第八届全国社会媒体处理大会

事理图谱的构建及应用

丁效

哈尔滨工业大学

社会计算与信息检索研究中心

2019.8.18 深圳

大数据驱动行业的发展变革

- **人工处理->智能计算分析**
- **海量的金融信息，超过了人类处理的极限**

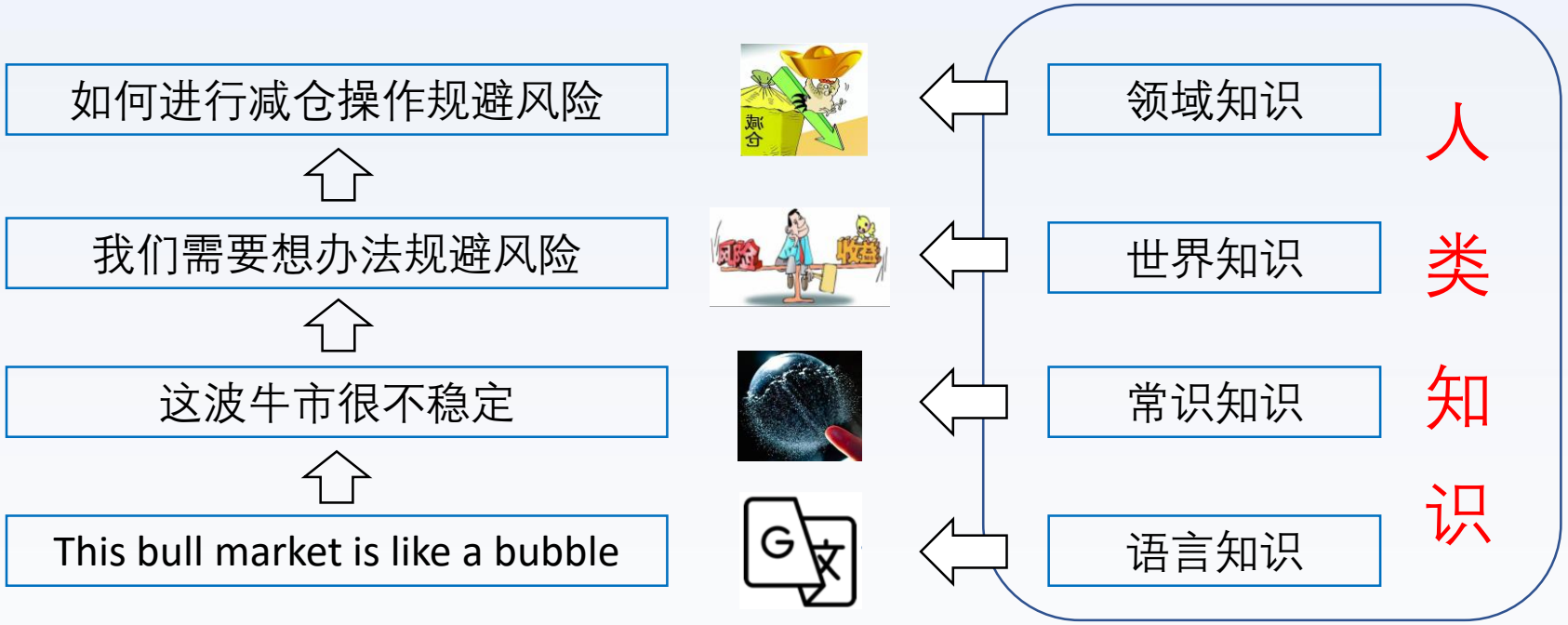
“20世纪90年代，一个基金经理把市场当天产生的研报、舆情、新闻、交易数据看完，大概需要10个小时；2010年，大概需要10个月的时间；2016年，把当天市场信息看完，需要20年的时间，相当于整个职业生涯”

