装备实体的词库中可见，“行动”在战争名称的词库中可见，并假设四种实体类型的顺序为装备，任务，单位，战争，则生成的后缀特征中，航母的每个字对应的向量为[1, 0, 0, 0]，行动的每个字对应的向量为[0, 0, 0, 1]，其余字对

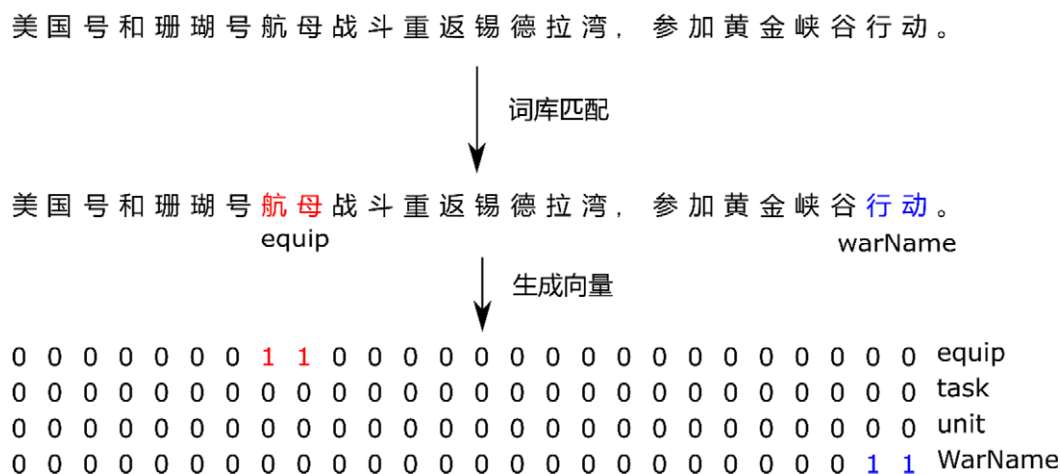


Fig. 3.4 The diagram of generation of the suffix feature vector

3.2.3 双向门控循环单元

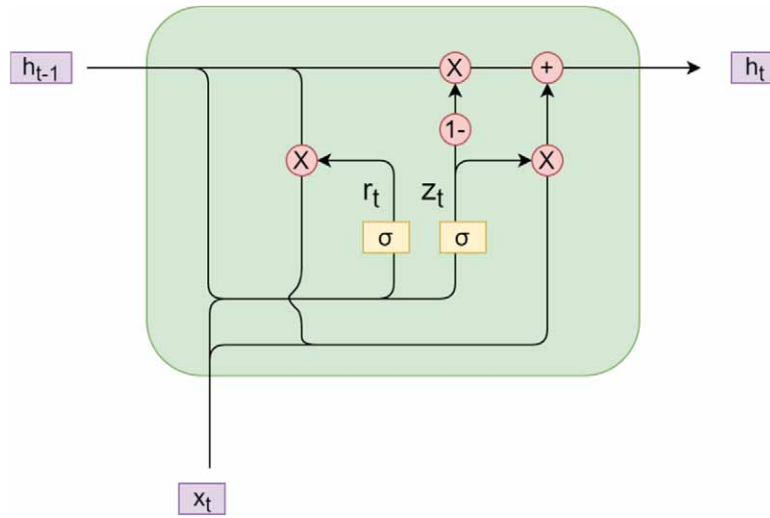


图 3.5 门控循环单元

Fig. 3.5 The GRU network

GRU 的重置门主要是得出需要新加入的信息。首先输入上一时刻的隐藏状态 h_{t-1} 以及当前时刻的输入 x_t ，经过线性变换和 sigmoid 激活函数得到 r_t 。然后通过重置门的结果，计算得到 \tilde{h}_t 。

$$r_t = \sigma(W_r * [h_{t-1}, x_t]) \quad (3.1)$$

$$\tilde{h}_t = \tanh(W * [r_t * h_{t-1}, x_t]) \quad (3.2)$$

更新门的计算过程与重置门相似，输入上一时刻的隐藏状态 h_{t-1} 以及当前时刻的输入 x_t ，经过线性变换和 sigmoid 激活函数得到 z_t 。与 LSTM 的更新不同，通过令 h_{t-1} 和 \tilde{h}_t 在更新中所乘的系数加和为一，减少计算所用的参数。

$$z_t = \sigma(W_z * [h_{t-1}, x_t]) \quad (3.3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (3.4)$$

GRU 将细胞状态和隐藏状态合并，通过令更新过程中的系数加和为 1 的约束条件减少门控单元数量，在摆脱长期依赖导致的梯度消失或梯度爆炸问题的同时，降低神经网络训练成本，虽然最终训练效果与 LSTM 相差不大，却能加快训练速度。

GRU 属于循环神经网络的一种变种，能够沿着序列前进方向接受来自前一时刻的信息不断得出结果。但是在人们阅读自然语言的时候，不仅仅是从前往后阅读，还要根据语境语法联系前文来理解真正的含义，所以逆向的循环神经网络也会提供一些重要的信息。将正向的 GRU 和逆向的 GRU 的输出直接连接起来，就得到了 Bi-GRU 模型，其计算方法如下：

$$Bi-GRU(x) = \text{concat}(\overrightarrow{GRU}(x), \overleftarrow{GRU}(x)) \quad (3.5)$$

3.2.4 自注意力机制

在深度学习中，注意力机制的广泛定义是，输入三个矩阵，Q (query)，K (key)，V (value)，通过 Q 查询 K 的值，生成结果中每个位置对于 V 中每个位置向量的权重 weight，然后通过权重 weight 和 V 来获得结果。期间的各种操作可以通过不同方式来实现。本文使用的自注意力机制，计算方式如下：

$$Q = W_Q x + b_Q \quad (3.6)$$

$$K = W_K x + b_K \quad (3.7)$$

$$V = W_V x + b_V \quad (3.8)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.9)$$

其中 x 为上一层的 Bi-GRU 的输出， W_Q 、 W_K 、 W_V 是线性变换的权重矩阵， b_Q 、 b_K 、 b_V 是线性变换的偏置向量。如果不经线性变换，Q、K、V 三个矩阵相同，则注意力机制就没有意义。所以我们需要通过线性变换使得这三个矩阵不同。自注意力输入简单，能够捕捉句子中不同字或词之间的相互关系，能够获取比循环神经网络更远的信息，提高对于数据集中较长实体的抽取能力。同时 Q、K、V 三者的线性变化的参数作为可训练参数，使得自注意力机制能够对数据更好地拟合。

3.2.5 CRF 条件随机场

条件随机场 (conditional random field, CRF)，是一种判别式概率模型。在深度学习的自然语言处理中，CRF 应用在序列标注任务中，通过训练能够得到一个状态转移矩阵，存储了序列标注中不同标签之间的转移概率。在预测的时候，使用维特比算法，得出概率最大的序列。

之所以需要 CRF，是因为在序列标注任务中，标签之间有一些依赖关系。例如在 BIO 标注方法中，B 标签和后面的 I 标签代表的实体类别应该一致，句子的开头应该是 B 标签或 O 标签，实体的开头应该是 B 标签而不是 I 标签。这些标签之间的前后依赖关系能够通过 CRF 来学到。

尤其在本文的军事图谱数据集中，事件元素实体的长度大部分都超过 5，导致 B 标签的数量远超 I 标签的数量；而实体标签占整个数据集中字数的比例较小。两个因素一起导致 B 标签占整个数据集中的比例较小，但是通过 CRF 的状态转移矩阵能够根据已经抽取的 I 标签推理出 B 标签，使得一个实体的 B、I 标签抽取完整，提高模型抽取效果。

3.2.6 损失函数

事件元素实体抽取的模型最后为 CRF，所以模型的损失函数也使用 CRF 的损失函数。损失函数的计算公式如下：

$$y = (y_1, y_2, y_3, \dots, y_n) \quad (3.10)$$

$$s(X, y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=1}^n P_{i, y_i} \quad (3.11)$$

$$loss = -s(X, y) + \log (\sum_{\tilde{y} \in Y_X} e^{s(X, \tilde{y})}) \quad (3.12)$$

其中 y 为需要预测的真实值， A 为 CRF 中需要的转移矩阵， P 为 CRF 之前的输出矩阵。第二个公式计算的 $s(X, y)$ 为需要预测的真实值的路径得分。然后通过 softmax 和对数方法，得到第三个公式，也就是损失函数的计算方法。

3.3 事件触发词抽取

本节的目的是抽取作战事件，参战事件，移动事件，部署事件四种事件触发词，得出每个事件触发词的位置及其类型。在上节中，使用了与军事专业领域相关的实体的后缀特征，抽取了事件元素实体。由于这些实体与军事专业领域密切相关，本文使用基于事件元素实体的注意力机制来抽取军事相关的事件触发词。

