



中国科学院大学
University of Chinese Academy of Sciences



知识图谱关键技术与问答系统

李航航

2019年3月

提纲

▶ 知识图谱关键技术

- 实体、关系识别与抽取
- 知识融合（实体链接、知识合并、知识更新）
- 知识推理
- 知识图谱存储与可视化

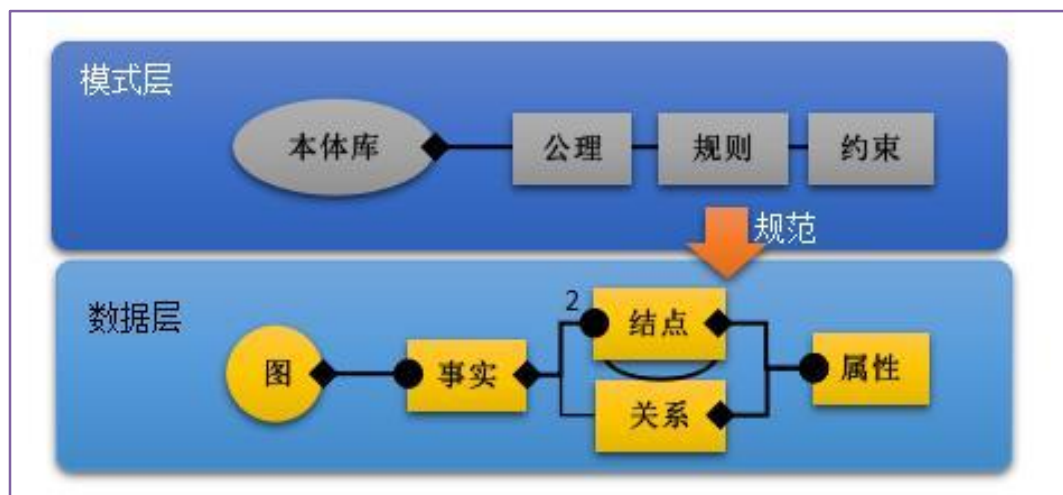
▶ 事理图谱

▶ 应用：基于KG的问答系统

- 处理流程
- 基于KG的优势
- 关键技术
- 应用

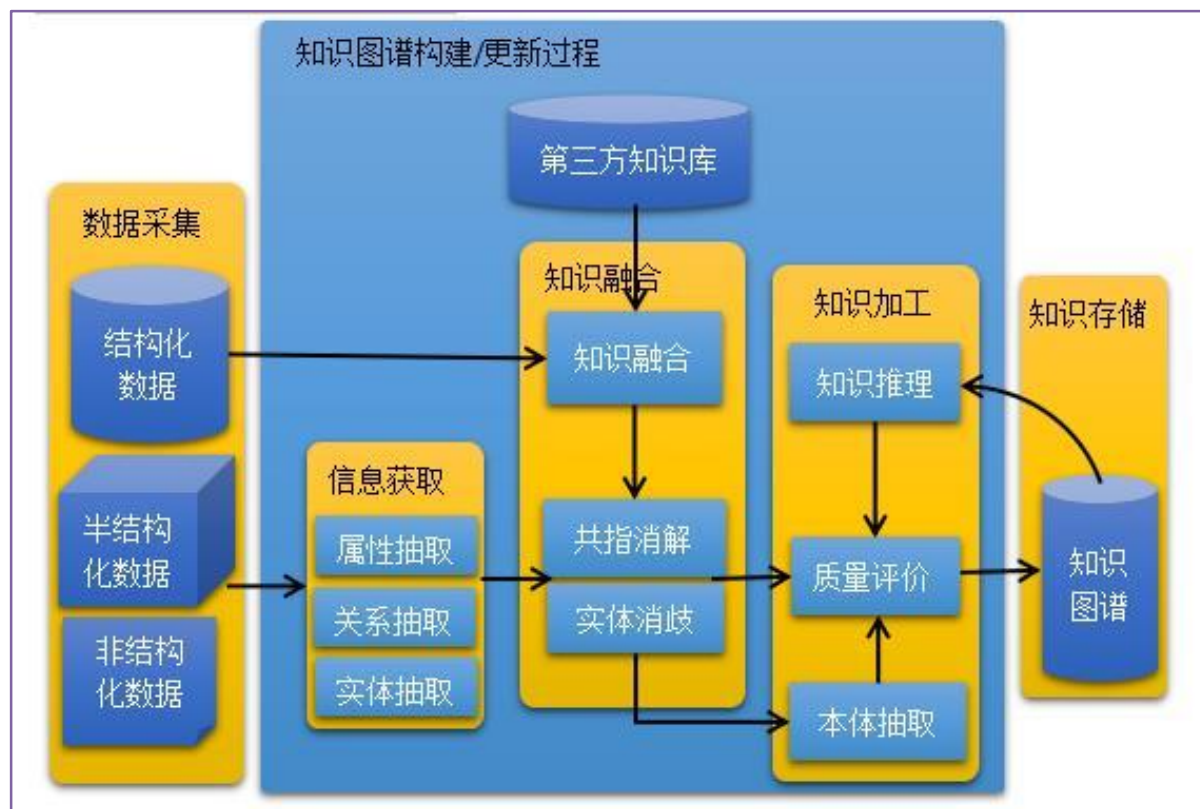
知识图谱

知识图谱基本组成



知识图谱

知识图谱总体架构图



知识图谱关键技术--知识抽取

■ 命名实体识别

命名实体识别 (Named Entity Recognition, NER) 是从一段自然语言文本中找出实体，并标注出其位置以及类型。

1. 传统的机器学习方法 (序列标注问题)

■ Hidden Markov Models (HMM)

未知参数：实体类型

优势：考虑了上下文信息

劣势：假设了可观测变量之间相互独立，限制观测变量是词语本身，限制了特征的选择，如词频、位置等

知识图谱关键技术--知识抽取

- Maximum Entropy Markov Models (MEMM)
生成模型变为判别模型，只计算给定可观测变量下隐藏变量的概率
优势：打破了HMM的观测独立假设
劣势：由于局部归一化，会造成标注偏差问题
- Conditional Random Fields (CRF)
将最大熵马尔科夫模型里面的条件概率转化为特征函数的形式,分解为两部分:转移特征和状态特征
优势：打破了观测独立假设，克服了标注偏差
劣势：训练速度较慢

2. 结合神经网络的典型方法

- IDCNN-CRF（迭代扩张卷积结合条件随机场）

Dilation Convolution

扩张卷积的好处是不做pooling损失信息的情况下，加大了感受野，让每个卷积输出都包含较大范围的信息。在自然语言处理中需要较长的sequence信息依赖的问题中，都能很好的应用。

Iterated Dilated Convolution

步骤：

知识图谱关键技术--知识抽取

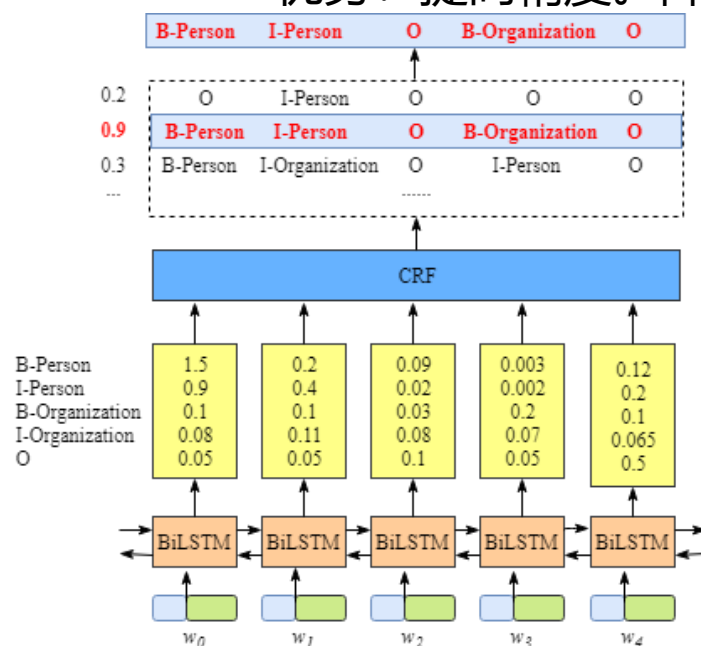
- ① 采用滑动窗口，窗口内做卷积，卷积层后接**扩张卷积**并继续进行L次迭代扩张卷积，这一过程称为“block”。
优势:保留更多信息。不增加网络层数，感受野以指数倍增加。
- ② 采用Multi-Scale Context Aggregation机制，将上一层Block输出作为下一层的输入。每一个Block具备整个输入的全局信息。
- ③ 改进的损失函数。每个Block后都可以直接接入CRF层并计算序列的标签概率。
将标签作为整体进行优化。
优势：提高精度。降低梯度弥散问题。