——《智能革命》

本页ppt借鉴于刘知远老师

面临挑战

- 对自然语言的深度理解需要复杂知识的支持



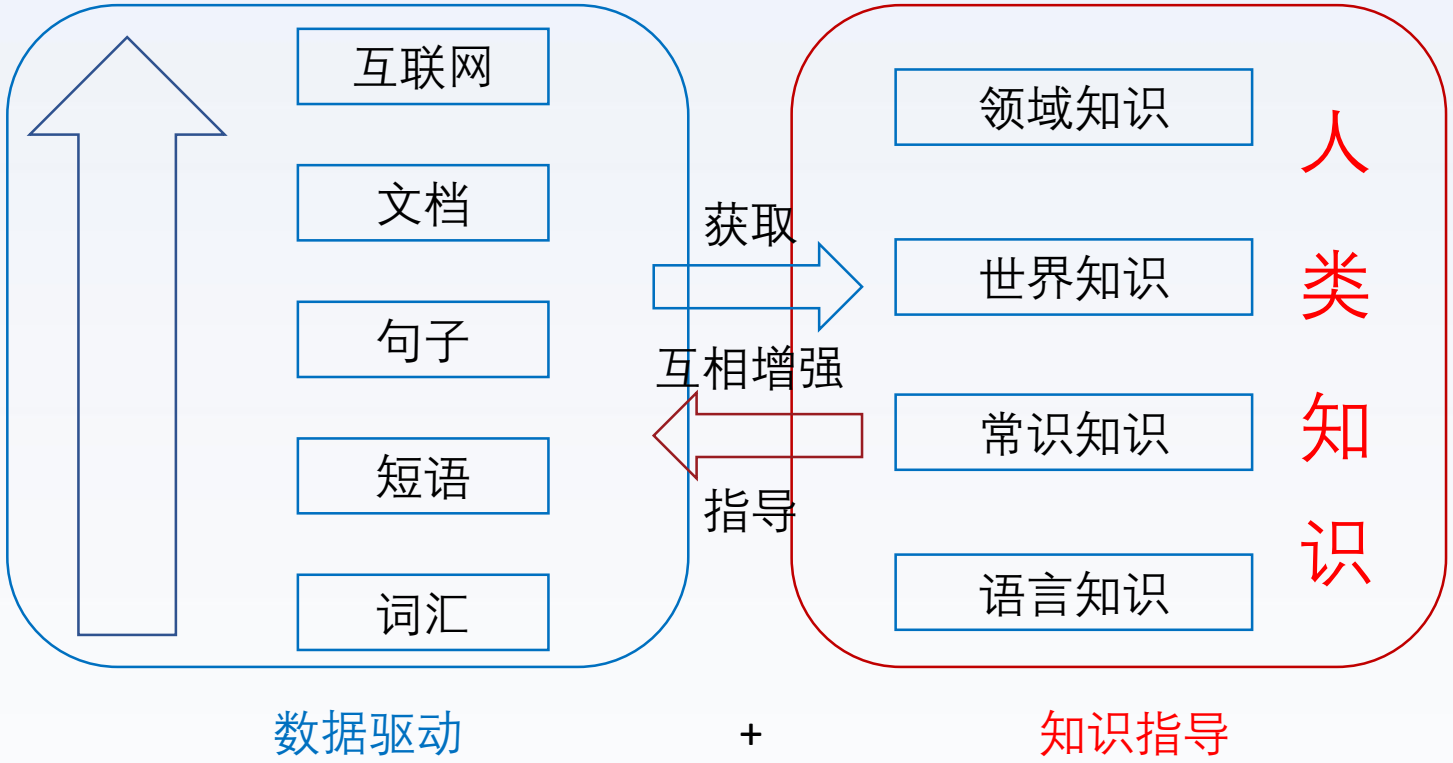
这波牛市就像泡沫

亟需知识支持实现NLP从字面意思到言外之意的跃迁

本页ppt借鉴于刘知远老师

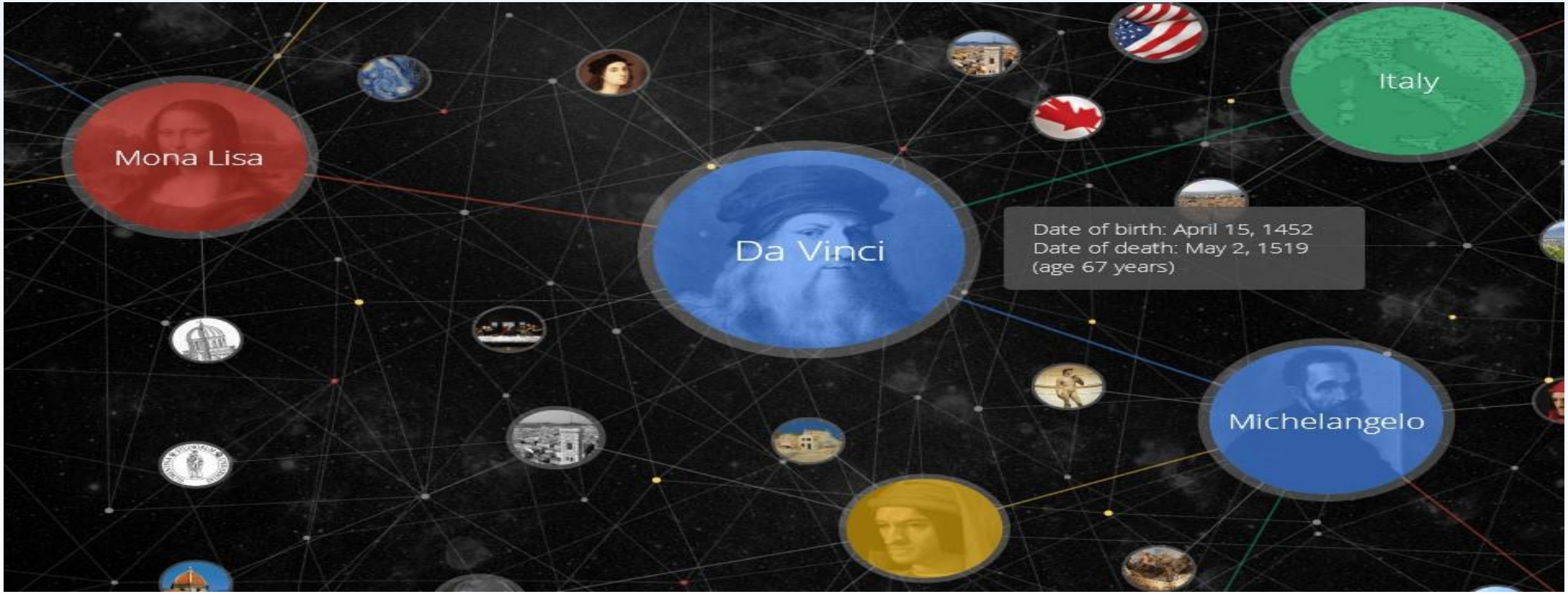
自然语言的特点

- 自然语言文本蕴含丰富的语言知识和世界知识



技术挑战

- 知识库的构建、表示及应用



知识图谱的构建，从人工到自动

人工搜集数据
人工标注数据

CYC

Cycorp

WordNet

WordNet
A lexical database for English

Wikipedia

500万条术语
700万条逻辑

117659同义词集
206941注释

500多万文档（英文）
2000万贡献者

文本/规则存储数据（知识库）

机器爬取数据
机器自动构建

yago select knowledge

DBpedia

freebase

WIKIDATA

Google Knowledge Graph

三元组存储数据（机器推理）

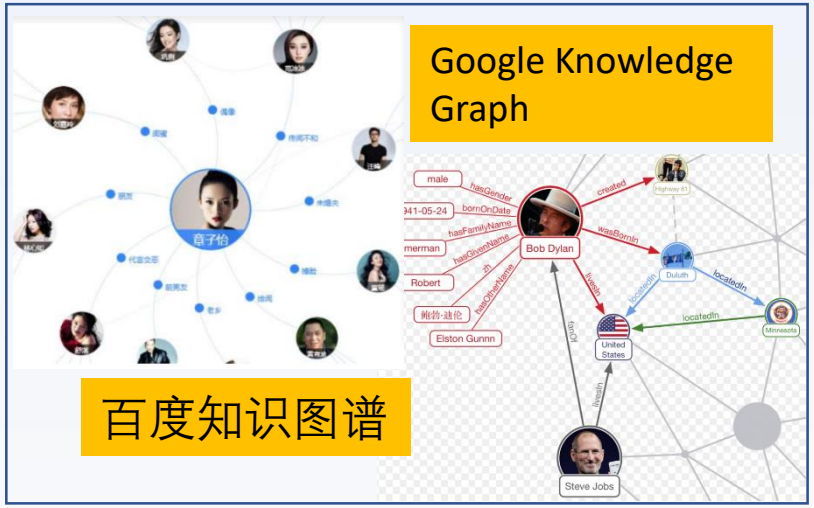
封闭域知识图谱 vs 开放域知识图谱

封闭域

DBpedia：
4 类基本关系
Part-of/kind-of/
instance-of/
attribute-of

关系类型人为设定
- 优点：易于扩充、无冗余
- 缺点：灵活性差、不便应用

开放域



关系类型自动挖掘
- 优点：关系灵活、覆盖全面
- 缺点：冗余不一致、难以支持逻辑推理

哈工大研制的音乐领域知识图谱：“八维音乐资讯”（2010年）

八维音乐 周杰伦 八维搜索

音乐首页 | 音乐新闻 | 音乐人物 | 歌曲专辑 | 挑歌 | 预告