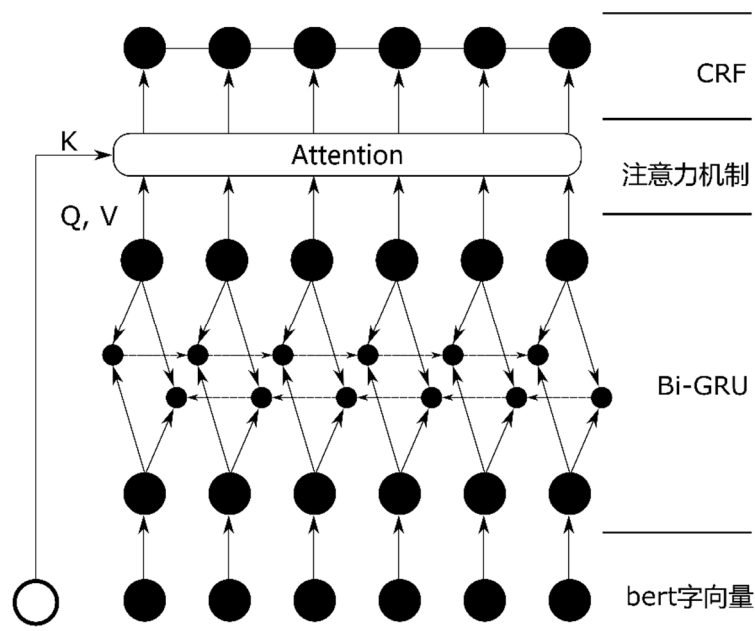


图 3.6 事件触发词抽取模型结构图

Fig. 3.6 The structure of event trigger extraction model

3.3.1 模型总体结构

总体模型以 Bi-GRU+CRF 为基础,通过序列标注的方式来抽取四种事件触发词。对于字向量,选用 bert 字向量。最初模型的输入仅有 bert 字向量,然后经过 Bi-GRU 神经网络模型提取上下文信息,来到注意力机制。注意力机制中,使用事件元素实体信息作为 query, Bi-GRU 的输出作为 key 和 value,通过事件元素实体的信息来提高事件触发词抽取的效果。最后经过线性变换和 CRF 模型,得到模型的最后输出。模型的最后仍是 CRF 模型,所以事件触发词抽取的损失函数和事件元素实体抽取的损失函数计算方法一致。

3.3.2 模型输入

模型最初的输入只有 bert 字向量。作为先进且应用广泛的动态字向量,使用 bert 可以为模型效果带来明显提升。bert 模型使用以自注意力机制为基础的 transformer 作为模型的主要框架,base 版模型有 12 层 transformer, large 版模型有 24 层 transformer,训练时需要大量显存和大量时间。同时也证明了,可以通过注意力机制设计一个很深的模型,达到比较好的效果。

3.3.3 基于事件元素实体的注意力机制及线性混合

在数据集标注的事件元素实体中,装备实体,任务实体,单位实体,战争名称实体四种实体一般是专业名词。同时,这些事件元素实体也与数据集中标注的事件触发词具有隐含的语义关系,例如装备实体标注了各种武器装备,与武器装备相关的事件,基本是作战训练,使用装备移动;单位实体标注了各种军事作战单位,让人们很容易联想到与作战,部署相关的事件中。但是事件元素实体和事件触发词之间并没有像上节的后缀特征一样直接的联系,于是本文采用了注意力机制的方案,将事件元素实体信息加入模型中。

在注意力机制中,以 Q , K , V 三个输入为基础。通过 Q 和 K 来获取对于每个位置的权重,在通过这些权重来查询 V 中的值,得到最终输出,这就是注意力机制的基本流程。而本文注意力机制的实际输入,主要有之前 Bi-GRU 的输出,以及事件元素实体的信息。本文将 Bi-GRU 的输出转化为 K , V 两个矩阵,将事件元素实体的信息转化为 Q 矩阵。

将事件元素实体信息转化为 Q 矩阵的方法与之前生成后缀特征的方法类似。首先获取四种事件元素实体的位置,然后通过相同的方式生成对应的向量。由于注意力机制需要 Q 和 K 矩阵的向量长度一致,所以又对事件元素实体信息进行一次线性变换,最后得到 Q 矩阵。

将 Bi-GRU 的输出从中间分割, 会得到两个矩阵, 本文将这两个矩阵作为注意力机制输入的 K, V 矩阵。但是 Bi-GRU 的输出是正向 GRU 和逆向 GRU 的输出连接直接连接起来, 直接分割会将正向和逆向的信息分开, 而我们需要的是 K, V 两个矩阵各自都包含有足够的正向和逆向信息, 所以这里需要使用一个线性变换将正向和逆向输出混合之后, 再进行分割, 获得 K, V 矩阵。

在获得 Q, K, V 三个矩阵之后, 接下来进行注意力机制的计算。计算注意力采用常见的点乘方法。综上, 从 Bi-GRU 到注意力机制的计算方法如下:

$$K, V = W_{mix}x_{bi-gru} + b_{mix} \quad (3.13)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.14)$$

其中 x_{bi-gru} 为 Bi-GRU 的输出矩阵。

如果查看事件元素实体信息转化为 Q 矩阵的方法, 会发现线性变换之前的 Q 矩阵中大部分都是 0, 所以经过线性变换之后 Q 矩阵中的信息还是比较稀疏, 导致直接使用注意力机制作为后续输入效果不佳。与之相对的信息比较稠密的形式就是源于字向量的信息。所以最后还需要将注意力机制的输出与 Bi-GRU 的输出连接到一起, 基于事件元素实体注意力机制的信息只作为辅助, 提升最终的抽取效果。

3.4 实验结果与分析

3.4.1 事件元素实体抽取实验

在事件元素实体抽取中, 后缀特征在模型最初和 bert 一起输入一次, 在 CRF 前又输入一次, 两次又重复不同的次数。本文对于这两次输入的次数进行了简单的对比实验, 实验不同输入次数对于模型最终效果的影响。

(1) 实验参数设置

表 3.3 数据集中实体数量统计

Tab. 3.3 The statistics of the number of entities in dataset

实体类型	装备实体	任务实体	单位实体	战争名称	数值实体	地点实体	时间实体
训练集	4105	119	1251	341	934	1021	1659
测试集	848	23	245	61	185	233	341
总数	3257	96	1006	280	749	788	1318

事件元素实体抽取采用本文标注的军事图谱数据集，将 80% 的数据作为训练集，剩余的作为测试集。其中训练集共 2858 条数据，测试集共 715 条数据，模型的事件元素实体使用 BIO 标签方式，各个实体在训练集和测试集，以及总数的统计如表 3.3。

实验的超参数设置：

表 3.4 超参数设置

Tab. 3.4 The setting of hyperparameters

参数名称	参数值
bert 字向量	768
Bi-GRU	512
自注意力机制	512
Dropout	0.5
学习率	1e-5

(2) 实验结果及分析

首先本文在 CRF 前没有输入后缀特征的情况下，对于和 bert 字向量连接的后缀特征重复次数进行对比实验。

表 3.5 模型最初时后缀特征重复次数对比实验

Tab. 3.5 Comparison experimental results of different repeat times of suffix feature at model start

后缀特征重复次数	Precision(%)	Recall(%)	F1-score(%)
0	74.86	80.73	77.68
1	76.97	79.61	78.26
2	76.51	79.27	77.87
5	74.60	78.60	76.55

实验最初在只重复一次的情况下提升微弱，猜想可能是因为后缀特征向量长度只有 4，与 bert 字向量的长度悬殊导致的。但是在不同重复次数的对比下，发现只重复一次 F1 值提升最高。但是经过之前对于后缀特征和真实结果之间的契合度统计发现，后缀特征对应的位置由 0.924 的概率是对应实体的一部分，所以后缀特征信息还没有被充分利用。既然只要被后缀特征匹配上，就有很大概率是对应的实体，而且这个概率远远大于当前模型识别能力，那么将后缀特征在模型即将输出的位置直接输入，会有什么效果？本文在前面后缀特征与 bert 字向量连接且重复一次（效果最高）的条件下，在模型快要结束的位置，也就是 CRF 之前进行输入，并重复多次，实验效果如下：

表 3.6 CRF 前后缀特征重复次数对比实验

Tab. 3.6 Comparison experimental results of different repeat times of suffix feature before CRF

后缀特征重复次数	Precision (%)	Recall (%)	F1-score (%)
0	76.97	79.61	78.26
10	75.31	82.53	78.76
20	76.60	82.11	79.26
40	74.94	82.71	78.63

最终发现在 CRF 之前重复 20 次效果最佳，次数过多过少效果都会下降。而经过两次输入，神经网络充分使用了后缀特征的信息，相比最初的 Bi-GRU+自注意力机制+CRF 效果已经有很大提升。

鉴于进行一次训练需求的时间过长，所以本文没有针对模型输入时重复多少次和 CRF 之前重复多少次分别作为两条坐标，形成一个二维统计网格，而是在模型输入时重复的最优次数的基础上再优化 CRF 之前重复的次数。所以这里只能得出一个初步结论。初步结论是在模型开始输入时，后缀特征只需重复一次即可最优；在 CRF 之前需要后缀特征则需重复 20 次可达到最优。在 CRF 之前输入后缀特征能够提升模型效果可能是因为，后缀特征对应的位置由 0.924 的概率是对应实体的一部分，远远超过模型的识别性能，而 CRF 在学习了 BIO 标签之间的隐含关系之后，同样能够进行简单的序列标注，比直接进行线性变换效果会有提升。而后缀特征长度只有 4，相对于注意力机制的输出向量长度较短。通过不断重复可以增加后缀特征的影响力，但是注意力机制的输出信息也很重要，所以两者之间存在一个平衡，体现在实验中就是重复 20 次后缀特征模型效果最好。而在模型开始时后缀特征只重复一次效果就是最高，可能是因为 Bi-GRU 中的门控机制抑制了多次重复的信息，也有可能是注意力机制有能力过滤掉此类冗余信息，也有可能是从模型最初的输入到最后的输出距离过长的原因，但是在没有经过验证之前无法下定论。

