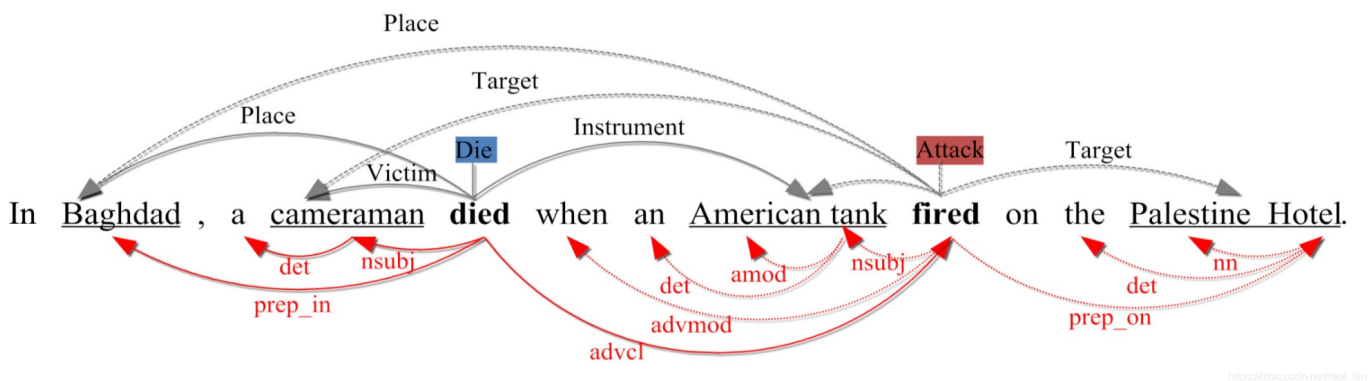


事件抽取

1. 基本概念

事件抽取是从非结构化的文本中提取上述结构化的事件要素信息的自然语言处理任务，可以在各种下游任务，如信息检索（Glavas和S najder，2014）和知识库群体（Ji和Grishman，2011）种发挥重要作用。事件抽取和命名实体识别、关系抽取共同构成了NLP中信息抽取的主要任务。

事件一般是需要包含事件的参与者（角色），并且事件往往可以被描述为一系列状态的改变。在事件抽取的过程中，一个事件往往被更形式化地定义为包含了事件触发器（event trigger），事件类型(event type), 事件元素(event argument) 和事件元素角色(event argument role)，因此事件抽取的任务就是识别出并且进行结构化组织。



在经典的事件定义中，组成事件的各元素包括: 实体、触发词、事件类型、事件属性、论元及论元角色等。其中，事件触发词标识着事件的发生。

- 事件提及（Event mentions）：描述事件的短语或句子包含触发器和相应的参数；
- 实体（Entity）：实体是语义类别中的一个对象或对象组。实体主要包括人、组织、地点、时间、事物等。在图内。2、“巴格达”、“摄影师”、“美国坦克”、“巴勒斯坦饭店”等字样为实体；
- 触发词（Trigger）：表示事件发生的核心词，多为动词或名词；
- 事件类型（Event type）：ACE2005 定义了8种事件类型和33种子类型。其中，大多数事件抽取均采用33 种事件类型。事件识别是基于词的34 类(33类事件类型+None) 多元分类任务，角色分类是基于词对的36 类(35类角色类型+None) 多元分类任务；
- 事件属性：包括事件极性（Polarity）、时态(Tense)，是衡量事件是否真实发生的重要依据。通过极性，事件分为肯定、否定、可能事件。通过时态，事件分为过去发生的事件、现在正在发生的事件、将要发生的事件以及其他无法确定时态的事件；
- 事件论元（Event argument）：事件论元为事件主体(Subject)、客体(Object)、时间(Time)、地点(Location)等，是表达事件重要信息的载体；
- 论元角色（Argument role）：事件论元在事件中充当的角色。共有35类角色，例如，攻击者、受害者等。

比如文本：北京时间 3 月 27 日晚上 7 点 15 分，英国首相鲍里斯约翰逊确诊感染了新冠肺炎。事件触发词为"感染"；事件论元中，主体为英国首相鲍里斯约翰逊，客体为新冠肺炎，时间为北京时间 3 月 27 日晚上 7 点 15 分，地点无；事件属性中，事件极性为肯定，事件时态为现在。

2. 类型

根据事件类别是否预定义，事件抽取可以被分为两大类：限定域抽取（close-domain extraction）和开放域抽取（open-domain extraction）。根据抽取的方法，可以分为基于模式匹配的事件抽取（由词汇、语法等方面的专家知识驱动的方法）和基于机器学习的事件抽取（机器学习驱动的方法）。根据文本粒度，可以分为句子级事件抽取和篇章级事件抽取。

3. 基于深度学习的事件抽取

3.1 事件抽取的四种子任务

经典的事件提取任务可以被分为四个子任务：触发器识别（Trigger Identification）、事件类型分类（Trigger classification）、参数识别（Argument Identification）和参数角色分类（Argument Role Classification）。

- 1) Trigger Identification: 触发器是事件提取中能够清晰表达事件发生的核心单元。触发器标识子任务是从文本中查找触发器；
- 2) Trigger classification: 触发器分类是确定每个句子是否符合现有触发器的事件。此外，如果句子是一个事件，我们需要确定该句子所属的一个或多个事件类型；
- 3) Argument Identification: 参数识别是从文本中识别事件类型中包含的所有参数。自变量识别通常取决于触发器分类和触发器识别的结果；
- 4) Argument Role Classification: 参数角色分类基于事件提取模式中包含的参数，每个参数的类别根据识别的参数进行分类。

3.2 范式

从事件抽取子任务的组织方式来看，现有的事件抽取方法大多可分为两类主流范式：基于流水线的范式（Pipeline-based Paradigm）和基于联合的范式（Joint-based Paradigm）。基于管道的方法利用了分治算法的思想；因此，它的优点是简化了每个子任务，可以为后续的子任务提供信息。相反，它的缺点是传播级联错误，而且总体性能很大程度上依赖于前面的子任务。基于联合的方法独立考虑子任务，不会在子任务之间传播错误。缺点是不能利用之前子任务的信息，需要更大规模的精细标记数据来训练模型。

3.2.1 Pipeline-based Paradigm

基于流水线的方法将所有子任务视为独立的分类问题。所需的分类器包括：

- 1) 触发器分类器用于确定该术语是否为事件触发器和事件类型；
- 2) 自变量分类器用于确定单词是否为事件的自变量；
- 3) 参数角色分类器用于确定参数的类别。

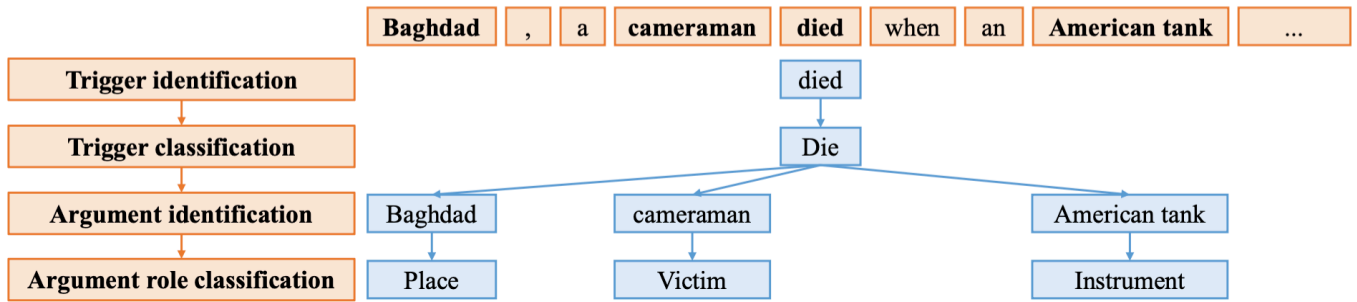


Fig. 4: An example of pipeline-based event extraction.