■ Bi-LSTM-CRF

步骤:

- ① 词向量表示。将句子中的词转化为词向量，作为网络的输入。
- ② 上下文表示。对词向量的序列进行BiLSTM，输出维度是tag siz。
- ③ 解码。使用CRF算法找到得分最高的标签序列和计算所有标签序列的概率分布。

CRF会考虑输出label之间的顺序性。



知识图谱关键技术--知识抽取

■ Bi-LSTM-CRF

优势：

- LSTM考虑的是输入序列X的上下文信息。CRF还可以考虑tag之间的依赖关系信息（对输出作一定的限制）。
- 命名实体识别准确率高、鲁棒性强。

知识图谱关键技术

评价指标

精确率 (Precision)

$$P = TP / (TP + FP)$$

■ 召回率 (Recall)

$$R = TP / (TP + FN)$$

■ F值

为了综合评价系统的性能,通常还计算召回率和准确率的加权几何平均值,即F指数,计算公式如下:

$$F = 2 / (1/P + 1/R)$$

表1 指标参数说明

	识别为正例	识别为负例
实际为正例	TP	FN
实际为负例	FP	TN

知识图谱关键技术--知识抽取

■ 实体关系抽取

关系抽取(Relation Extraction, RE)主要目的是从文本中识别实体并抽取实体之间的语义关系。

如：爱因斯坦创立了相对论

<爱因斯坦, 创立, 相对论>

一、基于远程监督的思想

Distant Supervision的提出主要基于以下假设：两个实体如果在知识库中存在某种关系，则包含该两个实体的非结构化句子均能表示出这种关系。

1. **Distant supervision** for relation extraction without labeled data [Stanford University, Mike Mintz etc, 2009]

提出远程监督的思想，减少人工依赖。

将纽约时报新闻文本与大规模知识图谱FreeBase进行实体对齐。

远程监督假设：一个同时包含两个实体的句子蕴含了该实体对在Freebase中的关系，并将该句子作为该实体对所对应关系的训练正例。

知识图谱关键技术--知识抽取

2. Modeling Relations and Their Mentions without Labeled Text [Conference Paper • September 2010, Sebastian Riedel etc]

在2的方法上基础上增强了远程监督的假设。

3. Multi-instance Multi-label Learning for Relation Extraction [Stanford University, Mihai Surdeanu etc, 2012]

引入多实例、多标签学习来缓解远程监督的噪音问题。

思路：

该方法的假设是，同一个包中，一个sentence只能表示 (E1, E2) 的一种关系，也就是只能给出一个label，但是不同的sentence可以表征 (E1, E2) 的不同关系，从而得到不同的label。多label标注的label值不是正或负，而是某一种关系。它为同时挖掘一个实体对的多种关系提供了可能的实现途径。

知识图谱关键技术--知识抽取

二、引入神经网络模型提取

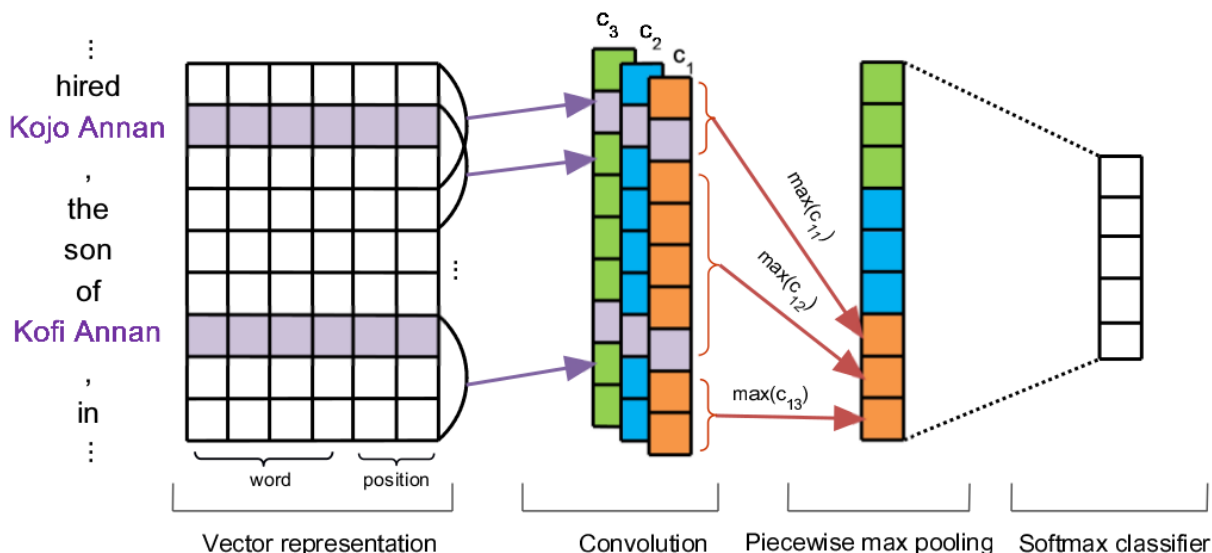
Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks [PCNN, Zeng et al. 2015 EMNLP]

主要思想:

主要解决Wrong Label问题以及传统统计模型特征抽取过程中出现的错误。

一是多实例学习。从instance level 到 bag level, 训练数据一共有T个bags, C种关系。

二是PCNNS。使用一个CNN + Piecewise Max Pooling 去抽取特征。



知识图谱关键技术--知识抽取

二、引入神经网络模型提取

特点：

- ① 文本向量化表示。使用了预训练的Word-Embedding，加入了Position Embedding。
- ② 卷积层。采用了常见的针对文本的卷积核设计，单向滑动。
- ③ 池化层。按照分段进行max pooling的，而PCNN的P就是这个代表，将句子按照两个entity进行分割，分割得到三段，这三段分别进行max pooling。
- ④ 分类器。使用了Softmax进行类别判断。

知识图谱关键技术--知识抽取

二、引入神经网络模型提取

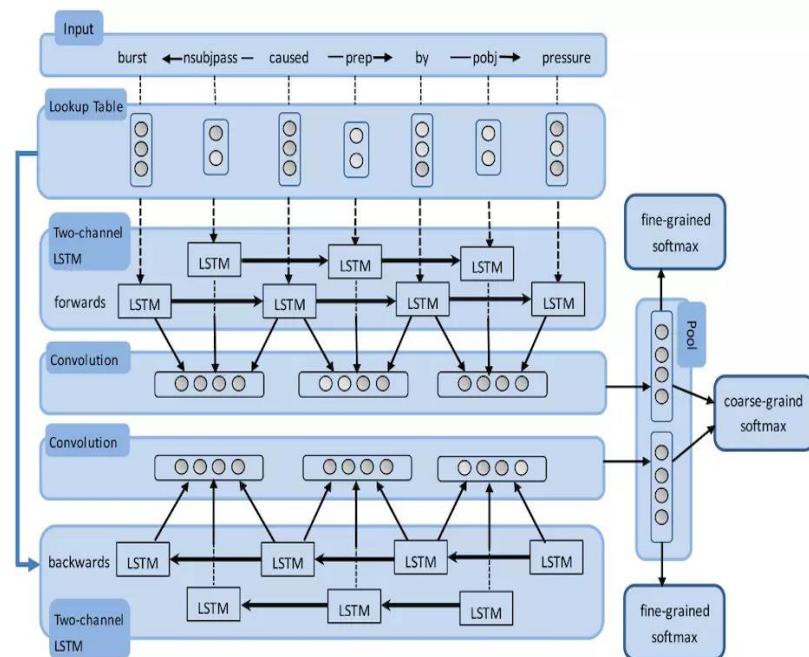
Bidirectional Recurrent Convolutional Neural Network for Relation Classification. [BRCNN, Rui Cai 2016 ACL]