最新动态

举行演唱会 更多>>

2012年 中国风巡演

（地点：）

周杰伦

性别：男

国家/地区：港台

生日：1979年1月18日

中国风巡演		2012年	查看来源
中国风亚洲巡演		2011年11月1日	查看来源
唱响今世缘，周杰伦群星南京大型演唱会	南京五台山体育场	2011年10月29日	查看来源
周杰伦中国风演唱会		2011年10月13日	查看来源
动感地带之夜周杰伦2011世界巡回演唱会	沈阳奥体中心	2011年8月27日	查看来源

发行专辑 更多>>

《跳痛》	2010年6月5日	查看来源
《划时代》	2010年5月14日	查看来源
《超人不会飞》	2010年4月26日	查看来源

有关周杰伦 (艺术家) 的关系 更多>>

专辑

魔杰座	查看来源
跨时代	查看来源
我很忙	查看来源
惊叹号	查看来源
范特西	查看来源

歌曲

青花瓷	查看来源
稻香	查看来源
菊花台	查看来源
千里之外	查看来源
双截棍	查看来源
本草纲目	查看来源
牛仔很忙	查看来源
千山万水	查看来源
不能说的秘密	查看来源
说好的幸福呢	查看来源

唱片公司

杰威尔音乐有限公司	查看来源

有关周杰伦 (艺术家)的印象 更多>>

总体印象： 中国风 天王 r&b 不好 简单

音乐： 独立 简单 奇幻 好听 激烈

歌： 好听 很好听 不错 再正常 比较好

2014年末《大词林》正式发布

- 2014年11月27日，哈工大正式发布《大词林》
- 《大词林》是自动构建的基于上下位关系的大规模开放域中文知识图谱

2014-11-27 08:56 被转发过

刘挺

10年前，我实验室（ 网页链接 ）研制了《同义词词林扩展版》。今天我们推出了《大词林（BigCilin）》测试版（ 网页链接 ），《大词林》借鉴《词林》的顶层结构，能从网上自动挖掘命名实体及实体间的上下位关系，是时刻都在自我更新的活的中文词汇语义知识库。欢迎大家试用指正！