3.2.2 Joint-based Paradigm

该方法根据第一阶段中的候选触发器和实体来识别触发器和自变量。在第二阶段，为了避免错误信息从事件类型传播，同时实现了触发器分类和自变量角色分类。它将触发器“死亡”划分为死亡事件类型，将自变量“巴格达”划分为放置自变量角色等。

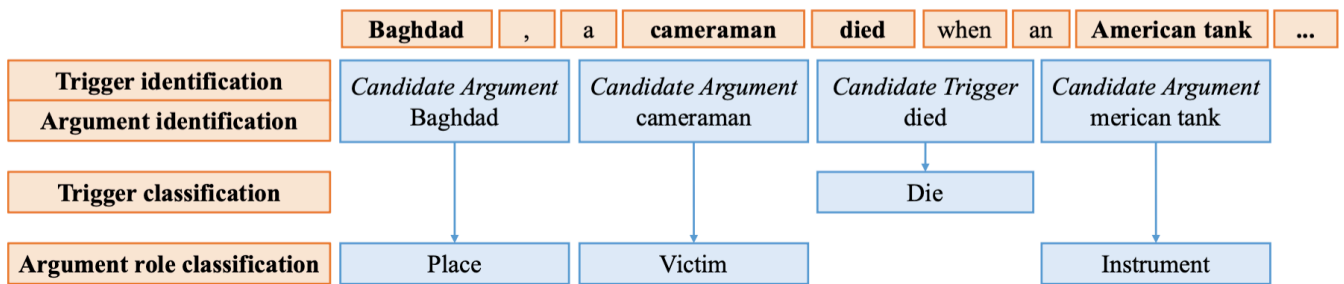


Fig. 5: The simplified architecture of joint-based event extraction paradigm.

3.3 经典的深度学习事件抽取模型

传统的事件提取方法在学习深度特征方面具有挑战性，这使得依赖于复杂语义关系的事件提取任务难以改进。最近的事件提取工作基于深度学习架构，如卷积神经网络（CNN）[39]、[101]、递归神经网络（RNN）[95]、[102]、图神经网络（GNN）[82]、[96]、[103]、[104]、[105]、Transformer[37]、[106]、[107]或其他网络[58]、[78]。深度学习可以捕捉复杂的语义关系，并显著改进多个事件提取数据集。

3.3.1 CNN-based Models

为了在不使用复杂的自然语言处理工具的情况下自动提取词汇和句子级别的特征，Chen等人[39]介绍了一种称为DMCNN的单词表示模型。它捕捉单词的有意义的语义规则，并采用基于CNN的框架来捕捉句子级别的线索。然而，CNN只能捕捉句子中的基本信息，它使用动态多池层来存储基于事件触发器和自变量的更关键的信息。事件提取是一种由具有自动学习特征的动态多池卷积神经网络实现的两阶段多类分类。第一阶段是触发器分类。DMCNN对句子中的每个单词进行分类，以识别触发因素。对于具有触发器的句子，此阶段应用类似的DMCNN为触发器分配参数并调整参数的角色。图6描述了论点分类的体系结构。词汇级特征表示和句子级特征提取用于获取词汇线索和学习句子的组成语义特征。

有线电视新闻网归纳出了句子中k语法的基本结构。因此，一些研究人员还研究了基于卷积神经网络的事件提取技术。Nguyen等人[108]使用CNN来研究事件检测任务，与传统的基于特征的方法相比，该任务克服了复杂的特征工程和错误传播的限制。但它广泛依赖于其他受监督的模块和手动资源来获得特征。在跨域泛化性能方面，它明显优于基于特征的方法。此外，为了考虑非连续k-gram，Nguyen等人[102]引入了非连续CNN。CNN模型通过具有丰富的局部和全局特征的结构化预测应用于基于管道和基于联合的范式，以自动学习隐藏特征表示。与基于管道的方法相比，基于联合的范式可以缓解错误传播问题，并利用事件触发器和参数角色之间的相互依赖性。

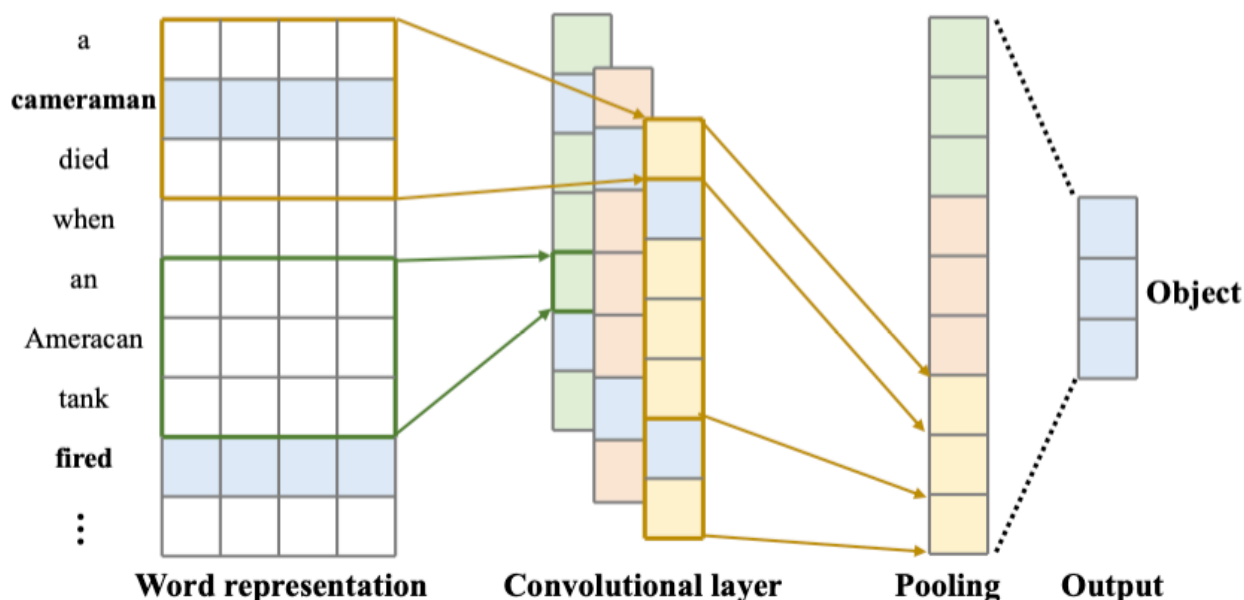


Fig. 6: The architecture of CNN-based argument extraction. It illustrates processing of one instance with predicted trigger 'fired' and candidate argument 'cameraman'.

3.3.2 RNN-based Models

除了基于CNN的事件提取方法外，还对RNN进行了其他一些研究。RNN用于对序列信息进行建模，以提取事件中的痕迹，如图7所示。JRNN[40]是在基于联合的范式中提出的用于事件提取的双向RNN。它有编码阶段和预测阶段。在编码阶段，它使用RNN来总结上下文信息。此外，它还预测了触发和自变量在预测阶段的作用。

以前的方法在很大程度上依赖于特定语言的知识 and 现有的NLP工具。冯等人[109]开发了一种混合神经网络，用于在特定序列和信息片段的背景下捕获信息，并将其用于训练多语言事件检测器。该模型使用双向长短期记忆

(LSTM) 来获得需要识别的文档序列信息。然后使用卷积神经网络获取文档中的短语块信息，将这两种信息结合起来，最终识别触发器。该方法对多种语言（英语、中文和西班牙语）具有鲁棒、高效和准确的检测能力。在跨语言泛化性能方面，复合模型优于传统的基于特征的方法。深度学习中的树结构和序列结构比序列结构具有更好的性能。为了避免过度依赖词汇和句法特征，依赖桥递归神经网络 (DBRNN) [95] 基于双向RNN进行事件提取。