主要思想：

双向递归卷积神经网络模型（BRCNN），一个基于最短依赖路径（SDP）的深度学习关系分类模型。BRCNN分为两个RCNN，一个前向（SDP为输入），一个后向（反向的SDP为输入）。

基本原理：

- ① 将SDP中words与words之间的依赖关系使用词嵌入表示，SDP中的words之间的依赖关系和words分开到两个独立channel的LSTM。
- ② Convolution层：相邻词对应的LSTM输出和它们的依赖关系的LSTM输出连结起来作为输入。
- ③ max pooling层。
- ④ Softmax分类。两个RCNN各使用一个分类器，合并使用一个分类器。



知识图谱关键技术--知识抽取

二、引入神经网络模型提取

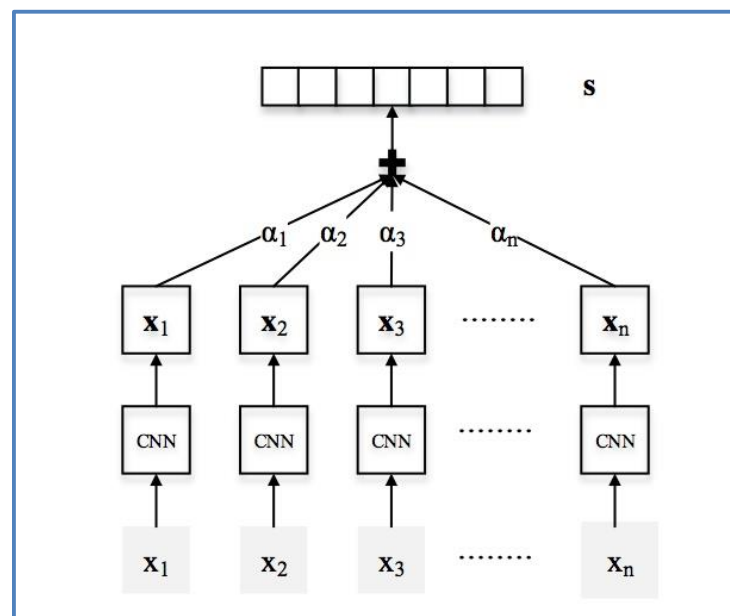
Neural Relation Extraction with Selective Attention over Instances [**CNN+ATT/PCNN+ATT**,清华 Lin et al. 2016]

主要思想:

- 在句子层面使用Attention机制来尽量减少错误的Label。
- 运用CNN将关系用Sentence Embedding的语义组合来表示, 以此充分利用训练知识库的信息。

步骤:

- ① 句子编码。word embedding 和 position embedding后依次通过CNN (或PCNN) 中的卷积、最大池化、非线性变换, 最后拼接所有的编码向量。
- ② Attention Model.对包含同一个实体对的所有句子的信息计算一个和向量, 得到该句子对表示关系的重要程度的指标。



知识图谱关键技术--知识抽取

二、引入神经网络模型提取

Deep **Residual Learning** for Weakly-Supervised Relation Extraction [**ResCNN-9**, Yi Yao Huang 台湾国立大学 EMNLP 2017]--可改进

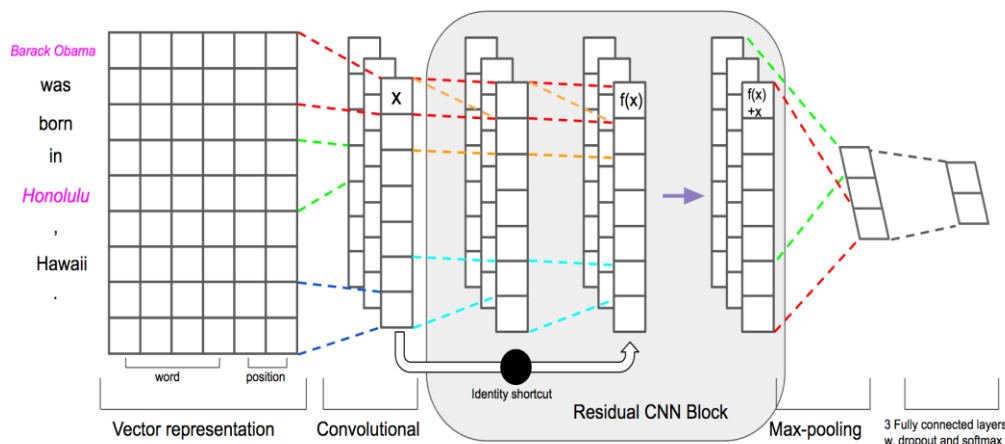
(弱监督关系抽取的深度残差学习方法)

主要思想:

探索更深的 CNN 模型对远程监督模型的影响, 本文设计了基于残差网络的深层 CNN 模型。实验结果表明, 较深的 CNN 模型比传统方法中只使用一层卷积的简单 CNN 模型具有较大的提升。

步骤:

- ① 向量化表示。表示词向量和位置向量表示
- ② 卷积层。为得到句子更高级的表示, 采用多个卷基层堆叠在一起。为了解决训练时梯度消失的问题, 在低层和高层的卷基层之间建立 Shortcut Connection。
- ③ 最大池化层。
- ④ 全连接层。



知识图谱关键技术--知识抽取

二、引入神经网络模型提取

Relation Classification via Multi-Level Attention CNNs. [Multi-Attention CNN (Wang et al. ACL, 2016)]

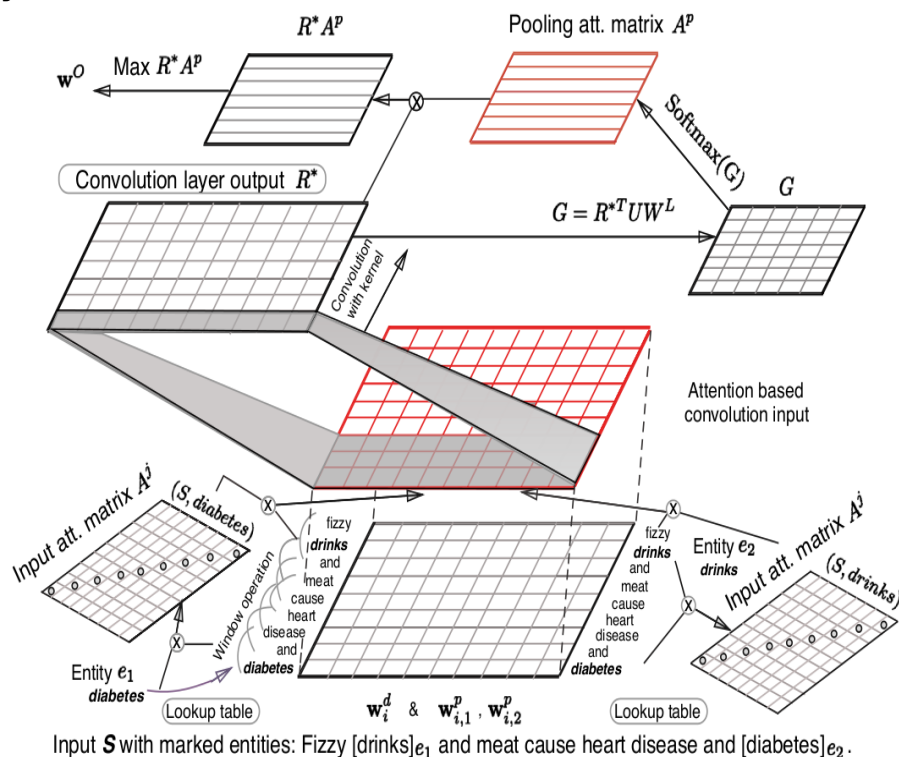
(基于多Attention机制CNN网络的实体关系抽取)

主要思想：

将Attention机制加入到神经网络中，对于反映实体关系更重要的词语给予更大的权重，辅以改进后的目标函数，从而提高关系提取的效果。

步骤：

- ① 词向量表示。每个词的表示由三部分组成，分别为：词的的两个相对位置表示与词嵌入表示。加入Attention机制，表示该词与给定实体间联系的强弱。
- ② 卷积层处理。
- ③ 基于注意力机制的池化处理。



知识图谱关键技术--知识抽取

二、引入神经网络模型提取

Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. [BLSTM + ATT, 中科大自动化所 Zhou ACL 2016]