3.4.2 事件触发词抽取实验

本文在事件触发词抽取中，基于 Bi-GRU+CRF 模型，基于事件元素实体建立注意力机制。

(1) 实验参数设置

事件触发词抽取采用本文标注的军事图谱数据集，将 80% 的数据作为训练集，剩余的作为测试集。模型的输出的事件触发词使用 BIO 标签方式。其中训练集共 2858 条数据，测试集共 715 条数据，各个事件触发词在训练集和测试集，以及总数的统计如下。

表 3.7 数据集中事件触发词数量统计

Tab. 3.7 The statistics of the number of event triggers in dataset

触发词类型	作战事件	参战事件	移动事件	部署事件
训练集	668	53	153	23
测试集	168	18	33	7
总数	836	71	186	30

实验超参数设置:

表 3.8 超参数设置

Tab. 3.8 The setting of hyperparameters

参数名称	参数值
bert 字向量	768
Bi-GRU	512
注意力机制	512
Dropout	0.5
学习率	1e-7

(2) 实验结果及分析

为了评价本文的事件触发词抽取模型的效果, 选择了以下几种方法进行对比:

DMCNN 模型, 通过多个动态卷积网络提取句子层面特征, 以及当前词的周围词向量连接组成的语义特征, 来对当前词进行分类。输入上有词向量, 位置特征, 事件类型特征; 缺点是要求以词为单位, 中文事件抽取时分词可能会对结果造成很大影响, 无法使用 bert 字向量。

JRNN 模型, 输入为 word2vec 词向量, 实体类型, 深度依赖树特征; 通过 Bi-RNN 作为编码器。**JRNN** 认为一句话中不同的事件触发词之间具有某种约束关系, 同一个事件下的事件元素的元素角色之间也具有某种约束关系, 于是通过矩阵将已经抽取的事件触发词和事件元素记录, 通过之前的记录和 Bi-RNN 的输出来抽取事件触发词和事件元素。

NPN 模型, 输入为字向量和词向量的混合, 通过对不同的词分类来找到事件触发词的位置。**NPN** 认为中文的事件触发词具有某种特定的内在形式, 同时中文的事件触发词长度有限。但是这种输入导致无法应用在外文上, 也无法使用 bert 字向量。

TLNN 模型, 考虑到中文的多义词现象, 使用 HowNet 作为外部知识库, 同时输入字和词的信息, 通过 LSTM 编码, 然后通过序列标注的方式抽取事件触发词。**TLNN** 由

于使用了外部知识库，具有一定的分辨多义词的能力，但是不如动态字向量分辨多义词的能力强。而且由于复杂的输入，导致无法使用 bert 字向量。

PLMEE 模型，着重解决了一个事件元素扮演多个元素角色的问题；在事件触发词抽取上，使用 bert 词向量作为输入，然后使用 MLP 进行分类，使用交叉熵作为损失函数。使用 bert 词向量为模型效果带来了较大提升，但是最终的分类形式更适合英文，中文还需要转换为 BIO 标签。

表 3.9 事件触发词抽取结果对比

Tab. 3.9 Comparison experimental results of event trigger identification and classification

方法	触发词识别			触发词分类		
	精确率/%	召回率/%	F-值/%	精确率/%	召回率/%	F-值/%
DMCNN	76.3	65.1	70.3	73.6	63.2	68.0
JRNN	67.7	75.2	71.3	65.3	72.4	68.7
NPN	78.7	69.3	73.7	74.6	65.2	69.6
TLNN	73.2	77.6	75.3	70.4	72.4	71.4
PLMEE	81.8	82.4	82.1	78.2	78.6	78.4
Ours	82.0	78.2	80.1	78.3	74.4	76.3

实验结果如表，相较于未使用 bert 字向量的方法，本文方法在事件触发词抽取上有明显提升。与通过神经网络模型自动抽取事件特征进行事件触发词抽取的 DMCNN 模型相比提高 F-值提高 9.8%，和 JRNN 模型相比 F-值提高 8.8%。与使用字向量和词向量融合的 NPN 相比，F 值提高 6.4%。与使用了外部知识库的 TLNN 相比，提高了 4.8% 左右。但是与同样使用 bert 的 PLMEE 模型相比，召回率较低。

经过以上的简单分析，本文的模型的性能比 PLMEE 模型较低。这可能是因为本文将事件元素实体输入到注意力机制时重点考虑了位置信息，但没有充分发挥语义特征信息的作用，在事件元素实体的语义特征信息提取方面还略显不足。虽然 PLMEE 在 bert 词向量之后直接使用分类器对词进行分类，模型结构比较简单。但是相对于其他模型来说，本文的模型性能有较大提升。相对于使用 word2vec、glove 等静态字向量的 DMCNN 和 JRNN 模型，本文各项指标都有较大提升，这说明 bert 会为事件抽取带来更好的效果。同时 JRNN 相对于 DMCNN 提升微弱，说明通过复杂的结构实现事件触发词和事件元素的联合抽取可能不会有很大提升。而 NPN 和 TLNN 两个在字向量方面，虽然都是使用静态的字向量，但是 NPN 综合使用字向量和词向量，TLNN 考虑了多义词的因素，但是都无法与 bert 动态字向量相比。bert 不仅仅能够为多义词提供不同的语义向量，还

能够根据语境生成对应的向量，而且足够的向量长度还可以提供更多的语义信息。所以，中文事件触发词的抽取未来应该基于 bert 字向量，以字为单位。

除了性能之外，DMCNN 和 JRNN 以词为单位进行抽取，在分词的时候受分词结果影响，也无法使用中文 bert 字向量。NPN 将字向量和词向量结合；TLNN 使用外部知识库，同时也使用了词向量，两者都很难接入 bert 字向量。所以与这些方法相比，本文的模型更能贴合百科文本中的军事图谱构建需求。

3.5 本章小结

本章首先描述了在本文标注的军事图谱数据集中事件元素实体以及事件触发词标签的概况，描述了本章的事件元素实体抽取以及事件触发词抽取任务。然后介绍了两项任务的各种细节。

在事件元素实体抽取中，本文通过观察数据集，发现了军事专业相关实体的共性，即实体的后缀词取词范围小能够轻松统计。于是基于 Bi-GRU+CRF 模型设计了本文的模型。通过向模型中输入后缀特征来提升模型效果。同时实验了在不同处重复不同次数对于模型效果的影响。

在事件触发词抽取中，本文观察发现，与军事专业相关的实体往往与军事事件密切相关。于是通过注意力机制将军事专业相关的实体信息输入到模型中。通过与其他模型对比发现，与未使用 bert 的方法对比有很大提升，但是与通用使用 bert 的方法还稍有不足。可能是本文的模型在使用事件元素实体的语义信息能力方面还有欠缺。

4 元素角色抽取

4.1 元素角色抽取问题描述

4.1.1 问题描述

在事件抽取中，一个事件是由一个事件触发词作为核心，与一个或多个事件元素一同组成。其中事件触发词是最能够代表事件发生的词，事件元素是事件的参与者；事件元素在整个事件中扮演的角色就是对应的元素角色。经过第三章神经网络模型的抽取，本文已经抽取了事件触发词和事件元素实体，接下来就需要确定在一个事件中事件元素的元素角色，以组成一个完整的事件。

本文的数据集中，有四种事件触发词和七种事件元素实体，每种事件对应了 5 到 6 种元素角色。事件触发词类型，事件元素实体类型，和元素角色之间具有复杂的约束关系，对应的关系如表 2.1-2.4 和图 3.1。这种复杂的约束关系对提高模型的精确度有很大帮助，可以消除那些在标注时候就已经明确的不可能存在的元素角色标签；不过约束信息的复杂程度较高，需要设计比较精巧的输入方式。

在事件中，一个事件触发词能够和多个事件元素之间存在关系。在常见的关系抽取模型中，一次需要输入一个句子，以及两个实体的位置，通过多分类方法得到关系的类型。但是在本文中，如果一次只输入一个事件触发词和事件元素实体，虽然可以充分利用约束信息，但会导致训练和预测时每得到一个元素角色都需要运算一次，运算次数和成本将大幅提升，同时如果对于每一对事件触发词和事件元素类型组合训练一个模型，需要的模型数量变多，模型训练成本和所需时间将成倍提高；如果一次能够输入一个句子中的一个事件触发词和全部事件元素实体，可以降低运算次数，节约训练时间和硬件资源，减少需要训练的模型数量，但很难利用复杂的约束信息。所以本文需要使用可以同时抽取多个关系的方法，同时还可以使用约束关系提升模型性能。