DBRNN是通过桥接语法相关的单词来增强的。DBRNN是一个基于RNN的框架，它利用依赖图信息来提取事件触发器和参数角色。

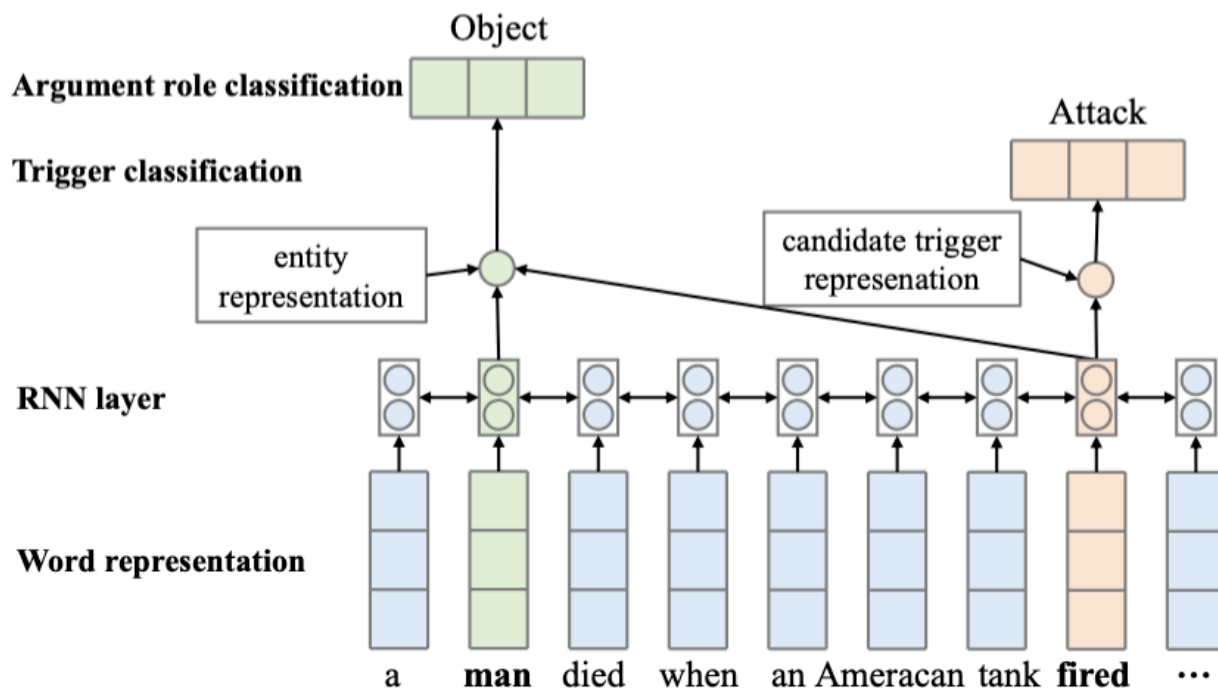


Fig. 7: The simplified architecture of RNN-based argument extraction for the input sentence “a man died when a tank fired in Baghdad” for candidate trigger “fired”.

3.3.3 Attention-based Models

利用深度学习模型自动提取事件特征和利用外部资源增强事件特征主要关注事件触发器的信息，较少关注事件自变量和词间依赖关系的信息。句子级序列建模在捕获非常长范围的依赖关系方面效率很低。此外，基于RNN和基于CNN的模型并不能完全模拟事件之间的关联。结构信息在保持机制中的建模逐渐引起了研究人员的注意。随着研究方法的不断提出，添加注意力机制的模型逐渐出现，如图8所示。注意力机制的特征决定了它可以在不考虑位置信息的情况下使用全局信息来对局部上下文进行建模。它在更新词的语义表示时具有良好的应用效果。

通过控制句子每个部分的不同权重信息，注意力机制使模型关注句子的重要特征信息，而忽略其他不重要的信息，并合理分配资源以提取更准确的结果。同时，注意力机制本身可以作为一种对齐，解释端到端模型中输入和输出之间的对齐，使模型更具可解释性。

一些研究人员还使用分层注意力机制来进行信息的全局聚合。联合多事件提取 (JMEE) [96]由四个模块组成：单词表示、句法图进化网络、自注意触发分类和自变量分类模块。通过引入语法快捷弧来增强信息流。使用基于注意力的图卷积网络对图信息进行联合建模，以提取多个事件触发器和自变量。此外，当联合提取事件触发器和自变量以解决数据集失衡时，它优化了有偏损失函数。

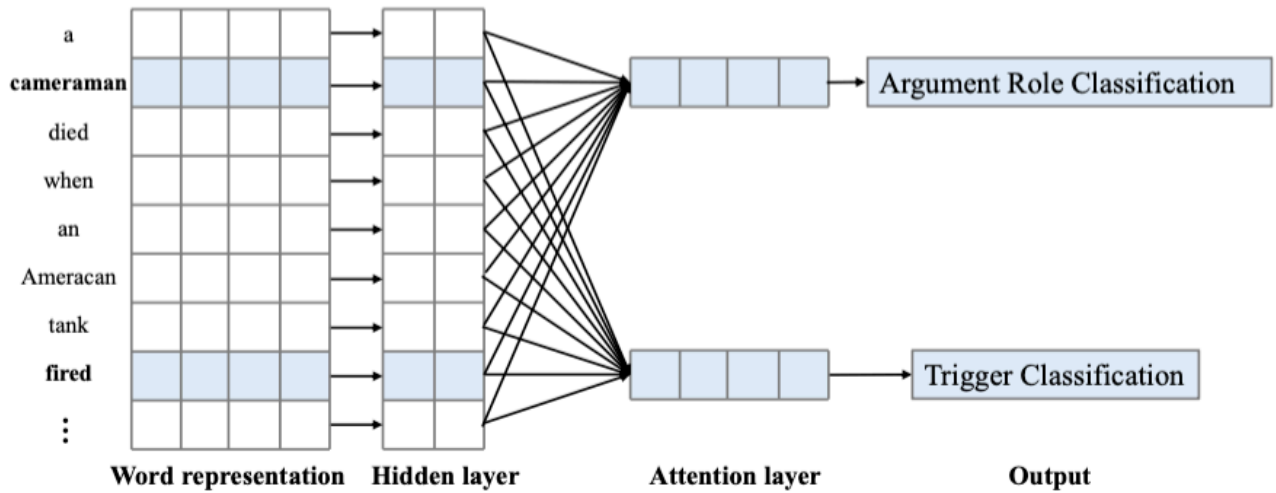


Fig. 8: The architecture of attention-based event extraction.

3.3.4 Graph Convolutional Network-based (GCN-based) Models

句法表征提供了一种有效的方法，可以将单词直接链接到其信息上下文，用于句子中的事件检测[33]，[82]，[110]，[111]。Nguyen等人[110]研究了基于依赖树的卷积神经网络来执行事件检测，是第一个将依赖树关系信息集成到神经事件检测中的人。该模型将所提出的模型与图卷积网络（GCN）[111]和基于实体提及的池化一起使用。他们提出了一种新的池化方法，该方法依赖于实体提及来聚合卷积向量。该模型对当前单词和句子中提到的实体的基于图的卷积向量进行池化。该模型聚合卷积向量以生成用于事件类型预测的单个向量表示。该模型是对来自实体提及的信息进行显式建模，以提高事件检测的性能。

在[77]中，文本分析会议知识库群体（TAC-KBP）时隙用于填充任务中提出的四元时间表示，该模型预测事件的最早和最晚开始和结束时间，从而表示事件的模糊时间跨度。该模型基于共享的参数和时间关系为每个输入文档构建了一个文档级事件图，并使用基于图的注意力网络方法在图上传播时间信息，如图9所示。实体用下划线表示，事件用粗体表示。Wen等人构造了一种基于输入文档事件-事件关系的文档级事件图方法。将提取文档中的事件参数。然后，根据诸如“之前”和“之后”之类的关键字以及事件发生的时间逻辑，按时间顺序排列事件。实体参数在不同的事件之间共享。模型实现将事件合并到一个更准确的时间表中。

The enemy have *now* been **flown out** and we're treating them including a man who is almost dead with a **gunshot wound** to the chest after we (Royal Marines) **sent** in one of our companies of about 100 men in here (Umm Kiou) *this morning*.

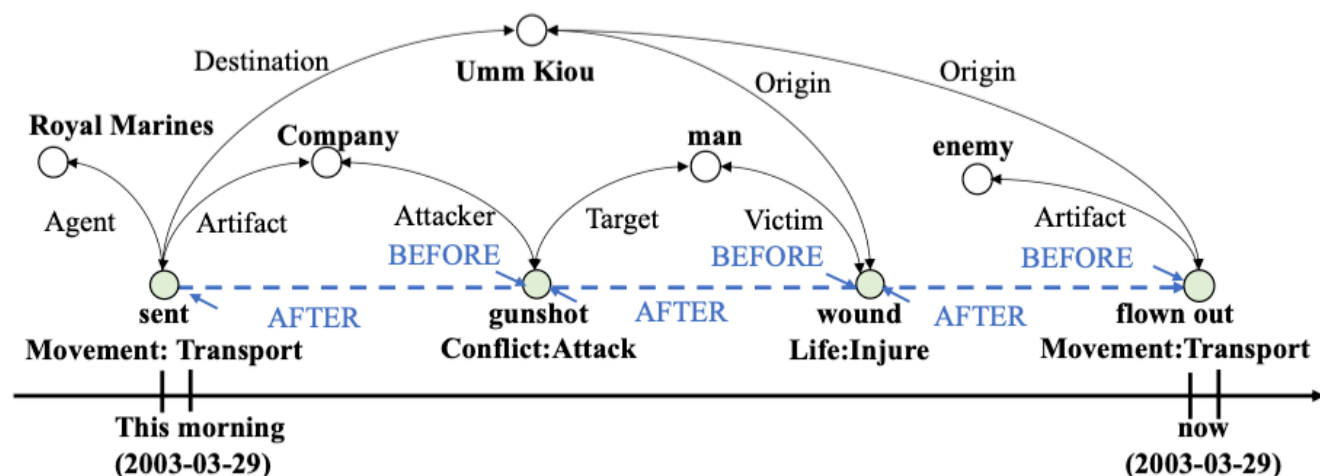


Fig. 9: The example event graph. The solid line consists of event arguments, while the dashed line graph is constructed based on time relationships [77].