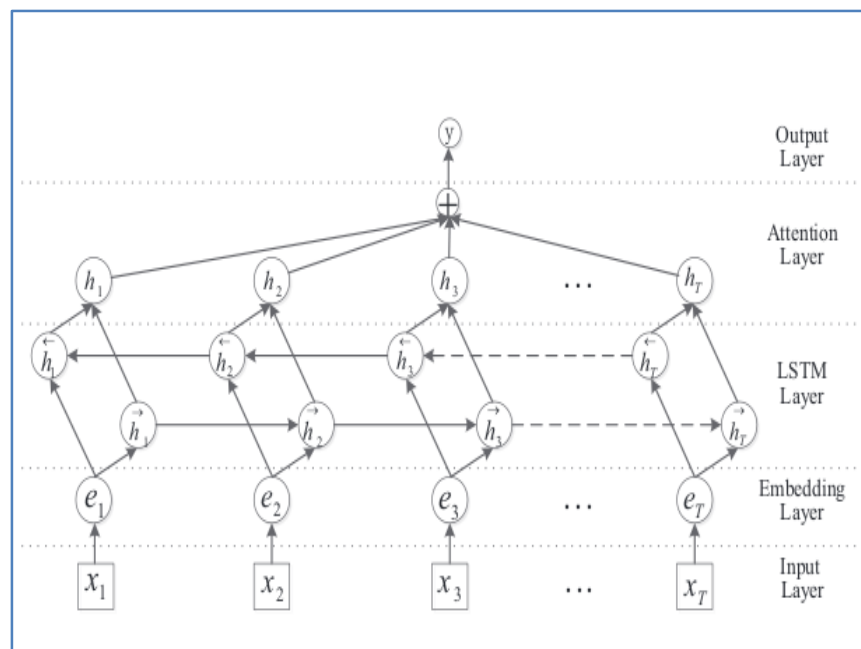
基于注意力的双向LSTM网络的关系抽取

主要思想：

运用双向lstm加attention自动提取包含整个句子语义信息关键词汇或者字符。

步骤：

- ① 词嵌入层。将每个词映射到低维空间。
- ② LSTM层。从词嵌入层中获取高级特征。
- ③ 注意力层。生成一个权重向量，通过与这个权重向量相乘。最终将词汇级特征合并为句子级特征。
- ④ 输出层。实现句子级特征向量的关系分类。



知识图谱关键技术--知识抽取

二、引入神经网络模型提取

一种实体和关系识别提取一次性解决的方法

CoType: Joint Extraction of Typed Entities and Relations with Knowledge Bases.[CoType,Ren, WWW,2017]

主要思想:

利用实体和关系的联合抽取来解决实体和关系提取的领域限制以及降低错误。主要提出以下技术:

- CoType框架基于Distant Supervision 和 Weakly Supervision对文本中的实体和关系联合抽取。
- 实体抽取的领域无关的文本分割算法。
- 联合嵌入的代价函数融合。

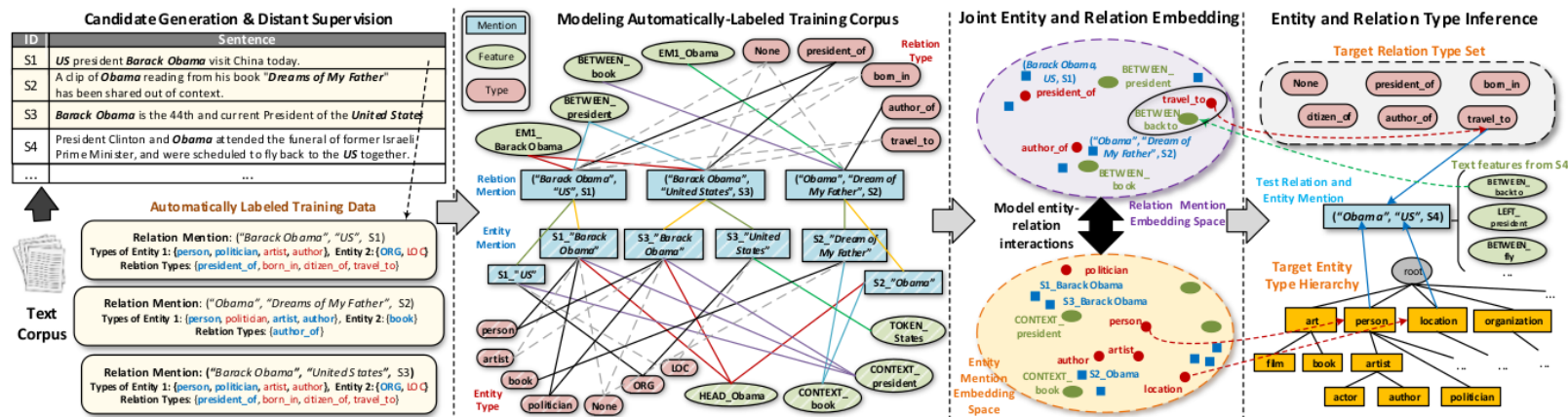


Figure 2: Framework Overview of CoType.

知识图谱关键技术--知识抽取

二、引入神经网络模型提取

CoType: Joint Extraction of Typed Entities and Relations with Knowledge Bases.[CoType,Ren, WWW,2017]

步骤:

- ① 输入。已标注词性的文本集 (D)，知识库，实体层级类型（为知识库实体层级类型子集），关系类型集合（为知识库的实体关系类型子集）。
- ② 生成。两个低维的向量空间（一个是实体的向量空间，一个关系向量空间），抽取的实体和关系。

框架分解:

- ① 候选集生成
 - 检测文本集中的实体选项 (Entity Mentions)，记为m
 - 实体选项和知识库中的实体做映射关联
 - 把实体选项在知识库中的映射关联实体的类型作为实体选项的类型,记为Y
 - 同一个句中的实体选项对，将它们在知识库中的映射关联的实体对之间的所有关系类型作为这两个实体选项对的候选关系，候选关系记为 z

知识图谱关键技术--知识抽取

二、引入神经网络模型提取

② 联合训练实体和关系向量空间

- 关系类型建模
- 实体类型建模
- 联合模型

② 实体类型和关系类型推理预测

通过第二部分训练得到的两个向量空间，向量表示第一步中unlinkable的部分，通过最近邻查找，预测unlinkable部分的实体和关系类型。

知识图谱关键技术--知识融合

1. 实体链接

实体链接（Entity linking, EL），目的是将文本中的实体映射到给定的知识库（KB），建立之间的关系，增强对实体的语义理解。

基本思想：

首先根据给定的实体指称项，从知识库中选出一组候选实体对象，然后通过相似度计算将指称项链接到正确的实体对象。

步骤：

1. 实体抽取
2. 候选集选取：在知识库中为该实体指称寻找一些可能成为该实体指称真正含义的实体。
3. 候选体实体排序：候选集确定后，并进行排序，达到实体消歧的目的。

技术点：实体消歧、共指消解

2. 知识合并

合并外部知识库，主要处理数据层和模式层的冲突

合并关系数据库，有RDB2RDF等方法

知识图谱关键技术--知识加工

- **本体构建**

本体编辑软件

自动化构建

实体并列关系相似度计算

中国科学院大学、清华大学、计算机

实体上下位关系抽取，生成本体

学校分为一类，计算机为一类

- **知识推理**

$A \rightarrow B$ $B \rightarrow C$

实体间的推理

属性值间的推理

本体的概念层次关系间的推理

主要方法：逻辑推理、基于图的推理、基于深度学习的推理

知识图谱关键技术--知识加工

质量评估

对知识的可信度进行量化，通过舍弃置信度较低的知识来保障知识库的质量。

知识更新

从逻辑上看，知识库的更新包括概念层的更新和数据层的更新。

更新方式：

全面更新

指以更新后的全部数据为输入，从零开始构建知识图谱。这种方法比较简单，但资源消耗大，而且需要耗费大量人力资源进行系统维护。

增量更新

以当前新增数据为输入，向现有知识图谱中添加新增知识。这种方式资源消耗小，但目前仍需要大量人工干预（定义规则等），因此实施起来十分困难。

知识图谱关键技术

知识存储

- 基于RDF存储：如Virtuoso

☐ include secondary database models

19 systems in ranking, March 2019

Rank			DBMS	Database Model	Score		
Mar 2019	Feb 2019	Mar 2018			Mar 2019	Feb 2019	Mar 2018
1.	1.	1.	MarkLogic	Multi-model	13.74	-1.25	+2.77
2.	2.	3.	Virtuoso	Multi-model	3.20	+0.26	+1.37
3.	3.	2.	Apache Jena - TDB	RDF	1.94	-0.24	-0.35
4.	4.	4.	Amazon Neptune	Multi-model	1.03	-0.03	+0.33
5.	6.	7.	GraphDB	Multi-model	0.93	+0.07	+0.48