2014-11-27 06:55 来自 微博 weibo.com

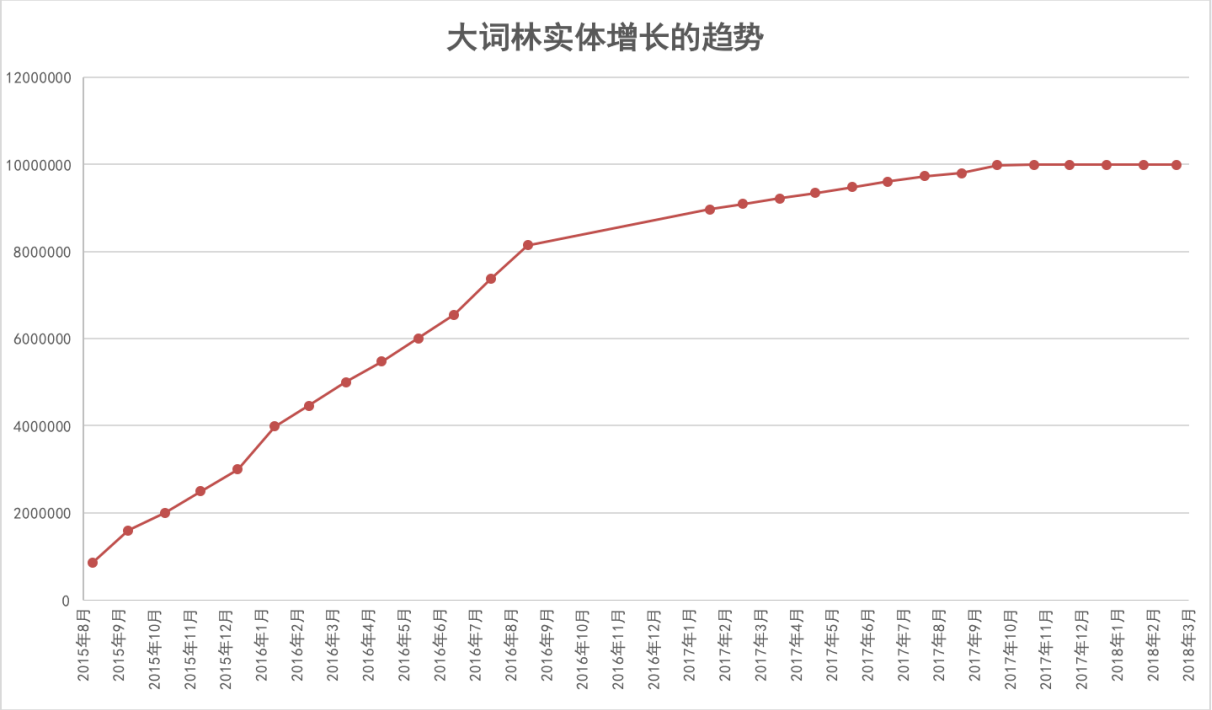
收藏 | 转发249 | 评论20 | 28

大词林(BigCilin, http://www.bigcilin.com)

总数据量 **10,223,423（一千万）**

实体词数量 **10,041,344（一千万）**

概念词数量 **182,079 (十八万)**



实体间关系数据：目前《大词林》拥有的实体间关系对数量为**12,258,278（一千两百万对）**。

实体属性-值对数据：目前《大词林》拥有的属性数量（实体属性对数量）为**45,854,287（四千五百万对）**。关系（属性）类型**161,983（十六万）**。