如果需要输入不确定个数的事件元素实体的位置信息，同时输入不确定个数的元素角色信息，那么文本依赖树抽取是一个很好的参照对象。文本依赖树抽取中，可以将每条关系弧同时进行多分类。但是在检索使用 bert 字向量的中文文本依赖树抽取时的相关论文时，发现文献数量很少，这是由于文本依赖树抽取是以词为单位，而中文 bert 是以字为单位。因为中文的词数量远超字的数量，自行训练一个中文 bert 词向量不是很现实。即使是中文 bert 字向量，训练也需要大量的计算资源。而且经过许多自然语言处理任务的证实，bert 字向量的使用能够显著提升自然语言处理的模型性能。所以本文需要寻找一种使用 bert 字向量获得词的信息的方法。

4.1.2 解决方案

针对以上三个问题，本文对应提出了解决方案。

首先是事件触发词类型，事件元素实体类型，和元素角色之间复杂的约束关系的使用。本文对应提出了掩盖矩阵的方式。在掩盖矩阵中，可能出现的元素角色位置为 1，不可能出现的元素角色位置为 0，每次使用模型时实时生成对应的掩盖矩阵。在模型最后通过点乘的方式使用掩盖矩阵，将不可能出现的元素角色关系掩盖掉，将复杂的约束关系通过掩盖矩阵的方式输入到模型中，提升模型的性能。

接下来是如何同时抽取多个关系的问题。通过上述的掩盖矩阵已经可以解决同时抽取多个关系可能带来的约束关系使用的难点。在一个事件中，以一个事件触发词为核心，与若干个事件元素之间存在关系，形成一种星型结构，而星型结构也可以视为结构简单的树状结构。在文本依赖树抽取，也需要同时抽取多个关系，同时要求关系之间能够形成树状结构，正好与本文的事件的星型结构符合，所以本文从文本依赖树抽取方法中寻找灵感。其中使用双仿射的文本依赖树抽取方法仍是比较先进的神经网络文本依赖树抽取方法。所以本文使用双仿射的角色分类器作为同时抽取多个关系的方案。

最后是寻找使用 bert 字向量得出词的信息的方法。首先本文已经知道事件触发词和事件元素的位置信息，包含了词的起始位置和结束位置。我们可以轻松想出多种方案，例如将词的所有字的向量加和平均，词的结尾向量减去起始向量，等等。首先经过第三章的后缀特征的分析我们认知到，中文的词通常是修饰部分在前，核心部分在后，重点含义在后面。同时在文献^[39]中，Xingchen Li 等人提出使用词的结尾字的向量代替整个词进行文本依赖树抽取，同时完成了分词工作。所以本文使用词的结尾字对应的向量代替整个词的含义，输入到双仿射模型中参与计算。

综上所述，本文使用掩盖矩阵的方式输入复杂的约束关系，选择双仿射的文本依赖树抽取方法进行多关系同时抽取，使用的结尾字对应的向量代替整个词的含义，解决了元素角色抽取的三个难题。

4.2 模型总体结构

模型总体基于 Bi-GRU。模型输入为中文 bert 字向量，经过 Bi-GRU 神经网络模型提取上下文信息。然后使用线性变换混合正向和逆向的信息，将向量从中间进行分割成两部分，一部分用来作事件触发词信息，另一部分作为事件元素实体的信息。接下来使用一个词的最后一个字的向量来代表整个词。然后将事件触发词的向量和句子中全部事件元素向量输入到双仿射的关系类型分类器中，同时抽取多个事件元素的元素角色。最后，将事件触发词类型，事件元素实体类型和元素角色之间的约束关系做成掩盖矩阵，

将模型最后的输出中不可能出现的元素角色类型掩盖起来，再使用 softmax 进行向量归一化，得到模型最终的输出。

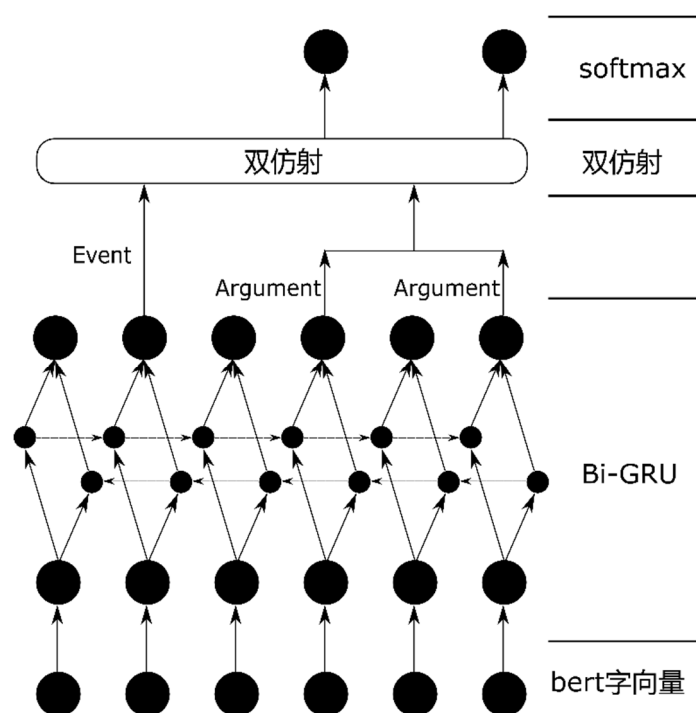


图 4.1 元素角色抽取模型结构图

Fig. 4.1 The structure of argument role extraction model

4.2.1 模型输入及 Bi-GRU

模型主要的和最初的输入是中文 bert 字向量。作为动态字向量，使用 bert 时与 word2vec 相比模型效果能获得很大提升。除此之外，模型还需要输入事件触发词和事件元素实体的位置信息，以及事件触发词、事件元素实体、元素角色之间的约束关系，这些除了 bert 字向量以外的信息需要在后续的步骤进行输入。其中位置信息通过将对应位置的向量挑选的方式输入，约束关系通过掩盖矩阵的方式输入。

在 bert 字向量之后，本文选择 Bi-GRU 作为编码器。门控循环单元 GRU 作为长短期记忆网络 LSTM 的一个变种，在继承了能够避免梯度爆炸或梯度消失的优点的基础上，优化减少了计算步骤，减少了训练参数的数量，使得模型的所需的存储空间减少，降低了训练和预测时的计算量，提高了训练和预测时所需的时间，却基本没有影响最终的模型效果。由于人们在阅读语言的时候，不仅是从前向后阅读，在思考后面的词汇的时候也需要联系前文，所以人们提出将文本序列正向逆向分布输入到循环神经网络一类

的模型中，将对应词或字的节点的正向逆向输出进行连接。如果使用 GRU 作为其中的循环神经网络类的模型，就会得到双向循环神经网络单元，即 Bi-GRU。而在双仿射深度文法依赖抽取方法原文中，使用了 Bi-LSTM 模型作为编码器。而 Bi-GRU 也像 Bi-LSTM 一样，同样能够有效地抽取上下文信息。

4.2.2 线性混合及双仿射

与之前的事件触发词抽取一样，在分割 Bi-GRU 输出的向量之前，需要进行一次线性变换，混合正向和逆向的信息。然后将得到的向量从中间分割，分割出来的两部分，一部分将事件触发词结尾位置的向量提取出来，另一部分将事件元素实体结尾位置的向量提取组成矩阵。经过线性混合之后，事件触发词向量和事件元素实体向量会同时带有 Bi-GRU 输出的正向和逆向的信息；如果没有线性混合，分割后得到的两部分就会分别是正向信息和逆向信息。然后使用事件触发词和事件元素词的结尾字代替整个词的信息，放入到后续的使用双仿射的文法依赖树角色抽取方法的关系类型分类器中。

上面的步骤中使用词的结尾处的向量来代替整个词。这是因为在中文中，一个词的核心部分一般在词的后面，形容修饰的部分放到前面。在上一章中，通过装备实体，任务实体，单位实体，战争名称实体四种实体的后缀特征来提升模型效果，其中后缀就是这些词的核心部分。

双仿射的输入有两个，一是事件触发词向量，二是一组事件元素实体向量，并且对于每一个事件元素实体，输出对应的一组向量表示每种元素角色的概率。假设事件触发词向量的长度为 v ，事件元素实体向量的长度为 n ，最后输出的元素角色种类数量为 l 。首先，双仿射模型里有两个可训练的参数，一个是权重张量，大小为 $v \times n \times l$ ，对于每种元素角色都可以有一个 $v \times n$ 大小的矩阵，假设第 i 类元素角色对应的权重矩阵为 W_i ；二是偏置向量 b ，长度为 l 。

计算时首先，对于每一类元素角色，使用权重矩阵，将事件触发词向量的长度转换，然后与事件元素实体向量进行点乘，对结果进行加和。

$$v_i = x_{event} * W_i * x_{argu}^T \quad (4.1)$$

其中 W_i 是第 i 类元素角色对应的权重矩阵， x_{event} 为事件触发词的向量， x_{argu} 代表其中一个事件元素实体的向量。最后结果获得一个数 v_i 。

$$s = v + b \quad (4.2)$$

每一种元素角色都有一个对应的数字 v_i ，连接在一起可以获得一个向量 v ，然后与偏置向量 b 加和，获得最终双仿射的输出结果 s 。

作为双仿射深度文法依赖树抽取方法的关系类型分类器，可以同时抽取多个一对一的关系。但是如果简化输入，例如将双仿射的两个输入中一个用事件触发词的输入重复多次代替，另一个输入句子中全部的事件元素实体，即可用多个一对一的关系模拟出一对多的关系，同时抽取一个句子中一个事件触发词与全部事件元素实体之间的关系。