- 基于图数据库存储：查询较高效，如neo4j

☐ include secondary database models

31 systems in ranking, March 2019

Rank			DBMS	Database Model	Score		
Mar 2019	Feb 2019	Mar 2018			Mar 2019	Feb 2019	Mar 2018
1.	1.	1.	Neo4j	Graph	48.58	+0.72	+7.68
2.	2.	2.	Microsoft Azure Cosmos DB	Multi-model	24.83	-0.03	+8.07
3.	3.	3.	OrientDB	Multi-model	6.13	+0.08	-0.33
4.	4.	4.	ArangoDB	Multi-model	4.26	-0.10	+0.14
5.	5.	5.	Virtuoso	Multi-model	3.20	+0.26	+1.37

知识图谱关键技术

知识图谱可视化常用工具

- ECharts

百度开源工具，API封装完善，简单好用，易上手，但不支持事件处理。

- Cytoscape.js

针对图形和网络，事件交互性的支持不错，同样易上手。

- D3.js

使用门槛较高，但支持事件处理器，D3的开销极小，支持大型数据集和交互动画的动态行为，支持图形丰富。

知识图谱关键技术总结

- 知识抽取
- 知识融合
- 知识加工
- 知识推理
- 知识存储
- 知识可视化

事理图谱

定义： 事理图谱（Event Evolutionary Graph, EEG）中的事件用抽象、泛化、语义完备的谓词短语来表示，其中含有事件触发词，以及其他必需的成分来保持该事件的语义完备性。

事理图谱与知识图谱的对比

	事理图谱	知识图谱
研究对象	事件 及其关系	实体 及其关系
组织形式	有向图	有向图
主要知识内容	事件间存在 顺承、因果、转移概率 信息	实体属性和关系、实体上下位信息等
知识特点	事件间演化规律的可能性度量	追求客观真实性

事理图谱

事理图谱的基本类型

事件	含义	形式化	事件应用	图谱场景	举例
因果事件	某一事件导致某一事件发生	A导致B	事件预警、分析	执果索因，执因索果	<地震，房屋倒塌>
条件事件	某事件条件下另一事件发生	如果A那么B	事件预警、分析	时机判定	<物美价廉，顾客多>
反转事件	某事件与另一事件形成对立	虽然A但是B	预测不测	反面教材	<起步晚,发展快>
顺承事件	某事件紧接着另一事件发生	A接着B	事件演化	未来事件推理	<去旅游,买火车票>

事理图谱

事理图谱的构建流程

语料收集---文本预处理---关系识别---关系节点抽取--事件表征

事理图谱

事理图谱构建的主要研究方向

- **事件抽取基本任务**

事件触发词检测 Event (trigger) detection

事件触发词分类 Event trigger typing （一般和detection一起做，归结为detection的一部分）

事件元素识别 Event Argument Identification

事件元素角色识别 Event Argument Role Identification

- **事件抽取的主要方法**

模式匹配方法

人工定义模板：词汇-语义模式、词汇-语法模式

机器学习方法

基于事件元素驱动、基于事件触发词驱动、基于事件实例驱动

混合事件抽取方法

模式匹配：事件识别+机器学习：事件分类

事理图谱

事件抽取的主要方法

Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks. [Y Chen, 2015, ACL] (DMCNN) ---动态多池化卷积神经网络的事件抽取

主要思想：

通过使用动态多池卷积神经网络自动学习特征来实现多类分类。

- 为了解决**词汇层**线索 (lexical-level clues) 问题，介绍了一种词编码 (word-representation) 方法。
- 为了解决**句子层**线索 (sentence-level clues) 问题，发明了一种动态多池化 CNN 方法。根据触发词以及参数词返回句子中每一部分的最大值。

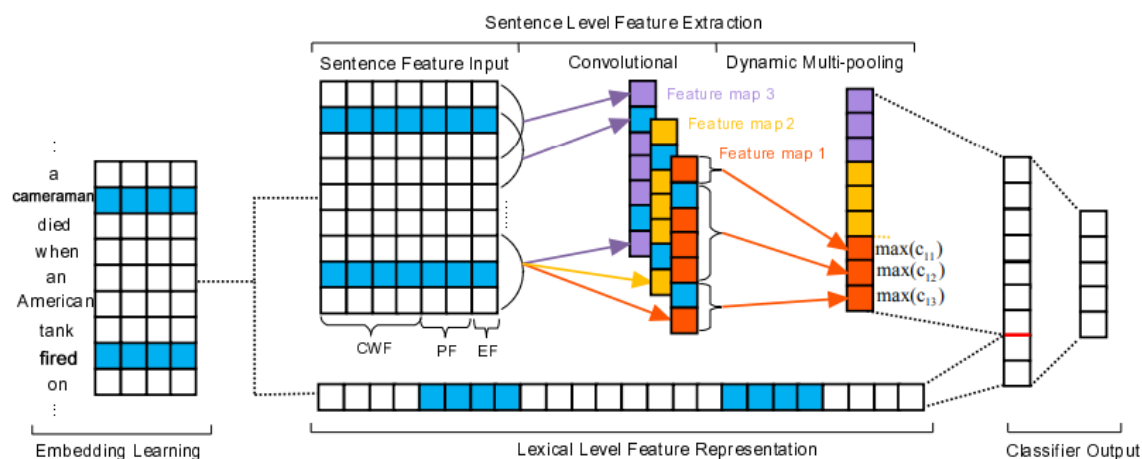
步骤：

- ① 词嵌入表示。
- ② 词汇级的特征表示。直接用词向量来发现词汇线索。
- ③ 句子级的特征抽取。提出DMCNN来学习句子中组成语义特征。
- ④ 论元分类输出。自动学习词汇和句子级特征全部串联成一个向量 $F=[L,P]$ ，为每一个论元候选角色计算置信度值。

事理图谱

事件抽取的主要方法

Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks. [Y Chen, 2015, ACL] (DMCNN) ---动态多池化卷积神经网络的事件抽取



事理图谱

事件抽取的主要方法

Exploiting Argument Information to Improve Event Detection via Supervised Attention Mechanisms. [Shulin Liu.etc, 2017, ACL, UCAS]

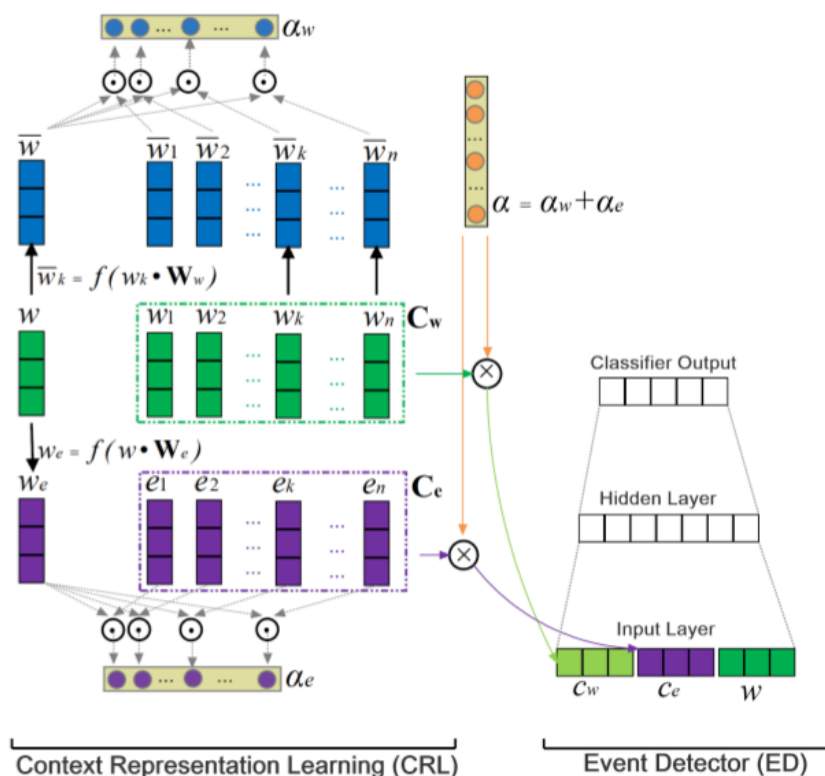
利用Argument信息以及Attention机制提升事件检测的效果

主要思想：

利用Argument信息对Trigger的识别和分类用。有监督的注意力机制。

步骤：

- ① Context Representation Learning (CRL)。在Trigger Candidate向量化表示的基础上，利用有监督的注意力机制进一步提取其中的有效信息
- ② Event Detector。
 - 输入层
 - 一个隐含层
 - 一个Softmax层