3.3.5 Transformer-based Models

利用在各种事件中扮演不同角色的一个论点来改进事件提取是具有挑战性的。杨等人[37]采用了一种根据自变量角色来分离自变量预测的方法，以克服角色重叠问题。此外，由于训练数据不足，该方法通过编辑原型和通过对质量进行排序来筛选开发的样本，从而自动生成标记数据。他们提出了一个框架，即基于预训练语言模型的事件提取器 (PLMEE) [37]，如图 10。PLMEE通过使用提取模型和基于预先训练的语言模型的生成方法的组合来促进事件提取。它是一个两阶段的任务，包括触发器提取和参数提取，由触发器提取器和参数提取器组成，两者都依赖于BERT的特征表示。然后利用角色的重要性重新加权损失函数。

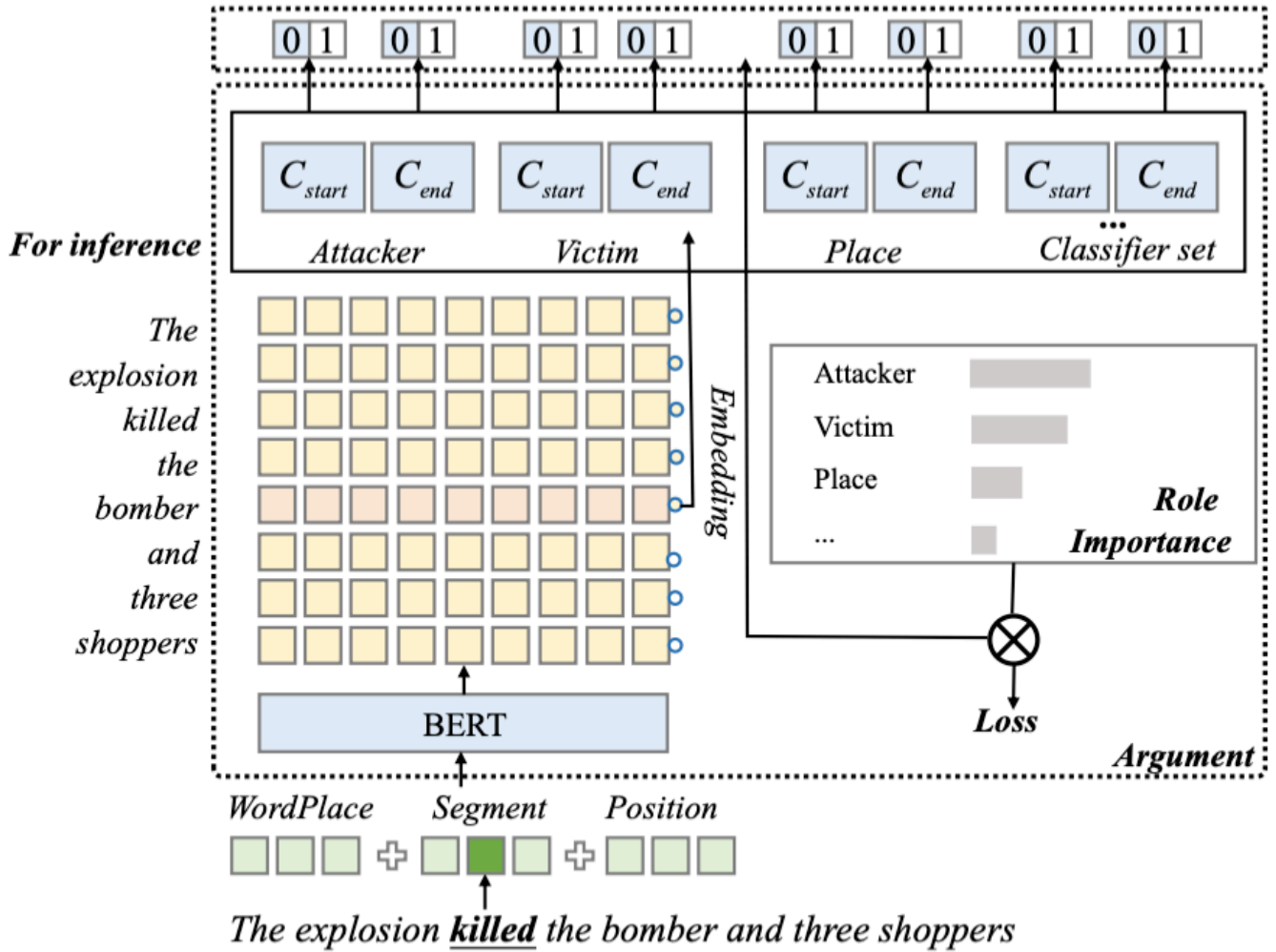


Fig. 10: The architecture of PLMEE [37] for extraction.

GAIL[84]是一种基于ELMo的[112]模型，利用生成对抗性网络来帮助模型关注更难检测的事件。他们提出了一种基于生成对抗性模拟学习的实体和事件提取框架。这是一种使用生成对抗性网络（GAN）的反向强化学习（IRL）方法。该模型通过使用IRL的动态机制，直接评估实体和事件提取中实例的正确和不正确标记。

DYGIE++[85]是一个基于BERT的框架，对句子和跨句子上下文中的文本跨度和捕获进行建模。许多信息提取任务，如命名实体识别、关系提取、事件提取和共同参考解析，都可以受益于句子之间的全局上下文或不依赖于局部的短语。它们将事件提取作为附加任务，并在事件触发器及其参数的关系图中进行跨度更新。跨度表示是在多语句BERT编码的基础上构建的。

总之，大多数传统的事件提取方法都采用人工构建的方法进行特征表示，并使用分类模型对触发器进行分类和识别自变量的作用。近年来，深度学习在图像处理、语音识别和自然语言处理等方面表现出了突出的效果。为了解决传统方法的不足，系统地讨论了基于深度学习的事件提取。在BERT模型出现之前，主流的方法是从文本中找到触发器，并根据触发器判断文本的事件类型。最近，随着BERT引入事件提取模型，基于全文的事件类型识别方法成为主流。这是因为BERT具有出色的上下文表示能力，在文本分类任务中表现良好，尤其是在只有少量数据的情况下。

4. 最新进展

4.1 低资源场景的事件抽取

由于数据出现中昂贵的注释，通常只有不足够的数据来用完全监督的方法训练准确的EE模型。在这项调查中，我们将这种情况称为资源匮乏。为了缓解数据稀疏性，现有研究探索了远程监督[4]方法来增强注释数据。此外，一些有前途的方法还研究了半监督方法[5]、[6]、[7]或多语言方法，以丰富监督信息。最近，一些研究在三种典型的低资源环境中探索了能效，包括few shots学习环境、零样本学习环境和增量学习环境。

4.1.1 Few shots学习场景

最近，few shots学习方法[8]，[9]在几种NLP任务中得到了广泛研究[10]，[11]，其目的是在极其有限的（few shots，如1 shot、3 shots等）观察到的训练样本下进行任务预测。在事件提取领域，现有的研究大多集中在few shots学习环境中应用的事件检测子任务上。第一类工作[12]，[13]，[14]，[15]，[16]，[17]旨在利用元学习方法在给定的候选触发器的情况下进行触发器分类。为了接近真实应用，第二行工作[18]、[19]、[20]仅使用纯文本数据联合进行触发识别和触发分类。其中，Cong等人[19]进一步学习了具有原型摊销条件随机场（CRF）的稳健序列标签转换分数，实现了显著的改进。