4.2.3 掩盖矩阵和 softmax

上节的双仿射已经输出了不同元素角色对应的概率权重，大多数的神经网络模型都会直接使用 softmax，将输出的向量转换为对应元素角色的出现概率。但是在标注时，我们定义了事件元素实体类型和元素角色类型之间的约束关系。本文通过使用掩盖矩阵将这些约束关系信息输入到模型中，提高模型效果。

模型中，对于每句话中的每个事件触发词都会生成对应的掩盖矩阵。掩盖矩阵中，可能会出现元素角色位置数字为 1，不可能出现的元素角色位置为 0。掩盖矩阵的生成方法如下：首先，将事件触发词，事件元素，元素角色的类型进行排序，之后的各种矩阵向量的对应的索引遵守这些排序；然后对于每种事件触发词准备一个事件元素实体类型和元素角色之间的约束关系矩阵（类似于表 2.1-2.4），其中可能会出现元素角色位置数字为 1，不可能出现的元素角色位置为 0；最后对于每个句子中的每个事件触发词，挑选对应的约束关系矩阵，然后按照句子中的事件元素实体顺序选出实体类型对应的向量，组成矩阵，最后加上全是 1 的一列代表没有关系的可能性，最后生成了对应的掩盖矩阵。最终得到的掩盖矩阵中，可能会出现元素角色位置数字为 1，不可能出现的元素角色位置为 0。

掩盖矩阵的最终目的，是使得最终输出的元素角色概率中，不可能出现的元素角色概率值最低。但是如果之间将双仿射的输出直接按位乘上掩盖矩阵，则不可能出现的位置被置为 0，但是其他位置还有可能会出现负数的情况，负数无疑比 0 小。所以我们需要在掩盖之前使用 relu 激活函数，使得小于零的位置全部变成 0，然后每个位置加上一个微小的固定值 0.01，再按位乘上掩盖矩阵，之前为负数的可能出现的元素角色位置则为固定值 0.01，最小的数就是不能出现的元素角色位置的 0 了，自然达成了不可能出现的元素角色概率值最低的目的。之后再使用 softmax，就可以得出每种元素角色出现的概率了。

而数据经过 softmax 之后，向量的值全部为正数，按位乘以掩盖矩阵之后也可以实现不可能出现的元素角色类型的位置值最低。但是 softmax 之后得出的向量一般被人们称为概率，是因为这个向量不仅所有的数值都大于 0，而且所有的数值加和为 1。如果将掩盖矩阵移到 softmax 之后，会使向量中部分大于零的数变为 0，减小了向量的总和。为了保证最后输出的向量能够被称为概率，本文选择在 softmax 之前使用掩盖矩阵。

4.2.4 损失函数

元素角色抽取模型最后为 softmax 模型，是一个多分类的问题。所以本文采取了多分类常用的交叉熵作为模型的损失函数。损失函数的计算方式如下：

$$loss = -\sum_{i=1}^n p(x_i) \log(q(x_i)) \quad (4.3)$$

其中 $p(x_i)$ 代表了需要预测的真实值， $q(x_i)$ 为模型的预测结果。

4.3 实验结果与分析

本文元素角色抽取采用了掩盖矩阵输入约束信息，所以实验中首先进行了有无掩盖矩阵的对比实验，之后再进行与其他同类模型的效果对比实验。

4.3.1 掩盖矩阵对比实验

本文军事图谱数据集在标注的时候已经规定了实体类型和元素角色之间的关系。当事件元素的实体类型和事件触发词的事件类型已经确定的条件下，只有固定的几种元素角色可能出现，其余不可能出现。于是本文将这种约束关系做成掩盖矩阵，将不可能出现的元素角色掩盖掉。

(1) 实验超参数设置

表 4.1 超参数设置

Tab. 4.1 The setting of hyperparameters	
参数名称	参数值
bert 字向量	768
Bi-GRU	512
Dropout	0.5
学习率	1e-5

(2) 实验结果与分析

表 4.2 掩盖矩阵对比实验

Tab. 4.2 Comparison experimental result of mask matrix			
	精确率/%	召回率/%	F-值/%
无掩盖矩阵	73.95	89.08	80.81
有掩盖矩阵	81.89	88.44	85.04

实验结果可以发现,掩盖矩阵可以大大提高模型的精确率,但是会略微降低模型的召回率,综合起来可以极大提升 F 值。通过推理可以发现,掩盖矩阵会去除某些错误的抽取选项,所以能够提升模型精确度;但是掩盖错误的同时会导致反向传播的时候,被掩盖的位置误差无法传播,可能会降低模型性能。最终通过实验发现,掩盖矩阵虽然损失了部分召回率,但能够极大提升效果。

4.3.2 不同模型之间的对比实验

本文在进行元素角色抽取之前,已经通过实体识别来确定事件元素的位置。虽然一个句子中的事件触发词与事件元素实体之间可能没有关系,但是这已经提供了一个确切的位置信息,方便了后续的抽取。而事件抽取的常见的评价标准里,与事件元素相关的是事件元素识别和事件元素分类两项。其中事件元素识别中,当一个事件元素的位置识别正确即认为事件元素识别成功;在事件元素分类中,需要一个事件元素的位置识别正确,同时元素角色识别也正确,才认为事件元素分类成功。所以本文将抽取结果进行换算,并按照这种常见标准来比较各种模型的优劣。

为了评价本文的模型效果优劣,选择了以下几个经典模型作为对比:

DMCNN 模型,通过动态多卷积模型来进行事件元素抽取。在模型运算中,会通过事件触发词和候选事件元素的结尾位置来对矩阵进行切分,所以模型要求输入的是词向量。即使输入了字向量,也无法区分候选词的起始位置,所以此模型对于本文的军事百科图谱构建中缺乏实际应用价值。

JRNN 模型,一个通过 **Bi-RNN** 的联合事件抽取模型,能够同时抽取事件触发词和事件元素,通过矩阵来保存已经抽取的事件触发词和事件元素信息,认为一个句子里事件触发词之间和事件元素之间具有某种约束关系。认为通过联合抽取能够减小二阶段抽取方法带来的误差。

PLMEE 模型,考虑了一个事件元素可能在一个事件里可能会有多个元素角色,于是模型最后的分类器经过精心设计,以抽取此类事件元素。模型的输入为 **bert** 字向量,然后连接分类器。其中 **bert** 字向量模型效果带来了较大提升。

首先,通过观察表中数据可以发现,使用 **bert** 字向量的模型性能明显高于未使用 **bert** 字向量的模型。这是因为 **bert** 字向量能够为多义词在不同语境下的含义得出不同的字向量,使用的注意力机制能够更好地使用长距离信息,输出的字向量长度为 768 能够携带更多语义信息,更深的模型具有更多的参数量能够更好地拟合数据集。其中 **PLMEE** 只通过 **bert** 字向量和简单的 **MLP** 分类器就获得了较高的模型效果,而本文的模型通过 **bert** 字向量和后续更多的复杂模型获取了更高的抽取效果。

表 4.3 事件元素抽取结果对比

Tab. 4.3 Comparison experimental results of event argument identification and classification

方法	事件元素识别			事件元素分类		
	精确率/%	召回率/%	F-值/%	精确率/%	召回率/%	F-值/%
DMCNN	57.7	58.3	58.0	55.6	56.4	56.0
JRNN	59.8	61.7	60.7	57.2	59.3	58.2
PLMEE	61.8	68.6	65.0	59.4	66.3	62.7
Ours	65.0	74.7	69.5	62.7	72.6	67.3

接下来,可以发现本文提出的模型性能超越了 PLMEE 模型性能。这可能有几方面的原因。一是本文的模型结构比 PLMEE 模型更加复杂。本文在 bert 字向量之后,通过 Bi-GRU 模型抽取上下文特征,双仿射深度文法依赖的角色分类器进行元素角色分类,以及掩盖矩阵去除不可能出现的元素角色,共三重提升,不断提高模型性能;而 PLMEE 在 bert 字向量之后仅有一层 MLP 分类器。二是,本文标注的军事图谱数据集中并没有 PLMEE 所关注的同一个事件元素在一个事件中出现多个元素角色的现象,所以 PLMEE 所精心设计的分类器和复杂的事件元素提取算法不仅对于本文数据集没有任何提升作用,而且还会提高空标签的比例,有可能降低模型性能。

而 DMCNN 和 JRNN 模型的性能低下,也有可能是因为语言不同导致的模型性能下降,以及 bert 字向量带来的性能提升。原本这两种模型是用于英文事件抽取,其中事件触发词和事件元素被默认为是一个词。而在使用本文的中文军事图谱数据集进行训练时只能进行一些变通,选用了词的最后一个字代替整个词。这样能够避免分词带来的误差。这也导致这些模型只能比较模型性能,完全无法直接应用到本文的军事图谱构建中。除此之外, bert 字向量可以为大部分自然语言处理任务较大提升,这是因为 bert 能够在不同语境下为字或词动态提供不同的向量,同时较长的字向量能够包含更多信息。

最后,从表中可以发现,本文的模型与其他方法对比中,准确率,召回率以及 F-值都优于其他模型。

4.4 本章小结

本章在上一章已经抽取事件元素实体和事件触发词的基础上,抽取了元素角色。输入上选择了中文 bert 字向量。以 Bi-GRU 作为编码器,双仿射深度文法依赖的角色分类器对元素角色进行分类,然后通过掩盖矩阵来去除在定义数据集标签时已经确定不可能出现的元素角色,最终获得了理想中的效果。