事理图谱

事件抽取的主要方法

Imitation Learning和GAN在事件抽取中的应用

Joint Entity and Event Extraction with Generative Adversarial Imitation Learning [T Zhang, H Ji, etc.2018] (基于生成对抗迁移学习的事件抽取)

(有启发性的一篇论文)

主要思想:

将 IE 任务中的两种 entity extraction 和 event extraction 结合起来, 提出一种端到端的基于模仿学习(imitation learning)的联合实体和事件抽取模型, 同时为抽取出的实体打上 argument role 的分类标签。并且在训练过程中使用了逆强化学习方法(inverse reinforcement learning)来提高模型的能力。



事理图谱

主要方法

使用模仿学习的思想用 Q-Learning 来做序列标记任务，找出哪些 word 是 entity 哪些是 trigger。之后用 policy gradient 来做事件对应 argument role 的判定，在训练过程中 extractor 的 reward 是由 GAN 来动态评估的。

RL Terms	Notations	EE Terms in supervised methods
Agent	A	Extractor
Expert	E	Ground truth from Human Annotators
Policy	π	Event model
Environment	e	Data set, corpus, documents or shared features
State	s	Features specified for a subtask
Action	a	Labels: event types, argument roles
Reward	r or R	Loss functions, which stimulate the models to update

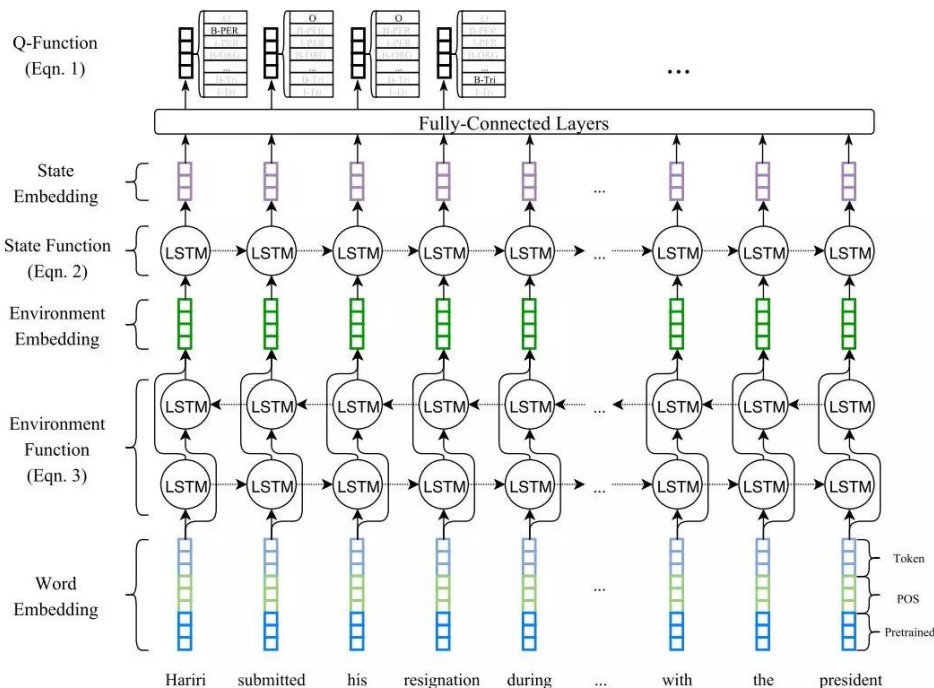
事理图谱

主要方法

IRL 的引入原因: 实际任务中设计 reward 很困难, 从认了专家提供的 ground-truth 去反推 reward 有利于该问题。于是, 对于给定状态空间 S , 动作空间 A , 并且给定决策轨迹数据集 π , 每个 π 由一个 s 和 a 的序列集合组成。

作者使用 Bi-LSTM 做 context embeddings, 然后输入到 forward LSTM 和全连接层去产生预测结果到 **Q-Table** 里面。

Q-table 主要获取所有分类的得分, extractor 就从 table 中获取 rank 得分最大的标记作为标记。这样在每一步都会计算一个 extractor 的标记和 ground-truth 的标记的差别, 以此来计算 reward。

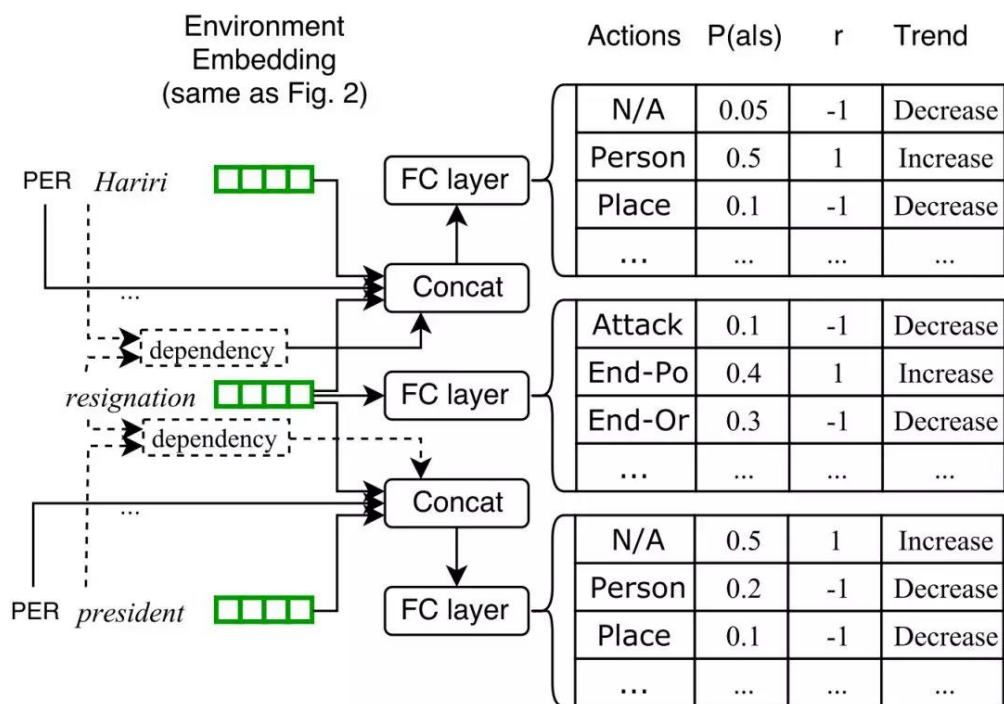


事理图谱

主要方法

Argument role:

考虑了 candidate argument 和 trigger 的上下文，得到 contextual embedding，同时也考虑了句子中各个实体的类型信息和 trigger 的类型信息。