上位词别名（同义词）数据：**129,424（十二万对）**。

实体词别名（同义词）数据：**15,988（一万五条记录）**

事理图谱的研究意义

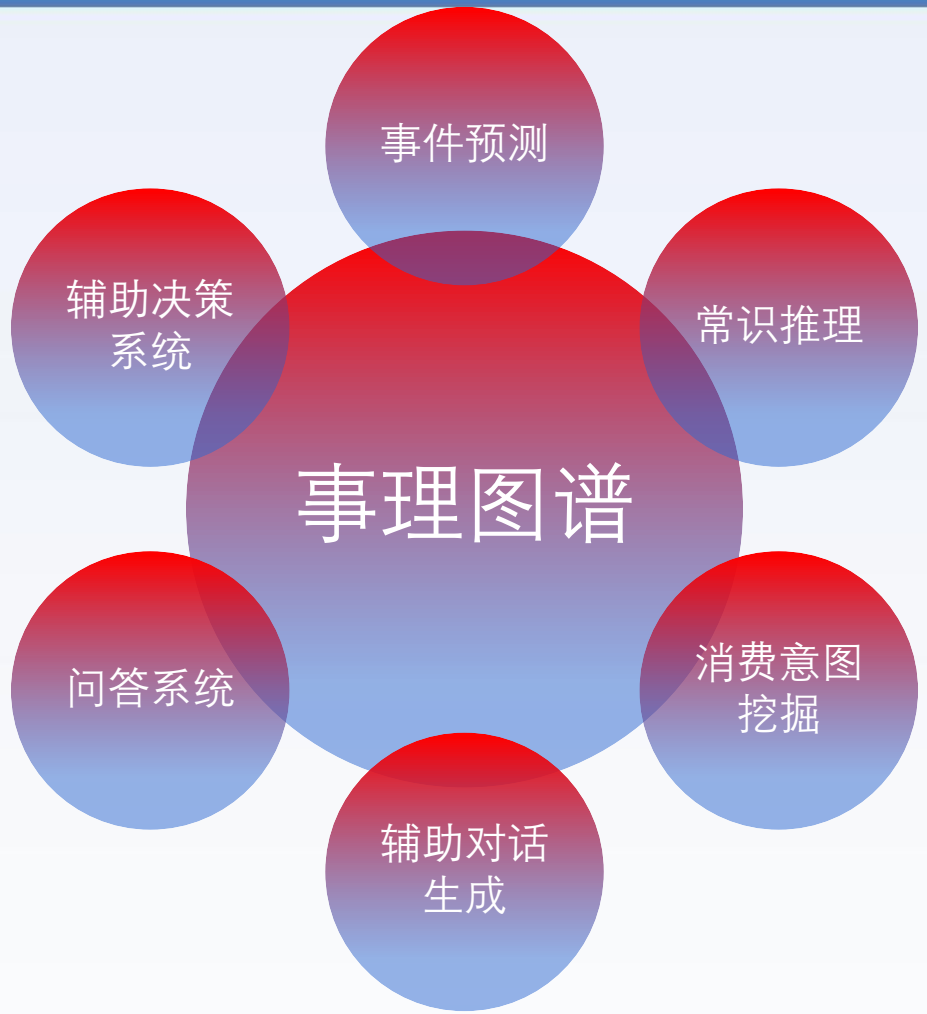
- 现有的知识库普遍是以“概念及概念间的关系”为核心的，缺乏对“事理逻辑”知识的挖掘
- 事理逻辑（事件之间的演化规律与模式）是一种非常有价值的人类知识，挖掘这种知识对我们认识人类行为和社会发展变化规律非常有意义

什么是事理图谱？

- 事理图谱：Event Logic Graph（ELG）
- 定义：事理图谱是一个事理逻辑知识库，描述了事件之间的演化规律和模式。结构上事理图谱是一个有向有环图，节点代表事件，有向边代表事件之间的顺承、因果、条件和上下位等逻辑关系。

事理图谱的应用

- 事理图谱可应用于事件预测、常识推理、消费意图挖掘、对话生成、问答系统、辅助决策等任务中
- 大规模事理图谱将和传统知识图谱一样，具有非常巨大的应用价值。



事理图谱与知识图谱的区别与联系

	事理图谱	知识图谱
研究对象	谓词性事件及其关系	体词性实体及其关系
组织形式	有向图有环图	有向图有环图
主要知识形式	事理逻辑关系，以及概率转移信息	实体属性和关系
知识的确定性	事件间的演化关系多数是不确定的	多数实体关系是确定性的

事理图谱中的事件定义

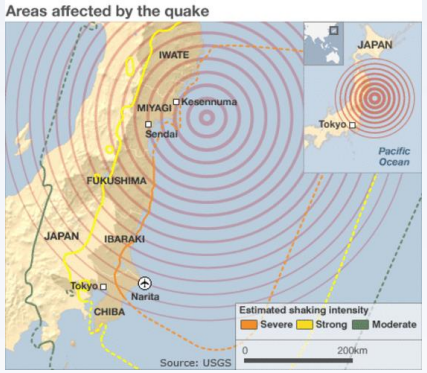
- 前人工作: 事件是特定时间、地点下的一个状态变化
  - 经典的事件抽取和分类任务, ACE 2005
  - 话题检测与跟踪

- 事理图谱中的事件:
  - 理论上，事理图谱中的事件是具有一定抽象程度的泛化事件。
  - 表示为抽象、语义完备的谓词性词或词组，也可以表示为可变长度的、结构化的(主体，事件词，客体)多元组
  - 其中必然包含一个事件词，标志事件的发生，例如：“跑步”；而主体和客体都可以在不同的应用场景下被省略，例如：“（元首，出访）”可以省略事件的客体；“（购买，机票）”可以省略事件的主体

事件关系可归纳：从事实到认知

- 例子：*A massive 8.9-magnitude earthquake hit northeast Japan on Friday, which cause a large amount of houses collapsed.*