4.1.2 Zero-shot 学习场景

以前的大多数监督EE方法都依赖于从手动注释派生的功能，这些功能无法处理没有附加注释的新事件类型。一个极具挑战性的低资源场景是在没有任何可用标记数据的情况下实现EE。为了研究这种情况的可能性，最近的研究[61]、[138]探索了用于事件提取的零样本学习（ZSL）。Huang等人[61]首先解决了这个问题，利用事件提及和类型的结构本体进行表示，并通过语义相似性测量进行预测。Lyu等人[138]进一步研究了新事件的迁移学习方法，该方法将EE公式化为文本隐含（TE）和问答（QA）查询（例如，“城市受到攻击”意味着“有攻击”），并利用预先训练的TE/QA模型进行直接迁移。尽管这些方法与监督方法仍有很大差距，但它们揭示了一个深刻的愿景，并为资源极低的EE提供了可能的改进方向。

4.1.3 增量学习场景

以前的大多数监督EE方法都依赖于从手动注释派生的功能，这些功能无法处理没有附加注释的新事件类型。一个极具挑战性的低资源场景是在没有任何可用标记数据的情况下实现EE。为了研究这种情况的可能性，最近的研究[61]、[138]探索了用于事件提取的零样本学习（ZSL）。Huang等人[61]首先解决了这个问题，利用事件提及和类型的结构本体进行表示，并通过语义相似性测量进行预测。Lyu等人[138]进一步研究了新事件的迁移学习方法，该方法将EE公式化为文本隐含（TE）和问答（QA）查询（例如，“城市受到攻击”意味着“有攻击”），并利用预先训练的TE/QA模型进行直接迁移。尽管这些方法与监督方法仍有很大差距，但它们揭示了一个深刻的愿景，并为资源极低的EE提供了可能的改进方向。

4.2 开放域事件抽取

在没有预定义事件模式的情况下，开放域事件提取被设计为从文本中检测事件，并且在大多数情况下，通过提取的事件关键字对类似事件进行聚类。事件关键字是那些主要描述事件的单词/短语，有时进一步分为触发器和参数。开放域事件提取[16]、[17]、[18]没有固定的参数角色模板。因此，论点往往是通过提取关键词来获得的。周等人[16]提出了一种基于公共新闻标题过滤不相关标题并进行初步事件提取的方法。价格和文本都被反馈到3D卷积神经网络中，以学习事件和市场走势之间的相关性。刘等人[17]使用无监督生成方法设计了一种新的潜在变量神经模型，以探索潜在事件类型向量和实体提及冗余。实验结果表明，它可以扩展到非常大的语料库。

5. 数据集

用于事件提取的标记数据集的可用性已成为快速发展背后的主要驱动力。在本节中，我们将总结这些数据集。

5.1 文档级

5.1.1 MUC-4

MUC-4是在第四次消息理解会议（1992）中提出的。MUC4是一个广泛使用的事件模式规划和事件抽取数据集，该数据集由1700个文档组成，定义了4种事件类型(Arson, Attack, Bombing, Kidnapping)和4种论元角色类型(Perpetrator, Instrument, Target, Victim)。

5.1.2 Google

谷歌数据集是全球事件、语言和语气数据库（GDELT）的一个子集事件数据库，事件相关单词检索包含11909篇新闻文章的30种事件类型的文档。

5.1.3 Twitter

推特数据集收集自2010年12月发布的推特，应用推特流媒体应用程序编程接口（API），包括20种事件类型和1000条推特。

5.1.4 NO.ANN, NO.POS, NO.NEG (DCFEE)

Yang 等人（2018）对四种类型的金融事件进行了实验：股权冻结事件、股权质押事件、股权回购事件和股权超重事件。通过自动生成数据，总共标记了2976个公告。NO.ANN表示可以为每个事件类型自动标记的公告数量。NO.POS代表正面样提及的总数。相反，NO.NEG代表负面提及的数量。

5.1.5 ChFinAnn (Doc2EDAG)

Zheng（2019）基于10年的ChFinAnn4文件和人类总结的事件知识库进行了基于远程监督（DS）的事件标记。这个新的中文股权收购事件数据集包括32040份文件和5种事件类型：股权冻结、股权回购、股权减持、股权增持和股权质押。

5.1.6 RAMS

Eber等人发布了“跨多个句子的角色”（Roles Across Multiple Pences, RAMS）。该任务旨在从5个语句窗口中识别给定事件触发器的事件自变量。该数据集包含3194个文档，其中9124个事件是基于139个事件类型和65个角色的本体从新闻中注释的。

5.1.7 WIKIEVENTS

WIKIEVENTS是由李等人（2021）发布的一个文档级基准数据集。该数据集收集自维基百科上描述真实世界事件的英文文章。

5.2 句子层次

5.2.1 Automatic Content Extraction (ACE)

ACE 2005是事件提取中使用最广泛的数据集。它包含一套完整的英文、阿拉伯文和中文培训数据，用于ACE 2005技术评估。语料库由语言数据联盟（LDC）为实体、关系和事件注释的各种类型的数据组成。它包括599个文档，包含8个事件类型、33个事件子类型和35个参数角色。其中，事件识别是基于词的34类(33类事件类型+None)多元分类任务，角色分类是基于词对的36类(35类角色类型+None)多元分类任务

5.2.2 Text Analysis Conference Knowledge base Filling (TAC KBP)

作为KBP中的一个独立组件任务，TAC KBP事件跟踪（从2015年到2017年）的目标是提取有关事件的信息，以便将其输入到知识库中。TAC KBP 2015用英语定义了9种不同的事件类型和38种事件亚型。TAC-KBP 2016和TAC-KBP 2017有三种语言的语料库：英语、汉语和西班牙语，它们拥有8种事件类型和18种事件亚型。

5.2.3 Rich ERE

它扩展了实体、关系和事件本体，并扩展了什么是可标记的概念。Rich ERE还引入了事件跳跃的概念，以解决事件共同引用的普遍挑战，特别是在文档内和文档之间的事件引用以及事件参数的粒度变化方面，为创建（分层或嵌套）跨文档的事件表示铺平了道路。

5.2.4 FSED

FSED数据集基于ACE 2005和TAC KBP 2017，是一个专门为few-shot场景定制的生成数据集。它包含70852次提及，涉及19种事件类型和100个事件子类型。

5.2.5 GNBusiness

GNBusiness从谷歌商业新闻收集新闻报道，以描述来自不同来源的每一个事件。从2018年10月17日至2019年1月22日，共获得55618篇商业文章，共288批13047个新闻集群。全文语料库以GNBusinessFullText的形式发布。

5.2.6 FSD

第一个故事检测（FSD）数据集是一个包括2499条推文的故事检测数据集。考虑到几个样本中提到的事件通常并不重要，研究人员过滤掉了少于15个样本中提及的事件。它包括2453条推文和20种事件类型。

5.2.7 FBI dataset

联邦调查局的都市级仇恨犯罪报告（FBI）数据集是通过抓取Patch“火灾与犯罪”类别中约37万篇未标记的新闻文章而建立的。它包含两个类别，用于对文本中是否存在特定的仇恨犯罪进行分类。此外，它还标注了仇恨犯罪文章的属性。

由监督学习训练得到的开放域事件抽取模型很难识别到大范围的事件类型，且监督学习所使用的人工标注数据成本较高。为了利用现有的未标记文本数据，有研究提出利用知识库自动生成标记数据的方法。survey讨论了两种生成标签数据的方法：通过识别key argument和key trigger来生成数据；二是仅通过识别key argument生成数据。

对比学习

1. 概念

对比学习（Contrastive Learning）是由Hadsell等人（2006）发起的，它是一种自监督学习，需要从无标注的图像数据中学习特征表示，并用于下游任务中。其指导原则是：通过自动构造相似实例和不相似实例，学习一个表示学习模型，通过这个模型，使得相似的实例在投影空间中比较接近，而不相似的实例在投影空间中距离比较远。即将样例与与它语义相似的例子（正样例）和与它语义不相似的例子（负样例）进行对比，希望通过设计模型结构和对比损失，使语义相近的例子对应的表示在表示空间更接近，语义不相似的例子对应的表示距离更远，以达到类似聚类的效果，见图1。