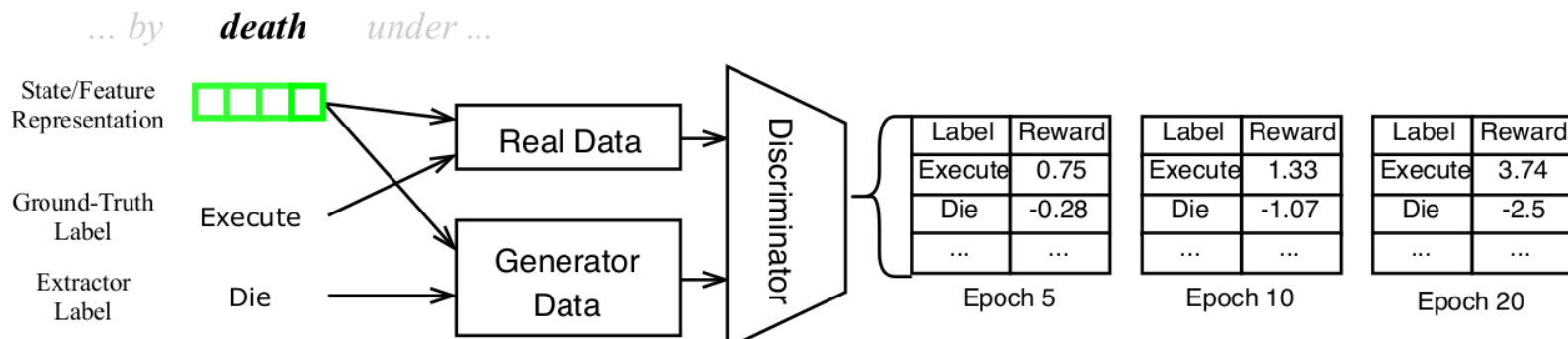


事理图谱

主要方法

GAN 调节Reward评分:

过不断对比 extractor 的标记结果和 ground truth 进行对抗。具体引入的原因是由于序列标注和角色标注都很复杂，所以reward 值应该是动态变化的。



- GAN 在训练过程中扩大了正确和错误标记的 rewards 值的差距，即对学习正确的 label 的例子将会给更高 reward 分，学习错误的label 的例子将会适当给更低的 reward 分，这样不断扩大学习中正负标记的 margins。
- 在训练的最后几个 epoch，就会获取一个比较好的动态的 reward 值，这个 reward 值就可以反过来导向 extractor 去进行正确标记。

事理图谱

• 事件因果关系抽取

事件因果关系识别是一种推断事件之间是否有因果关系的任务，主要是给定一对event pairs，根据文本中的信息及一些挖掘的因果特征来确定这一对事件是否存在着因果关系。

Improving Event Causality Recognition with Multiple Background Knowledge Sources Using Multi-Column Convolutional Neural Networks [MCNN. AAAI 17]

“smoke cigarettes” → “die of lung cancer”

主要思想：

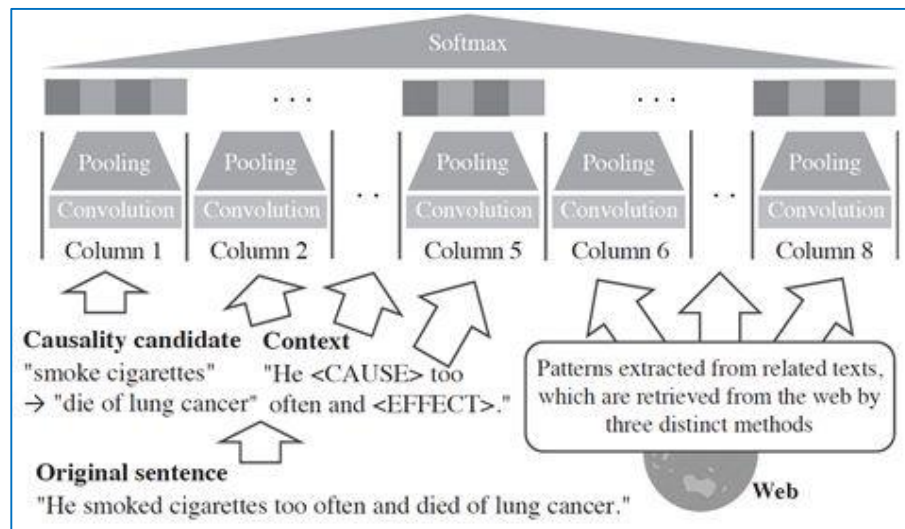
用 MCNN 来从海量语料中寻找一个事件对的因果关系paths，从而判断事件对之间是否有因果关系。**输入**是一对因果关系候选实体对，未知是否有因果关系。**输出**就是是否有因果关系。

事理图谱

• 事件因果关系抽取

步骤：

- ① 输入。MCNN 设置为8列，其中第一列放两个 candidate，第二列到第五列都放包含这两个事件 candidate 的句子的上下文。其余三列放从相关文本中的 patterns。MCNN 每一列都使用完整的 CNN 结构，只是 filter 的大小可以根据每个部分的不同内容来确定。
- ② 特征提取。MCNN 每一列都使用完整的 CNN 结构，只是 filter 的大小可以根据每个部分的不同内容来确定。
- ③ 分类输出。softmax分类器，是否具有因果关系。



• 事件抽取的评价指标

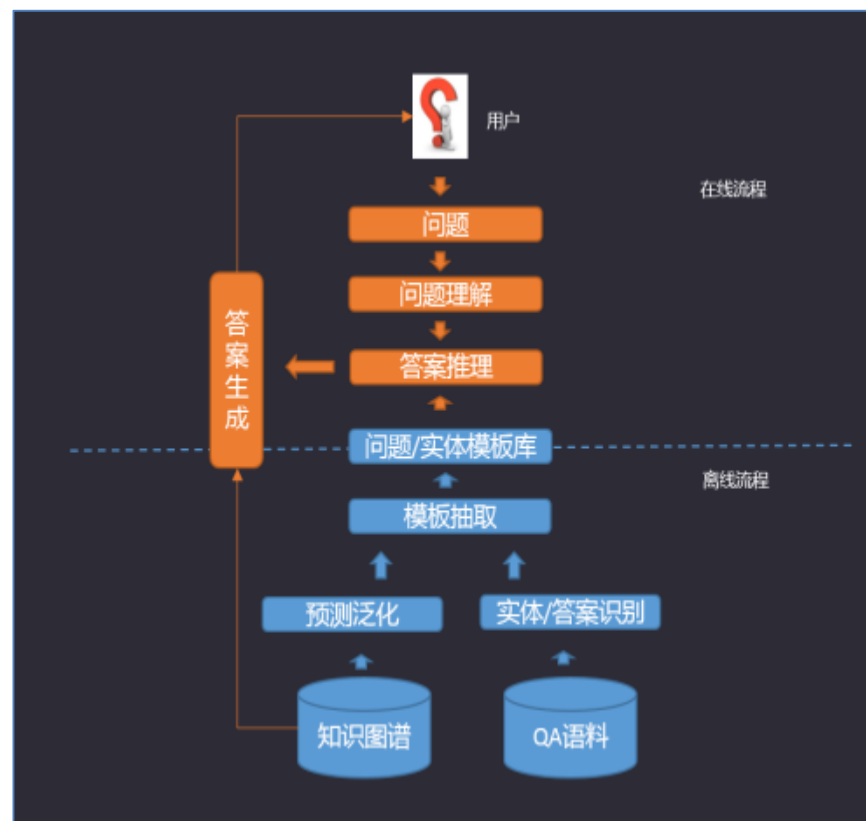
基于召回率 (记为 R) 准确率 (记为 P) 的微平均 (记为 F) 值法

基于丢失率 (记为 L) 误报率 (记为 M) 的错误识别代价 (记为 C) 法

基于知识图谱的问答系统

主要处理过程：

- ▶ 1、**自然语言语义理解**：主要是把用户说的话转成机器能理解执行的指令或查询，即做到“听得懂”。
- ▶ 2、**对话管理**：维护对话状态和目标，决定系统应该怎么说、怎么问下一句话，也就是生成一个应答的意图。
- ▶ 3、**自然语言生成**：就是根据系统应答的意图，用自然语言把这个应答意图表达出来。