（星期五，日本东北部发生8.9级大地震，造成大量房屋倒塌。）



常识性知识：地震 → 房屋倒塌

事理图谱的研究动机

- 从具体的、细节化的事件因果（事实）中归纳总结出事件因果律（认知）
- 利用事件因果律（认知）指导实践，例如事件预测和股票预测（事实）

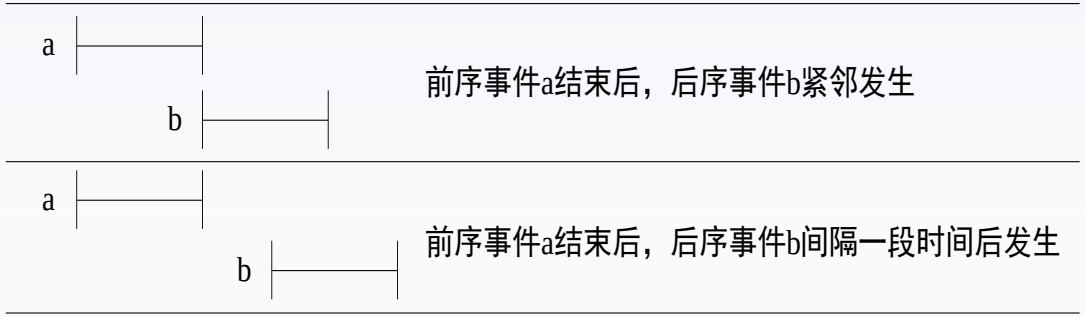
事件间顺承关系

- 顺承关系(Sequential)是指两个事件在时间上相继发生的偏序关系
- 两个前后顺承的事件之间存在一个介于0到1之间的转移概率，表示从一个事件按时序顺承关系演化到下一事件的置信度

吃过午饭后，小明到前台买单，然后离开了餐馆。

事理图谱中的顺承关系

- 借鉴TimeML时序关系类别中的before和after偏序关系，在事理图谱中的顺承关系包括两种情况：一种情况是顺承的前序事件a结束后，后序事件b紧接着发生；另一种情况是前序事件a结束后，隔一段时间后序事件b才会发生
- 事理图谱重点关注以下两种种顺承关系



事件间因果关系

- 因果关系(Causal)是指两个事件之间，前一事件（原因）导致后一事件（结果）的发生
  - 满足发生时间上的偏序约束
- 因果关系是顺承关系的子集；因果事件对之间存在一个介于0到1之间的因果强度值，表示该因果关系的置信度