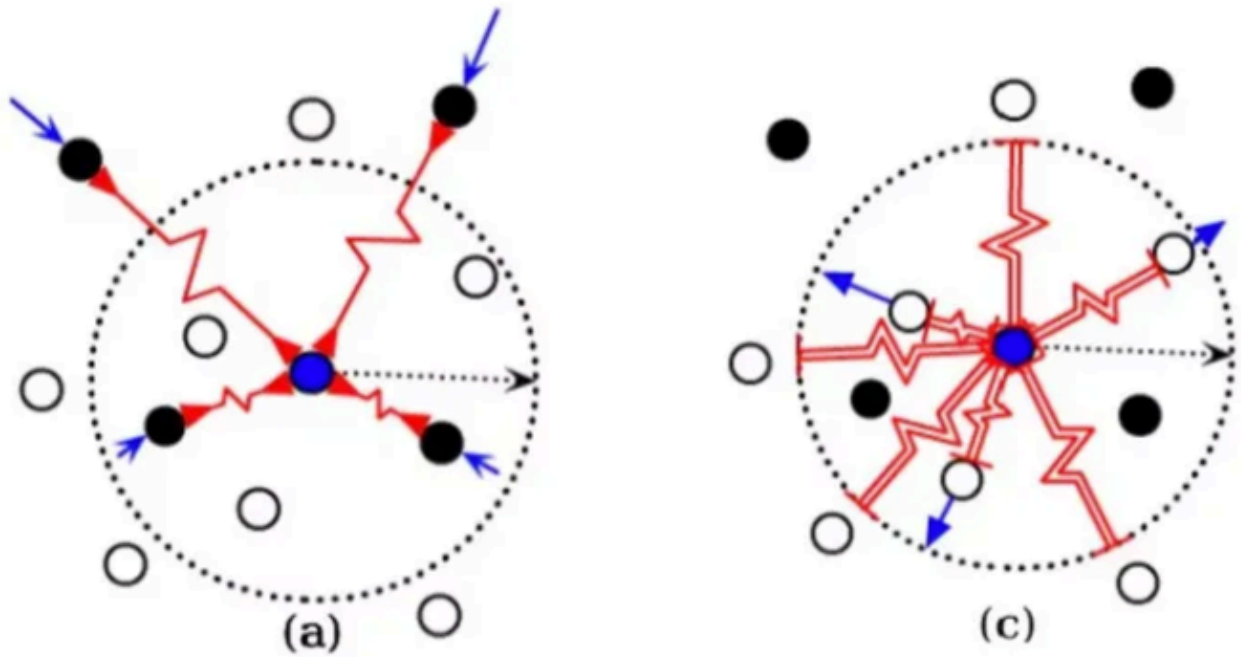


图1 对比学习中正负样例（黑球为正样例，白球为负样例）[8]

以图像为例，见图2，对比学习期望通过使同一类狗不同角度照片的表示相近，而不同种类动物对应的表示距离相远，使得学到的表示可以忽略掉角度（或光影等）变换带来的细节变动，进而学习到更高维度、更本质的特征（语义）信息。

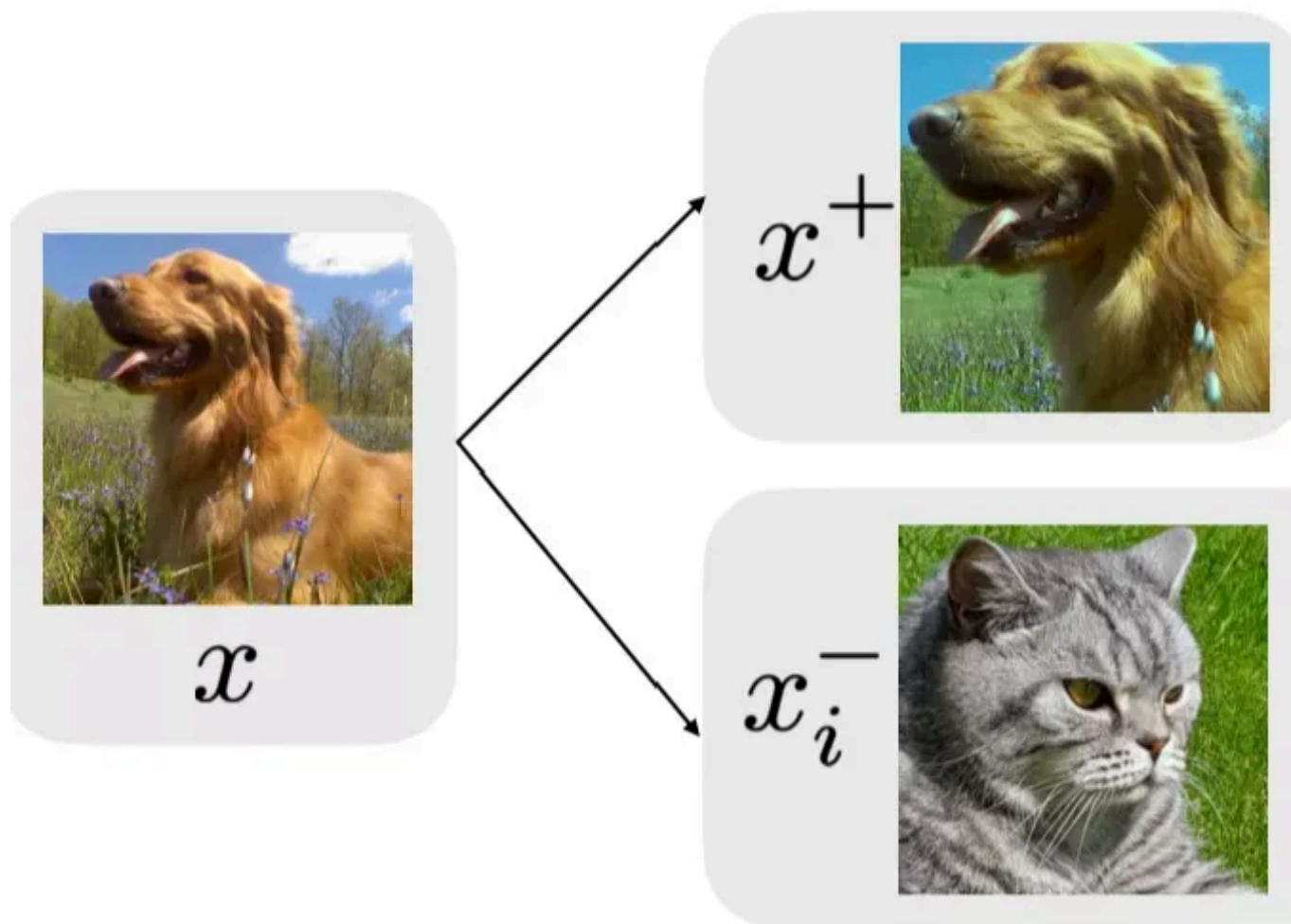


图2 一张狗的图片对应的正负样例[9]

这里我们引入两个重要的概念，对齐性(alignment)和均匀性(uniformity)[10]。由于对比学习的表示一般都会正则化，因而会集中在一个超球面上。对齐性和均匀性指的是好的表示空间应该满足两个条件：一个是相近样例的表示尽量接近，即对齐性；而不相近样例的表示应该均匀的分布在超球面上，即均匀性。满足这样条件的表示空间是线性可分的，即一个线性分类器就足以用来分类，因而也是我们希望得到的，我们可以通过这两个特性来分析表示空间的好坏。