基于知识图谱的问答系统

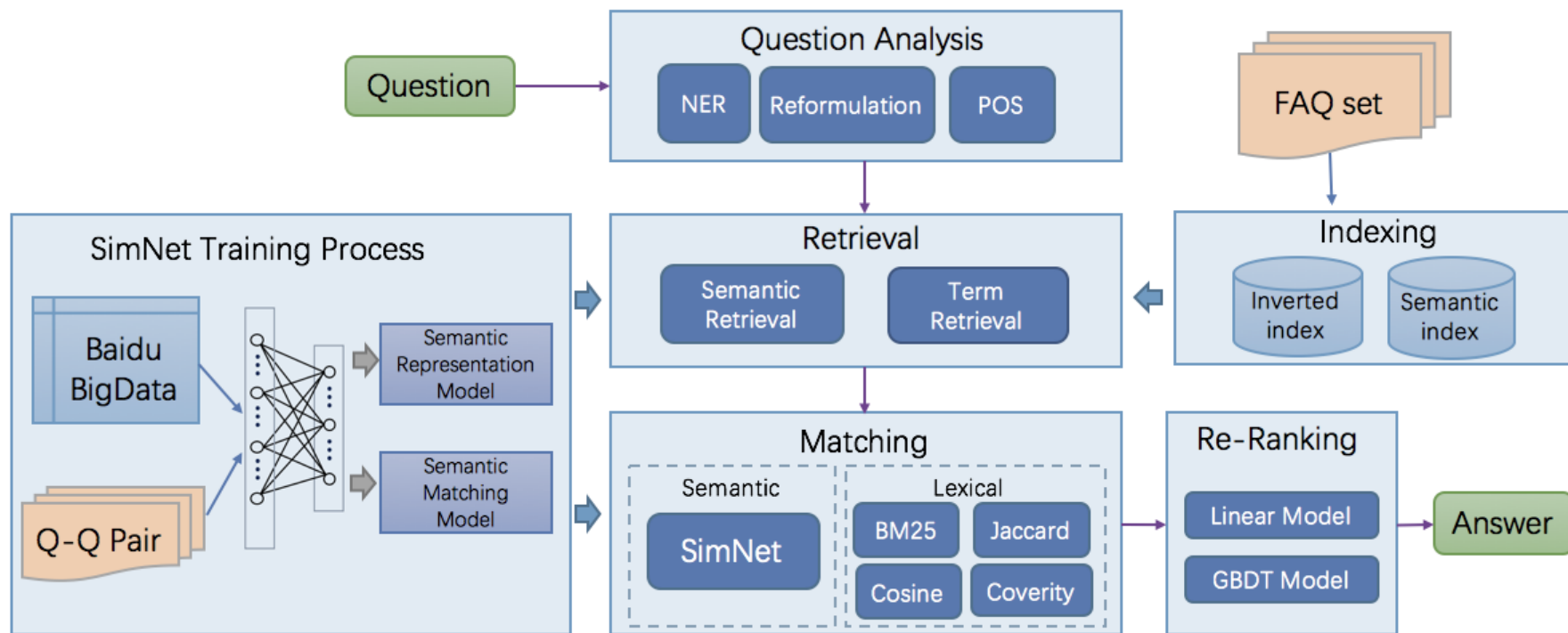


目的：解决事实类问题和简单的推理
基于KG的优势：

- ▶ 数据关联度-语义理解智能化程度高，语义理解程度是问答系统的核心指标。
- ▶ 数据精度-回答准确率高 知识图谱的知识来自专业人士标注，或者专业数据库的格式化抓取，这保证了数据的高准确率。

基于知识图谱的问答系统

问答系统总体架构：



源于百度相关技术框架

基于知识图谱的问答系统

基于KG的优势

数据结构化-检索效率高，知识图谱的结构化组织形式，为计算机的快速知识检索提供了格式支持。如SPARQL等

基于KG的QA系统的核心问题

- ① 如何理解问题语义，并用计算机可以接受的形式进行表示(问题的理解和表示)
以分类模型和序列标注为核心的语义理解模型能够精准的将自然语言转化为框架语义表示的结构化信息。
- ② 以及如何将该问题表示关联到知识图谱的结构化查询中(语义关联)

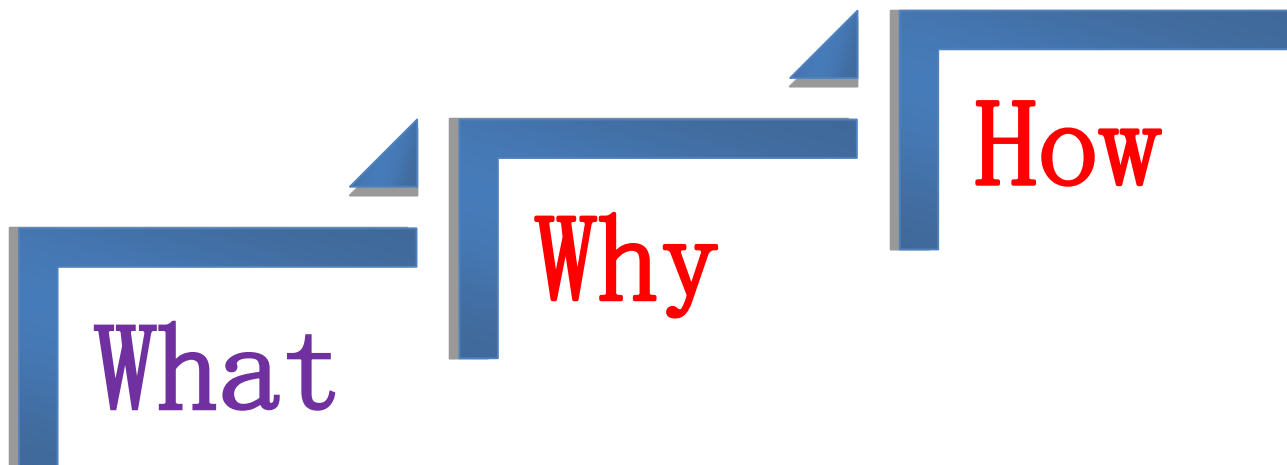
应用领域

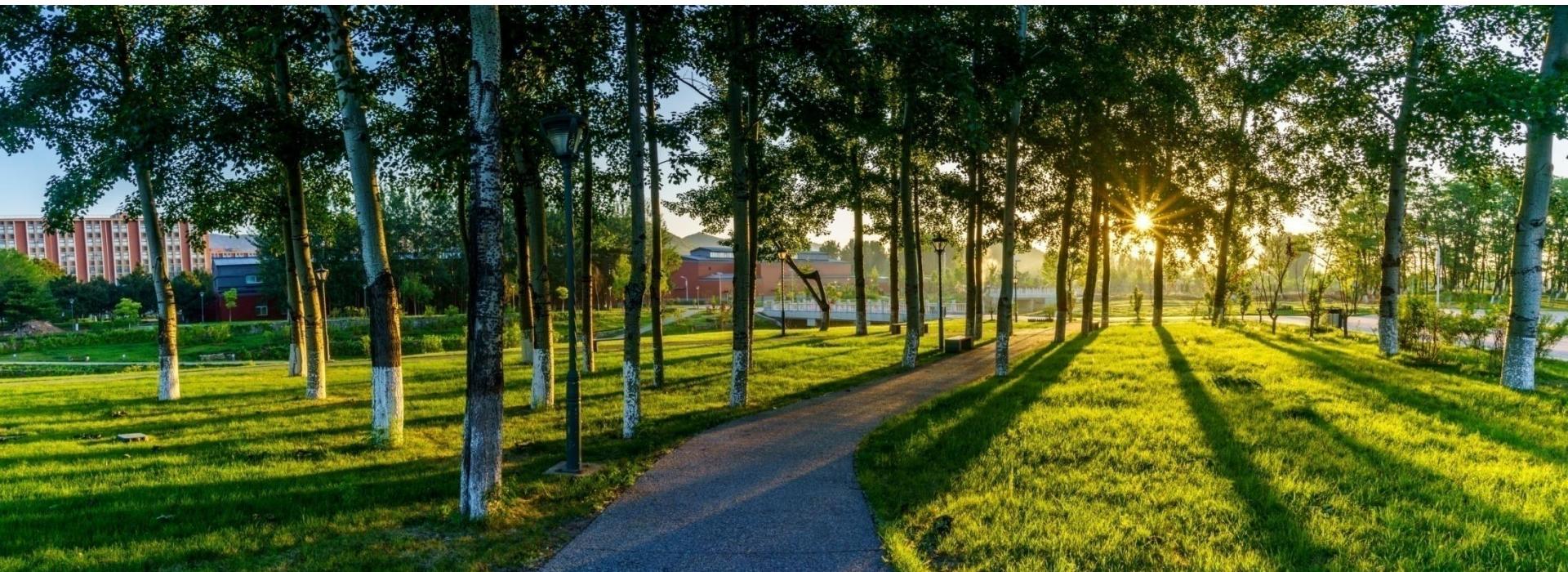
金融领域

医疗领域

.....

认知智能





THANKS



中国科学院大学
University of Chinese Academy of Sciences