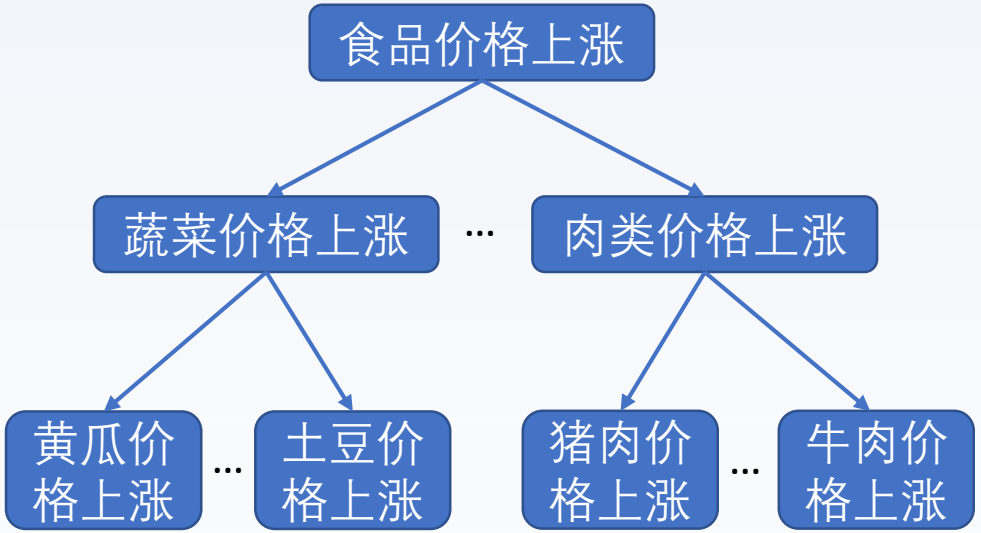
核泄漏引起了严重的海洋污染。

核泄漏 → 因果 → 海洋污染

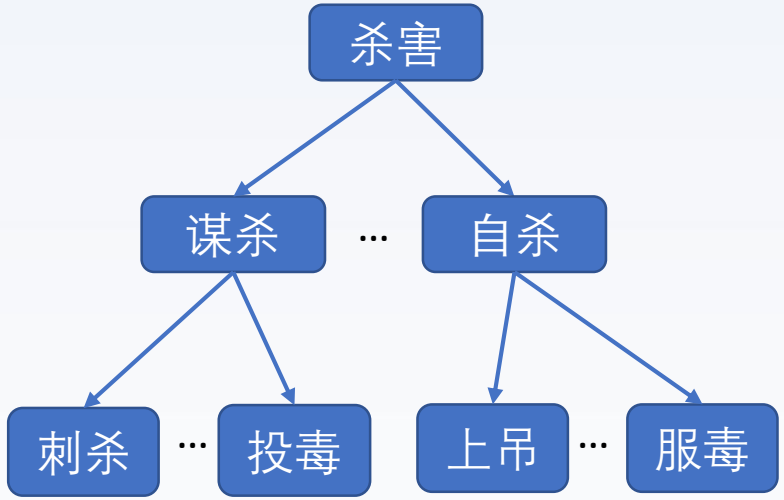
事件间上下位关系

- 事件之间的上下位关系有两种：名词性上下位关系和动词性上下位关系

名词性上下位关系



动词性上下位关系



事件间条件关系

- 条件关系是指前一个事件的发生是后一个事件发生的条件。

只有刻苦学习，才能取得好成绩。

刻苦学习 → 条件 → 取得好成绩

条件关系vs因果关系

- 二者在范围上并不相等，条件≠因果

如果天下雨，那么地上湿 = 因为天下雨，所以地上湿

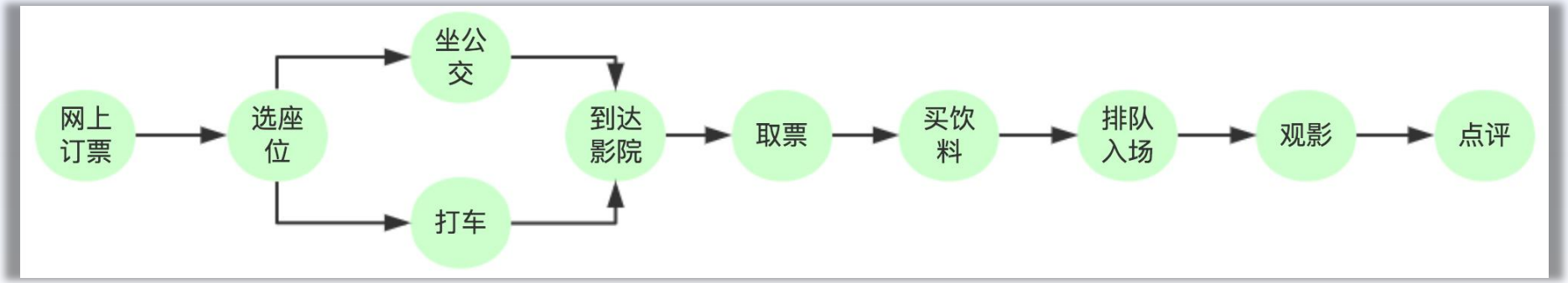
如果地上不湿，那么天没下雨 ≠ 因为地上不湿，所以天没下雨

- 二者论述对象不同
  - 条件关系属于思想中命题的某种逻辑关系，因果关系属于对客观事实的某种认识，原因≠理由
  - “原因”指的是事件之间的因果关系，是关于事实的，“理由”是前提与结论或论据与论点的内在联系，是关于逻辑的