本章首先描述了元素角色抽取问题。然后详细介绍了本文的元素角色抽取模型，着重介绍了双仿射深度文法依赖的角色分类器以及掩盖矩阵。最后设计了两组对比实验。通过掩盖矩阵的对比实验，来证实掩盖矩阵的效果；通过与其他模型的对比，验证本文模型的效果。

5 中文军事事件关系抽取系统

本文标注的军事百科图谱数据集中，在事件方面最终需要以事件关系为线将事件连接，以事件为节点构建事理图谱。而当前事件之间的关系研究还在初步阶段，很多实际项目中还是以模板方式为主，使用神经网络的算法还在起步阶段。本章使用事件关系连词构建模板，通过较为可靠的模板的方式抽取事件关系，基于第三章和第四章中的模型和算法抽取事件，构建一个中文军事事件关系抽取系统，能够从非结构化文本中抽取事件并抽取事件关系。

5.1 中文军事事件关系抽取系统框架

事件关系抽取是事理图谱构建的重要步骤。经过事件抽取和事件关系抽取，就已经可以初步形成一个事理图谱。军事事理图谱不仅可以总结文本中军事事件之间的关联，还可以通过推理来发现事件之间的隐含关系。通过观察宏观的军事事件关系，可以为国家的军事战略进行指导，包括军事系统升级，武器战略研发计划等；用于微观的军事事件时，则可以用于指导具体的军事行动，改良日常的军事训练。

在军事图谱构建过程中，初始的输入是各种非结构化的文本，得出的是结构化的图谱数据。一个常见的图谱构建流程大致包括节点抽取，关系抽取等步骤。在本文负责的与事件相关的部分中，包含了事件抽取以及事件关系抽取。其中事件抽取在本文中又分为了事件元素实体抽取，事件触发词抽取，元素角色抽取三步。

首先，直接从网页中爬取，或者来自于文档的原始语料，是无法直接输入到神经网络模型中的。原始语料需要经过分段分句，去除噪音文本等处理，才能得到能够直接用于后续流程的文本形式。同时其中事件元素实体抽取，事件触发词抽取，元素角色抽取三步使用的是基于 bert 字向量的神经网络模型，计算量大消耗时间长，而最终的事件关系抽取由于基于模板，消耗时间短。而且这里面某些步骤会对接下来的步骤有数据过滤作用，例如一个句子中没有事件或只有一个事件，那么不会进入到事件关系抽取中。这样，我们可以适当调整顺序，提前进行数据过滤，寻找一种减少计算量的方案。

但是几种神经网络模型的过滤效果必需在模型运算完成后才会发生，而模板抽取的过滤效果，如果句子中不存在模板中对应的关系连词，就不会被模板抽取，也就形成了一定的过滤效果。所以本文通过模板中的关系连词来对预抽取文本进行过滤。综上，事件关系抽取系统的流程图如下：

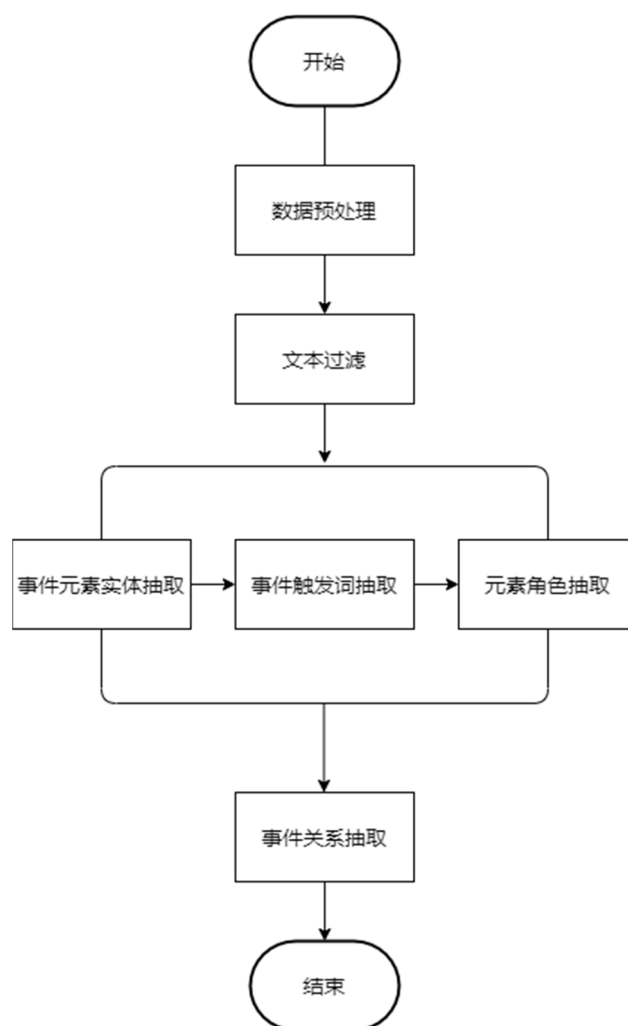


图 5.1 事件关系抽取流程图

Fig. 5.1 Flow chart of event relation extraction

5.2 数据预处理模块

数据处理模块的目的是将原始语料转变成后续模块需要的形式。本文的各种模型都是以句子为单位的抽取方式，所以数据处理的第一步就是分句。本文是基于中文，所以句子分割首先是分割换行符，其次是各种中文句子分隔符，如句号问号叹号等，通过这些分隔符完成句子分割。

数据处理第二步是去除噪音文本。本系统的输入是各式各样的军事相关百科文本。但是根据之前神经网络抽取的经验，未经处理的原始文本中含有大量噪音文本，这些文本需要进行过滤才能输入到神经网络模型中。而且来源不同的文本的噪音也是各种各样，

所以数据处理模块还需要一定的过滤方法的扩充能力。而之前处理数据集处理的流程中，过滤了以下噪音数据：

- (1) 各种空白字符，包括空格符，制表符。
- (2) 各种括号，以及括号里的内容，全部丢弃。
- (3) 各种英文单词。
- (4) 其他特殊符号，例如“/”等。

最后一步还需要校验句子的长度。如果第一步分好了句子，第二步却当成噪音全部去除了，会造成句子内容缺失；而长度过短的句子包含内容较少，难以抽取事件关系，所以最后需要将长度较短的句子去除。

最终，数据处理模块要获得语义流畅连贯的句子，句子中的绝大部分字的含义都能被后续的字向量清晰表达。这样的句子能够在保留原有中文事件的同时，提高后续事件抽取的成功率。

5.3 事件抽取模块

事件抽取模块的主要功能是从非结构化的句子中抽取事件的四个要素（事件触发词、事件元素、事件类型、元素角色），主要分为事件元素实体抽取，事件触发词抽取，数据过滤，元素角色抽取四步。

在本文第三章中，实现了事件元素实体和事件触发词的抽取。事件元素实体抽取模型通过后缀特征以及神经网络模型对事件元素实体进行抽取。其中的后缀特征需要构建对应的后缀词库。建议收集足够数量的实体词，然后进行倒序字频统计，这样很容易获得高频后缀，然后稍加人工筛选即可获得高质量的后缀词库。事件触发词抽取模型需要句子以及事件元素实体的位置信息，所以需要安排在事件元素实体抽取之后。在第三章中，本文基于神经网络，使用后缀特征，以及基于事件元素实体的注意力机制，完成了事件元素实体以及事件触发词的抽取，并且通过实验证实，本文的模型具有较高的抽取性能。

接下来需要对数据进行过滤，步骤如下：

- (1) 如果句子中没有抽取出事件触发词，则丢弃这条句子。
- (2) 如果句子中没有事件元素实体，不需要进行元素角色抽取。但如果句中仍有事件触发词，不需进行元素角色抽取，而且仍然保留句子。

在本文第四章中，使用基于双仿射文法依赖树的角色分类器，以及掩盖矩阵的方法，对元素角色进行抽取。在元素角色抽取中，需要使用到之前抽取的事件触发词和事件元素实体。如果一个事件只有触发词，没有事件元素，最终仍然把它当做一个事件处理。

最终，事件抽取模块将句子中的事件抽取出来，包括事件触发词位置，事件元素位置及对应的元素角色，和事件类型。

5.4 事件关系抽取模块

事件关系抽取模块的主要功能是以句子为单位，抽取句子中的事件之间的关系。事件关系与实体关系最重要的差别在于，实体关系是特定实体类别之间具有特定的关系类型，而事件关系是任意两类事件之间具有有限种事件关系类型，例如因果关系，顺承关系等。经过以上几步，最初的非结构化文本已经分句，并且抽取出句子中的事件。本文在观察军事数据集后，发现文中存在多种事件关系，本文爬取的军事百科文本内容，选择抽取了因果，转折，顺承三种事件关系。