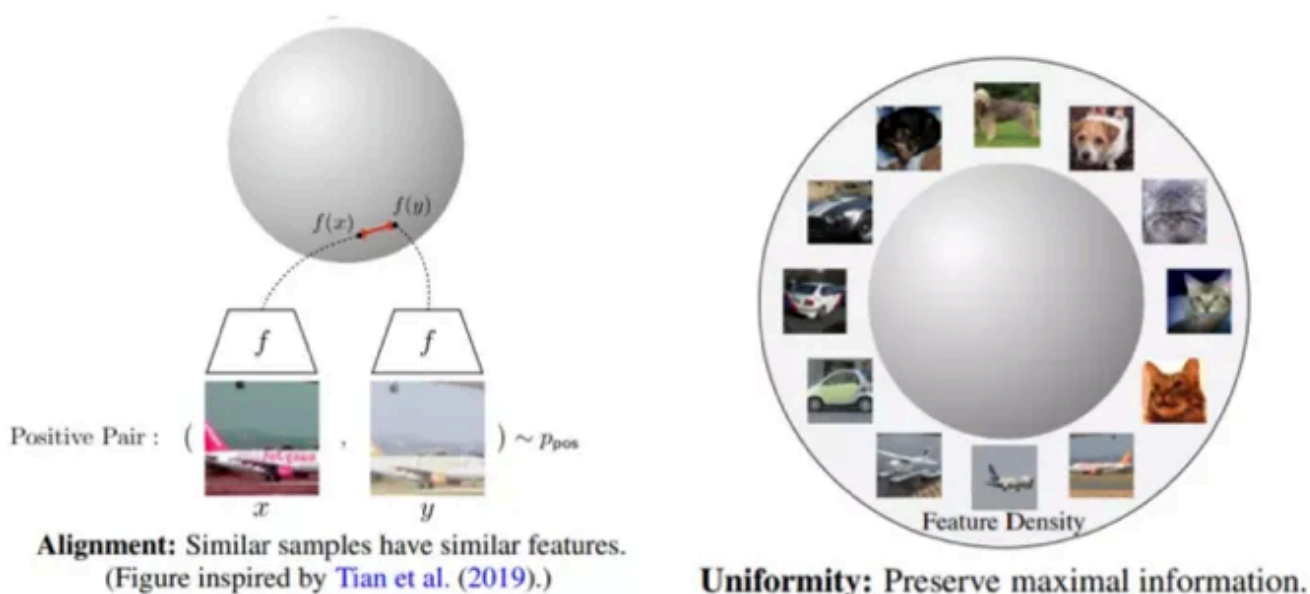


图3 对齐性与均匀性[10]

从上文的描述中我们可以轻易地发现对比学习有三个重要的组成部分：正负样例、对比损失以及模型结构。模型结构将在第三章介绍具体工作时详解介绍，这一章我们先来聊一聊正负样例和对比损失。

如上所述，正样例指的是与给定样例语义相近的样例，而负样例指的是与给定样例语义不相近的样例。对于有监督的数据，正负样例很容易构造，同一标签下的例子互为正样例，不同标签下的例子互为负样例，但对于无标签的数据，我们需要通过一定的方法得到正负样例。

自然语言领域的扰动大致可分为两类：词级别（token-level）和表示级别（embedding-level）。词级别的扰动大致有句子剪裁（crop）、删除词/词块（span）、换序、同义词替换等。表示级别的扰动包括加高斯噪声、dropout等。见图5。

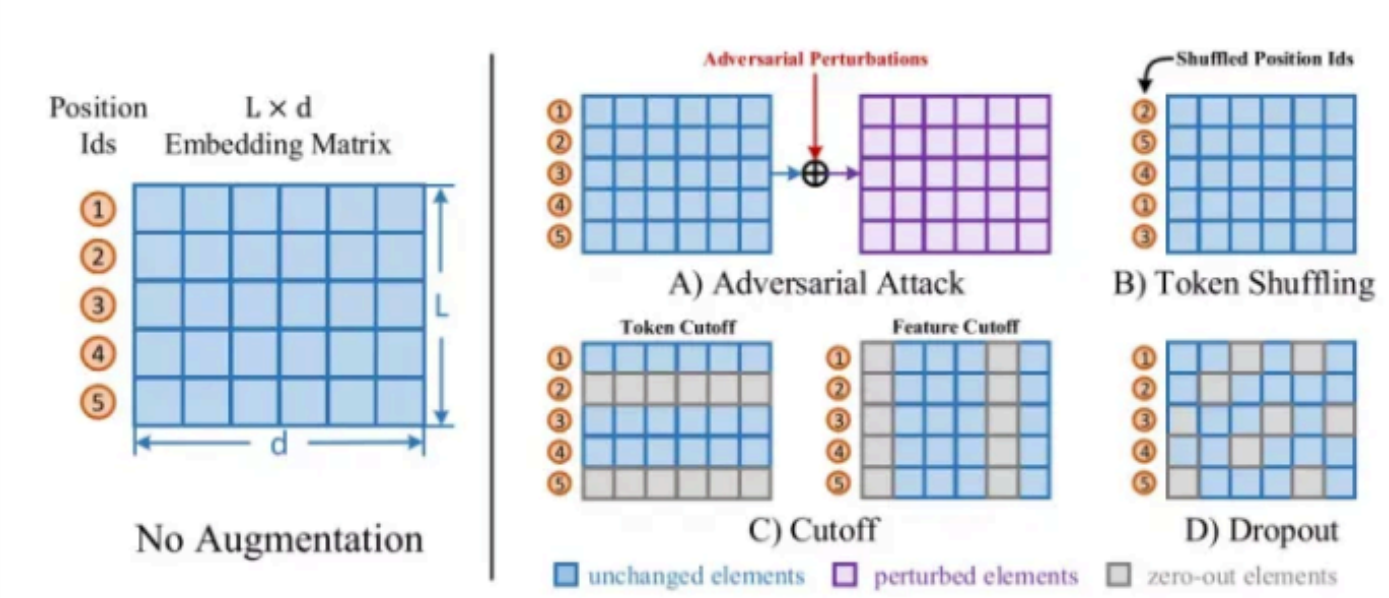


图5 自然语言中的一些增强方法[5]

但是不同于图像领域，对自然语言的扰动很容易改变语义，这就会引入错误正例（False Positive）从而对模型训练产生不好的影响。同时，错误负例（False Negative）也是一个不容忽视的问题，见图6。文献[9]的作者证实了错误负例会对对比学习产生负面影响，并提出了一种对比损失去抵消这种影响。

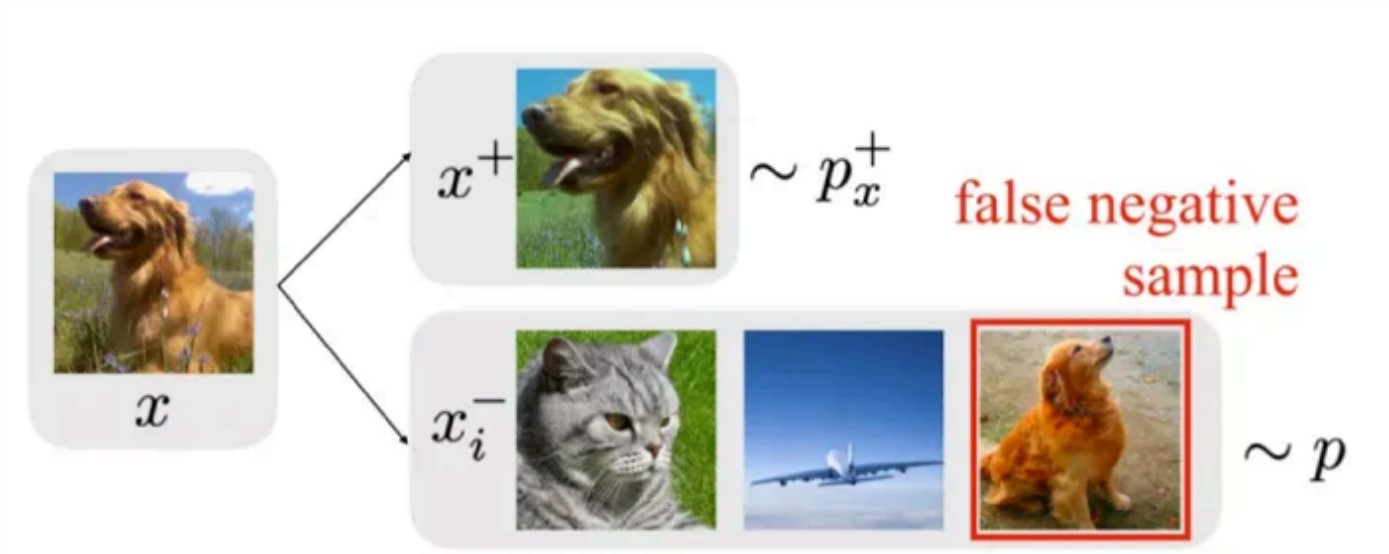


图6 错误负例的例子[9]

在有了正负例之后，我们需要给模型信号，激励他们拉近正样例间的距离，拉远负样例间的距离, 这就要通过设计对比损失来完成。常用的对比损失有普通对比损失、三元组损失（triplet loss）、InfoNCE损失（Noise Contrastive Estimation）等。

1.1 普通对比损失

$$\mathcal{L}((x_1, x_2), y) = yL_p(D(f_\theta(x_1), f_\theta(x_2))) + (1 - y)L_n(D(f_\theta(x_1), f_\theta(x_2)))$$