在中文中，表达事件之间关系的形式主要有三种：隐式表达，两个事件在语境中自然形成关系，如常见的一系列的顺承事件，或者两个事件在语义上会形成关系，例如开枪击中，开枪和击中在语义上就有一定的因果关系；半显式表达，使用一个关系连词引导一个事件，同时兼顾语境和语义上的作用，与另一个没有关系连词引导的事件关联起来；显式表达，直接用两个配对的关系连词引导有关系的两个事件。其中隐式表达很难通过模板的方式抽取，应该通过统计机器学习的方式进行抽取；半显式表达确定没有关系连词的事件时使用模板有一定缺陷，有很大可能会得到错误的结果；面对显式表达时，通过模板抽取会得到较高的准确率。本文通过总结可能出现的显式表达的事件关系连词组合，以正则表达式为工具构建模板，来圈定拥有因果，转折，顺承三种关系的事件触发词的所在范围。如果已经抽取出来的军事事件触发词落入这些范围中，则将相应的事件触发词通过相应的关系连接起来。

表 5.1 关系连词列表
Tab. 5.1 List of relative conjunctions

事件关系模式	关系连词列表	
	前词	后词
由因到果	因为，既然，由于	从而，为此，所以，以至于，于是，以致
由果到因	之所以	是因为，因为，是由于
转折关系	与其，宁可，宁肯	不如，也不，绝不
转折关系	虽然，尽管，虽，即使	但是，但，但也，但却，却
转折关系	不管，无论，不论，即使	都，总，也，始终
顺承关系	又，再，才，首先，先是	进而，其次，然后

基于以上所述，事件关系抽取模块主要分为五步：

(1) 预先根据选定的事件关系构建事件关系连词词库。在实际运行之前可以检查数据中的明显的具有事件关系的句子，将其中的事件关系连词加入到库中。

(2) 判断句子中事件触发词数量，小于 2 的不进行事件关系抽取。

(3) 将句子进行分词，以及词性标注，然后把词和词性通过“/”连接起来，然后将所有词通过空格连接起来，一句话组成一个字符串，方便后续正则模板匹配。分词可以保证匹配的事件关系连词在句子中作为整体存在，词性可以保证关系连词在句子中发挥真正的作用。

(4) 通过建立好的关系连词词库构建模板进行匹配，获取事件触发词所在的子句范围。本文以关系连词为基础构建事件关系模板。模板正则表达式为：

前词/[p|c]+\s(?:, |*)/w+ \s[.*]后词/[p|c]+\s(?: ?! !。 ; ;: :n|r, |*)/w+

其中 p, c, w 分别代表词性中的介词，连词，标点符号。s 匹配词之间的空格，() 内的部分为匹配到的内容。各种标点限制匹配到的内容范围，保障结果的准确性。匹配的部分为两个连词紧接的第一个子句。

(5) 最后检索第 3 步中匹配到的两个取词范围中是否有事件触发词存在。若两个范围中同时有事件触发词存在，然后将两个取词范围内的两组事件进行匹配组合，则认为这些事件组合之间具有事件关系连词代表的事件关系。

这种基于事件关系连词模板的事件关系抽取，虽然无法抽取隐式表达的事件关系，具有很高的准确率，但是召回率很低。使用正则表达式的方法运算量小运算速度快，很容易处理大量文本数据，获得准确率较高的事件关系抽取结果，构建出质量可靠的事理图谱。

5.5 基于关系连词的文本过滤模块

文本过滤模块，主要目的是过滤后续模块无法抽取事件关系的句子，减轻后续各个模块的运算量和计算压力，提高运行速度，位置紧接在数据预处理模块之后。

在事件关系抽取模块中，使用关系连词建立了事件关系模板。可见，如果句子中未出现这些关系连词，就无法抽取出事件关系。基于此，通过判断句子中是否有这些关系连词，如果没有则丢弃。

5.6 本章小结

事件关系抽取是事理图谱构建的关键步骤。接下来需要考虑如何实现从不同位置抽取的同一事件的合并。对于实体来说，字符类型相同的便可以简单合并；但是对于事件

来说，可能在抽取时会存在一义多词，事件元素抽取不全等各种情况。只需再解决这一点，就可以构建一个简单的事理图谱。

其次，使用模板进行抽取不会像机器学习一样，如果数据集质量足够高就会取得很好的结果；模板匹配只能够获得已经定义的模板对应的结果，精确度很高但是召回率较低。

综上，本章通过使用第三章和第四章中的模型进行军事事件抽取，通过基于关系连词的模板进行事件关系抽取，构建了中文军事事件关系抽取系统，可以应用于军事百科文本的图谱构建中。其中本文使用基于关系连词的文本过滤模块，过滤掉无法抽取事件关系的句子，为后续模块减少计算压力，提高整体的抽取速度。基于关系连词的模板的事件抽取方法，虽然召回率很低，但是抽取速度快，精度高。

表 5.2 事件关系抽取结果举例

Tab. 5.2 Examples of the results of event relation extraction

1943 年 12 月 25 日，大和号在特鲁克附近遭到美国潜艇的鱼雷攻击， 由于 战舰右舷第 165 号肋骨被一发鱼雷 命中 ，机械室及火药库进水，进水约 3000 吨， 导致 受损后的大和号 撤离 了作战海域。		
事件关系类型	关系连词	事件触发词
因果关系	由于，导致	[命中，撤离]
攻击机群分成两个集群投入攻击， 尽管 领航的长机被高炮 击落 ， 但 随后对战列舰的 攻击 接连 命中 目标。		
事件关系类型	关系连词	事件触发词
转折	尽管，但	[击落，攻击] [击落，命中]
即使 野猫战斗机 击落 了至少 3 架敌机， 但是 剩下的敌机仍然向着航母及其护卫舰 扑去 。		
事件关系类型	关系连词	事件触发词
转折	即使，但是	击落，扑去
南太平洋海战中 虽然 与瑞凤号合作 击伤 美军企业号和 击毁 大黄蜂号航空母舰， 但 翔鹤号飞行甲板在圣克鲁斯海战中被 命中 4 颗炸弹再次严重受损。		
事件关系类型	关系连词	事件触发词
转折	虽然，但	[击伤，命中] [击毁，命中]
冷战期间，西方不断释放高空侦察气球到苏联领空， 虽然 米格-17 也用于 阻击 高空侦查气球， 但 是仅有少数气球被 击落 。		
事件关系类型	关系连词	事件触发词
转折	虽然，但	[阻击，击落]
出厂后 又 载运 75 架 P-40 战斗机于 2 月 23 日 到达 卡萨布兰卡， 然后至 北大西洋，在新英格兰地区至加拿大新斯科舍省哈利法克斯之间的海域进行巡逻与训练任务。		
事件关系类型	关系连词	事件触发词
顺承	又，然后	[到达，至]
1941 年 11 月份，约克城号 又 进行了一次中立 巡逻 ， 然后 于 12 月 2 日 进入 诺福克海军基地。		
事件关系类型	关系连词	事件触发词
顺承	又，然后	[巡逻，进入]

结 论

本文爬取军事百科文本，标注了军事图谱数据集，并完成了图谱构建中的事件抽取任务，对于事件抽取任务中的事件元素实体抽取，事件触发词抽取和元素角色抽取任务进行了研究，最终使用这些模型和模板方案构建了一个军事事件关系抽取系统。

首先，本文从头条百科中爬取军事装备词条，建立了相应的军事图谱数据集。然后观察爬取的文本，定义军事相关标签。数据集中不仅标注了事件，同时将这些事件元素全部标注为实体，还标注了事件之间的关系和实体之间的关系，方便之后基于图谱构建更丰富的应用。

在事件元素实体抽取任务中，本文通过观察军事专业相关的四种实体，发现这些实体的后缀部分的词取词范围很小，容易构建词库。于是本文将这些后缀部分通过构建词库，建立军事专业相关实体的后缀特征，并将这些后缀特征输入到模型中。最后，本文通过对比实验证实了后缀特征的有效性。

在事件触发词抽取任务中，本文发现之前军事专业相关的四种实体，与军事相关事件具有语义相关性，能够让人们容易将两者联想起来。于是本文基于 Bi-GRU+CRF 模型，通过注意力机制注入事件元素实体信息，用以提高事件触发词抽取效果。最终本文进行了与其他典型模型的对比实验。通过实验得知，本文的模型略逊与同样使用 bert 字向量的 PLMEE 模型。这可能是由于本文模型未能充分利用事件元素实体的语义信息导致。但是本文模型的性能依旧远远超越了未使用 bert 字向量的模型。

在元素角色抽取任务中，本文利用双仿射深度文法依赖树抽取方法中的关系角色分类器，以及将不可能出现的元素角色掩盖的掩盖矩阵，提高模型性能。本文通过使用词的最后一个字的向量代替整个词的含义，使用 bert 字向量进行抽取。最后，本文进行了掩盖矩阵的对比实验和与其他经典模型的对比实验。通过掩盖矩阵的对比实验，本文证实了掩盖矩阵的有效性，掩盖矩阵能够极大提升模型的准确率和 F-值，虽然会略微降低模型的召回率。通过与其他经典模型的对比实验，证明了本文的模型能够有效抽取事件元素。

在军事事件关系抽取系统中，本文使用上面三个任务中的模型和算法抽取事件，使用基于关系连词的模板抽取事件关系。由于基于 bert 字向量的模型计算量大，计算时间长，本文在模型之前使用文本过滤模块减少抽取文本量。经过三个神经网络模型抽取事件之后，通过关系连词对因果，转折，顺承三种事件关系进行抽取。最终构建了一个能够从非结构化中文文本中，抽取军事事件关系的系统。