1.2 三元组损失 (triplet loss)

Triplet loss 最初是FaceNet[12]做人脸表示学习时提出的：给定一个三元组 (x, x^+, x^-) , x 被称做锚点(Anchor), x^+ 为正例, x^- 为负例。triplet loss设计的初衷也是希望 通过对比使得锚点与正例的距离更近，与负例更远，具体形式为：

$$\mathcal{L}(x, x^+, x^-) = \max(0, D(x, x^+)^2 - D(x, x^-)^2 + m)$$

1.3 InfoNCE损失 (Noise Contrastive Estimation)

NCE全称是噪声对比估计（Noise Contrastive Estimation），是文献[13]中提出的一种通过引入一个噪声分布，解决多分类问题softmax分母归一化中分母难以求值的问题。具体做法是把多分类问题通过引入一个噪音分布变成一个二元分类问题，将原来构造的多分类分类器（条件概率）转化为一个二元分类器，用于判别给定样例是来源于原始分布还是噪声分布，进而更新原来多元分类器的参数。听起来好像跟对比学习没什么关系，但如果把噪音分布的样本想成负样例，那这个二元分类问题就可以理解为让模型对比正负样例作出区分进而学习到正样例(原始分布)的分布特征，这即体现了对比学习的思想。

$$\mathcal{L}_q(u, v^+, \{v^-\}) = -\log \frac{\exp(u \cdot v^+ / \tau)}{\sum_{v \in \{v^+, v^-\}} \exp(u \cdot v / \tau)}$$

2. 应用

对比学习可以被广泛用于各种行为中的自我监督表征学习，如计算机视觉（Wu et al., 2018; Oord等人, 2018; Hjelm等人, 2019; 陈等人, 2020; 何等人, 2020）和图形（Qiu et al., 2020; 游等, 2020; 朱等人, 2020年）。在NLP的文本中，许多已建立的表征学习作品可以被视为对比学习方法，如Word2Vec（Mikolov et al., 2013）、BERT（Devlin et al., 2019; Kong et al., 2020）和ELECTRA（Clark等人, 2020）。

对比学习也被广泛用于帮助特定任务，包括问答（Yeh和Chen, 2019）、话语建模（Iter et al., 2020）、自然语言推理（Cui et al., 2019）和关系提取（Peng et al., 2021）。其中，Wang等人（2021）探索了一种用于事件抽取任务的对比预训练框架CLEVE，它由两个部分组成：事件语义预训练和事件结构预训练。该方法更好地利用了大型无监督数据中丰富的事件知识。在两个数据集上的实验表明，CLEVE可以在有监督和无监督事件抽取任务中实现显著改进。但是，这种方法没有探索其他类型的语义结构，如框架语义，以及无法克服语义解析器带来的无监督数据中的噪声。

References

EE

- [1] M. T. Chau, D. Esteves, and J. Lehmann, "Open-domain event extraction and embedding for natural gas market prediction," CoRR, 2019.
- [2] X. Liu, H. Huang, and Y. Zhang, "Open domain event extraction using neural latent variable models," in ACL, 2019.
- [3] M. Mejr and J. Akaichi, "A survey of textual event extraction from social networks," in LPKM, 2017.
- [4] H. Yang, Y. Chen, K. Liu, Y. Xiao, and J. Zhao, "DCFEE: A document-level chinese financial event extraction system based on automatically labeled training data," in ACL, 2018.
- [5] J. Ferguson, C. Lockard, D. S. Weld, and H. Hajishirzi, "Semi-supervised event extraction with paraphrase clusters," in NAACL-HLT, 2018.
- [6] Y. Zhou, Y. Chen, J. Zhao, Y. Wu, J. Xu, and J. Li, "What the role is vs. what plays the role: Semi-supervised event argument extraction via dual question answering," in AAAI, 2021.
- [7] L. Huang and H. Ji, "Semi-supervised new event type induction and event detection," in EMNLP, 2020.
- [8] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al., "Matching networks for one shot learning," Advances in neural information processing systems, 2016.
- [9] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in International Conference on Machine Learning, PMLR, 2017.
- [10] J. Sheng, S. Guo, Z. Chen, J. Yue, L. Wang, T. Liu, and H. Xu, "Adaptive attentional network for few-shot knowledge graph completion," in EMNLP, 2020.
- [11] Z. Ye and Z. Ling, "Multi-level matching and aggregation network for few-shot relation classification," in ACL, 2019.

- [12] V. D. Lai, F. Dernoncourt, and T. H. Nguyen, "Exploiting the matching information in the support set for few shot event classification," *Advances in Knowledge Discovery and Data Mining*, 2020.
- [13] V. D. Lai, F. Dernoncourt, and T. H. Nguyen, "Extensively matching for few-shot learning event detection," *CoRR*, 2020.
- [14] S. Shen, T. Wu, G. Qi, Y. Li, G. Haffari, and S. Bi, "Adaptive knowledge-enhanced bayesian meta-learning for few-shot event detection," in *ACL/IJCNLP, Findings of ACL*, 2021.
- [15] V. Lai, F. Dernoncourt, and T. H. Nguyen, "Learning prototype representations across few-shot tasks for event detection," in *EMNLP*, 2021.
- [16] J. Zheng, F. Cai, W. Chen, W. Lei, and H. Chen, "Taxonomy-aware learning for few-shot event detection," in *WWW*, 2021.
- [17] V. D. Lai, M. V. Nguyen, T. H. Nguyen, and F. Dernoncourt, "Graph learning regularization and transfer learning for few-shot event detection," in *ACM SIGIR*, 2021.
- [18] S. Deng, N. Zhang, J. Kang, Y. Zhang, W. Zhang, and H. Chen, "Meta-learning with dynamic-memory-based prototypical network for few-shot event detection," in *WSDM*, 2020.
- [19] X. Cong, S. Cui, B. Yu, T. Liu, Y. Wang, and B. Wang, "Few-shot event detection with prototypical amortized conditional random field," *CoRR*, 2020.
- [20] J. Chen, H. Lin, X. Han, and L. Sun, "Honey or poison? solving the trigger curse in few-shot event detection via causal intervention," in *EMNLP*, 2021.

Fourth Message Understanding Conference (MUC-4): Proceedings of a Conference Held in McLean, Virginia, June 16-18, 1992, 1992.

H. Yang, Y. Chen, K. Liu, Y. Xiao, and J. Zhao, "DCFEE: A document-level chinese financial event extraction system based on automatically labeled training data," in *ACL*, 2018.

S. Li, H. Ji, and J. Han, "Document-level event argument extraction by conditional generation," in *ACL*, 2021.

S. Zheng, W. Cao, W. Xu, and J. Bian, "Doc2edag: An end-to-end document-level framework for chinese financial event extraction," in *EMNLP-IJCNLP*, 2019.

CL

- [1] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
- [2] He, Kaiming, et al. "Momentum contrast for unsupervised visual representation learning." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
- [3] Chen, Xinlei, et al. "Improved baselines with momentum contrastive learning." *arXiv preprint arXiv:2003.04297* (2020).
- [4] Chen, Xinlei, Saining Xie, and Kaiming He. "An empirical study of training self-supervised visual transformers." *arXiv preprint arXiv:2104.02057* (2021).
- [5] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *International conference on machine learning*. PMLR, 2020.

- [6] Chen, Ting, et al. "Big self-supervised models are strong semi-supervised learners." arXiv preprint arXiv:2006.10029 (2020).
- [7] Contrastive Self-Supervised Learning <https://ankeshanand.com/blog/2020/01/26/contrastive-self-supervised-learning.html>
- [8] Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114 (2013).
- [9] Goodfellow, Ian J., et al. "Generative adversarial networks." arXiv preprint arXiv:1406.2661 (2014).
- [10] Caron, Mathilde, et al. "Unsupervised learning of visual features by contrasting cluster assignments." arXiv preprint arXiv:2006.09882 (2020).
- [11] Grill, Jean-Bastien, et al. "Bootstrap your own latent: A new approach to self-supervised learning." arXiv preprint arXiv:2006.07733 (2020).
- [12] Chen, Xinlei, and Kaiming He. "Exploring Simple Siamese Representation Learning." arXiv preprint arXiv:2011.10566 (2020).
- [13] Gao, Tianyu, Xingcheng Yao, and Danqi Chen. "SimCSE: Simple Contrastive Learning of Sentence Embeddings." arXiv preprint arXiv:2104.08821 (2021).