综上所述,本文的模型具有较好的效果,能够适合面向百科文本的军事图谱构建中。其中 bert 字向量为所有的模型带来了巨大提升,通过各种注意力机制也提高了模型的性能,充分军事百科文本数据集的各种特点提高模型效果。但是仍存在一些不足,如未能充分利用事件元素实体的语义信息,提升事件触发词的抽取效果。

参 考 文 献

- [1]郭喜跃, 何婷婷. 信息抽取研究综述[J]. 计算机科学, 2015, 42(2): 14-17, 38.
- [2]马晨曦, 陈兴蜀, 王文贤, 等. 基于递归神经网络的中文事件检测[J]. 信息安全, 2018, (5): 75-81.
- [3]陈斌. 基于长短期记忆网络的事件抽取研究与应用[D]. 中国矿业大学, 2019.
- [4]沈兰奔. 结合注意力机制与长短期记忆网络的中文事件抽取方法研究[D]. 北京交通大学, 2019.
- [5]杨航. 面向非结构化中文文本的篇章级事件抽取研究[D]. 哈尔滨理工大学, 2019.
- [6]Ellen Riloff. Automatically Constructing a Dictionary for Information Extraction Tasks[C]. Proceedings of the Eleventh National Conference on Artificial Intelligence, 1993, 811-816.
- [7]Ellen Riloff, Shoen Jay. Automatically Acquiring Conceptual Patterns Without an Annotated Corpus[J]. Proceedings of the Third Workshop on Very Large Corpora, 2002: 3.
- [8]Jun-Tae Kim, Moldovan Dan. Acquisition of Linguistic Patterns for Knowledge-Based Information Extraction. [J]. Knowledge and Data Engineering, IEEE Transactions on, 1995: 7713-724.
- [9]Hai Leong Chieu, Hwee Tou Ng. A Maximum Entropy Approach to Information Extraction from Semi-Structured and Free Text[J]. American Association for Artificial Intelligence, 2002: 786-791.
- [10]David Ahn. The stages of event extraction[C]. Association for Computational Linguistics, 2006: 1-8.
- [11]Heng Ji, Grishman Ralph. Refining Event Extraction through Cross-Document Inference[C]. Association for Computational Linguistics, 2008: 254-262.
- [12]Yu Hong, Zhang Jianfeng, Ma Bin, et al. Using Cross-Entity Inference to Improve Event Extraction[C]. Association for Computational Linguistics, 2011: 1127-1136.
- [13]Qi Li, Ji Heng, Huang Liang. Joint Event Extraction via Structured Prediction with Global Features[C]. Association for Computational Linguistics, 2013: 73-82.
- [14]Thien-Huu Nguyen, Grishman Ralph. Event Detection and Domain Adaptation with Convolutional Neural Networks[C]. Association for Computational Linguistics, 2015: 365-371.

- [15]Yubo Chen, Xu Liheng, Liu Kang, et al. Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks[C]. Association for Computational Linguistics, 2015: 167-176.
- [16]Thien-Huu Nguyen, Cho Kyunghyun, Grishman Ralph. Joint Event Extraction via Recurrent Neural Networks. San Diego, California[C], Association for Computational Linguistics, 2016: 300-309.
- [17]Hongyu Lin, Lu Yaojie, Han Xianpei, et al. Nugget Proposal Networks for Chinese Event Detection[C]. Association for Computational Linguistics, 2018: 1565-1574.
- [18]Ning Ding, Li Ziran, Liu Zhiyuan, et al. Event Detection with Trigger-Aware Lattice Neural Network[C]. Conference on Empirical Methods in Natural Language Processing, 2019: 347-356.
- [19]Jacob Devlin, Chang Ming-Wei, Lee Kenton, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Minneapolis[C], North American Chapter of the Association for Computational Linguistics, 2019: 4171-4186.
- [20]Zhisong Zhang, Kong Xiang, Liu Zhengzhong, et al. A Two-Step Approach for Implicit Event Argument Detection[C]. Annual Meeting of the Association for Computational Linguistics, 2020: 7479-7485.
- [21]Sen Yang, Feng Dawei, Qiao Linbo, et al. Exploring Pre-trained Language Models for Event Extraction and Generation[C]. Annual Meeting of the Association for Computational Linguistics, 2019: 5284-5294.
- [22]Rong Ma, Zeng Ying, Feng Yansong. Scale Up Event Extraction Learning via Automatic Training Data Generation[C]. Thirty-Second AAAI Conference on Artificial Intelligence, 2017: 6045-6052.
- [23]Lifu Huang, Ji Heng, Cho Kyunghyun, et al. Zero-Shot Transfer Learning for Event Extraction[C]. Annual Meeting of the Association for Computational Linguistics, 2018: 2160-2170.
- [24]Sepp Hochreiter, Jürgen Schmidhuber. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [25]Kyunghyun Cho, Bart van Merriënboer, Gulcehre Caglar, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[C]. Conference on Empirical Methods in Natural Language Processing, 2014: 1724-1734.
- [26]Laurent Itti, Christof Koch. Computational modelling of visual attention[J]. Nature Reviews Neuroscience, 2001, 2(3): 194-203.

- [27]冯建周, 宋沙沙, 王元卓, 刘亚坤, 武红颖, 龚昊. 基于改进注意力机制的实体关系抽取方法[J]. 电子学报, 2019, 47(08):1692-1700.
- [28]Xiangnan He, Zhankui He, Jingkuan Song, et al. NAIS: Neural Attentive Item Similarity Model for Recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2018: 2354-2366.
- [29]William Chan, Navdeep Jaitly, Quoc Le, et al. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, 2016: 4960-4964.
- [30]Kelvin Xu, Jimmy Ba, Ryan Kiros, et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention[J]. Computer Science, 2015: 2048-2057.
- [31]Artsiom Ablavatski, Lu Shijian, Cai Jianfei. Enriched Deep Recurrent Visual Attention Model for Multiple Object Recognition[J]. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 2017: 971-978.
- [32]Ashish Vaswani, Shazeer Noam, Parmar Niki, et al. Attention Is All You Need[C]. Advances in Neural Information Processing Systems 30, 2017: 5998-6008.
- [33]Tomas Mikolov, Chen Kai, Corrado Greg, et al. Efficient Estimation of Word Representations in Vector Space[C]. 1st International Conference on Learning Representations, 2013: 296-308.
- [34]Tomas Mikolov, Sutskever Ilya, Chen Kai, et al. Distributed Representations of Words and Phrases and their Compositionality[C]. Conference and Workshop on Neural Information Processing Systems, 2013: 3111-3119.
- [35]Tobias Schnabel, Labutov Igor, Mimno David, et al. Evaluation methods for unsupervised word embeddings[C]. Conference on Empirical Methods in Natural Language Processing, 2015: 298-307.
- [36]Jeffrey Pennington, Socher Richard, Manning Christopher. GloVe: Global Vectors for Word Representation[C]. Conference on Empirical Methods in Natural Language Processing, 2014: 1532-1543.
- [37]Matthew-E Peters, Neumann Mark, Iyyer Mohit, et al. Deep contextualized word representations[C]. North American Chapter of the Association for Computational Linguistics, 2018: 2227-2237.
- [38]Jacob Devlin, Chang Ming-Wei, Lee Kenton, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]. North American Chapter of the Association for Computational Linguistics, 2019: 4171-4186

- [39]Xingchen Li, Mingtong Liu, Yujie Zhang, et al. A Joint Model for Graph-based Chinese Dependency Parsing [C]. The Chinese National Conference on Computational Linguistics, 2020: 820-830

致 谢

在三年的研究生时光内，在最初我从对软件工程所知甚少，通过对理论基础知识的学习和不断实践，最终有了很大提升，完成了毕业设计。首先感谢大连理工大学软件学院给我们提高学习知识和项目实践的机会，让我收获学习很多，不仅有自然语言处理的知识技术，还有严谨的治学态度。

本论文的研究工作是在刘导师的指导下完成的。导师的渊博的专业知识、严谨的治学态度、精益求精的工作作风、诲人不倦的高尚师德、朴实无华平易近人的人格魅力都给我留下了深刻的印象。从论文选题到文章批改，刘导师都以规范、严禁、精确的作风要求精心指点。

同时还要感谢赵老师的指导。在刘导师之外，赵老师为我们精心指导，明确研究方向，批改毕设论文，统筹实验进行。此外，实验室的学长和博士，帮助我的研究生工作生活，将我带入自然语言处理的大门。

最后还要感谢学校的工作人员，在新冠肺炎疫情之下保障了学校的教育教学工作的正常运行，为我们的正常学习研究工作保驾护航，使得我们能够正常如期毕业。

大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目：面向百科数据的军事事件抽取

作者签名：高鸿博 日期：2021年6月5日

大连理工大学学位论文版权使用授权书

本人完全了解学校有关学位论文知识产权的规定，在校攻读学位期间论文工作的知识产权属于大连理工大学，允许论文被查阅和借阅。学校有权保留论文并向国家有关部门或机构送交论文的复印件和电子版，可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

学位论文题目： 面向百科数据的军事事件抽取
作者签名： 高鸿博 日期： 2021 年 6 月 5 日
导师签名： 李 日期： 2021 年 6 月 5 日