如果买票的人多，那么电影好看 √

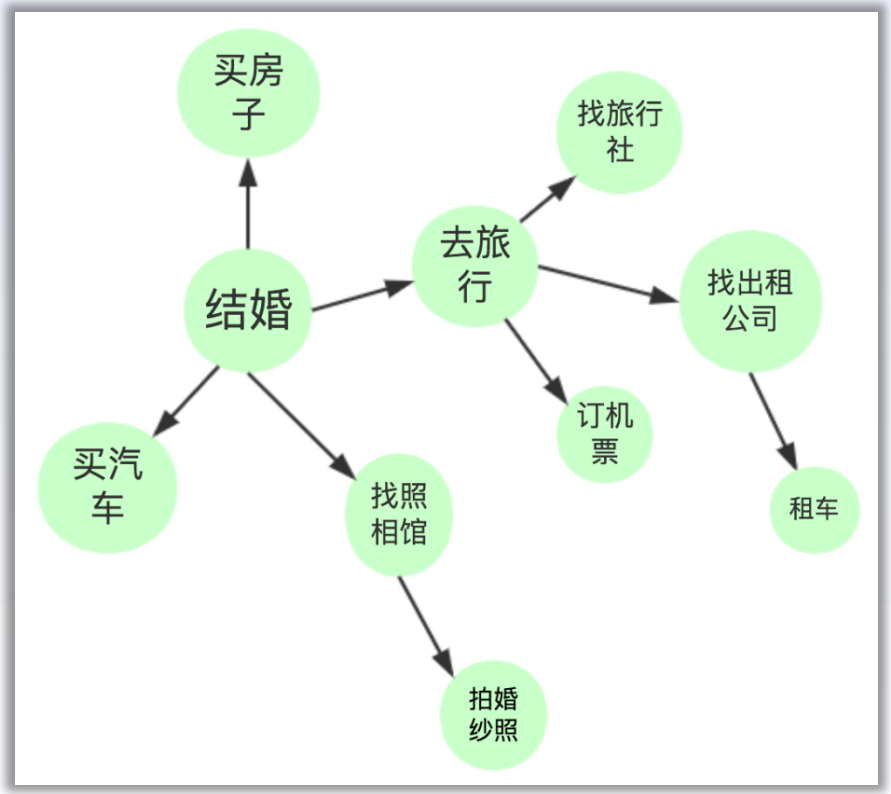
因为买票的人多，所以电影好看 ×

事理图谱的三种典型子图结构



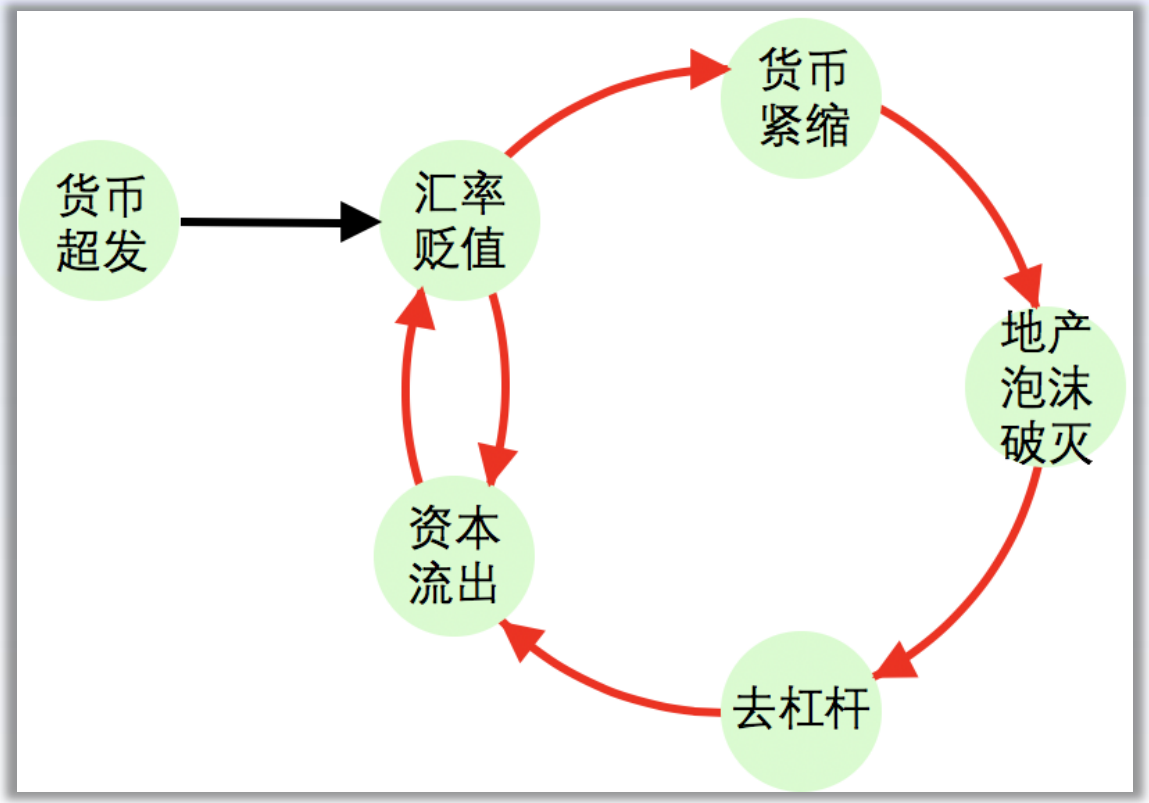
‘看电影’场景下的**链状**事理图谱

事理图谱的三种典型子图结构



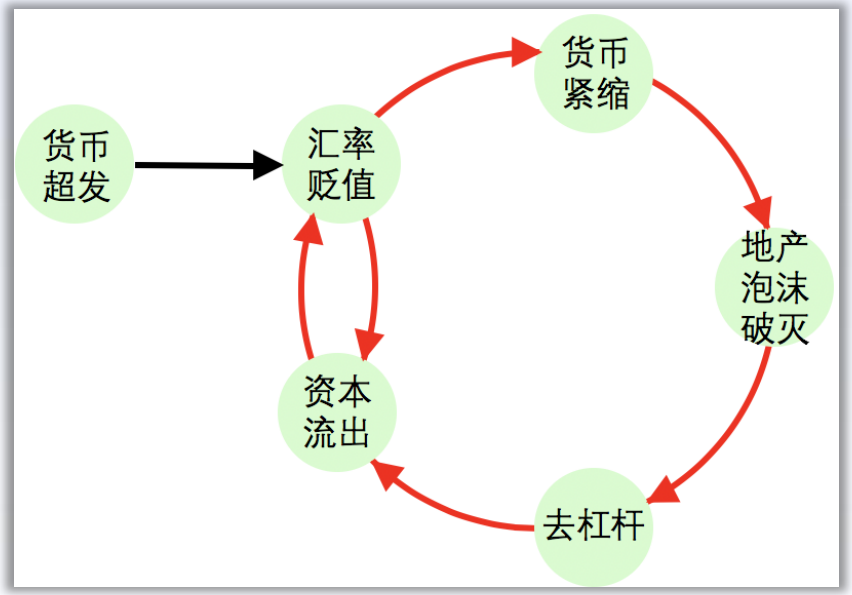
‘结婚’场景下的**树状**事理图谱

事理图谱的三种典型子图结构



金融领域的**环状**事理图谱

事理图谱的三种典型子图结构



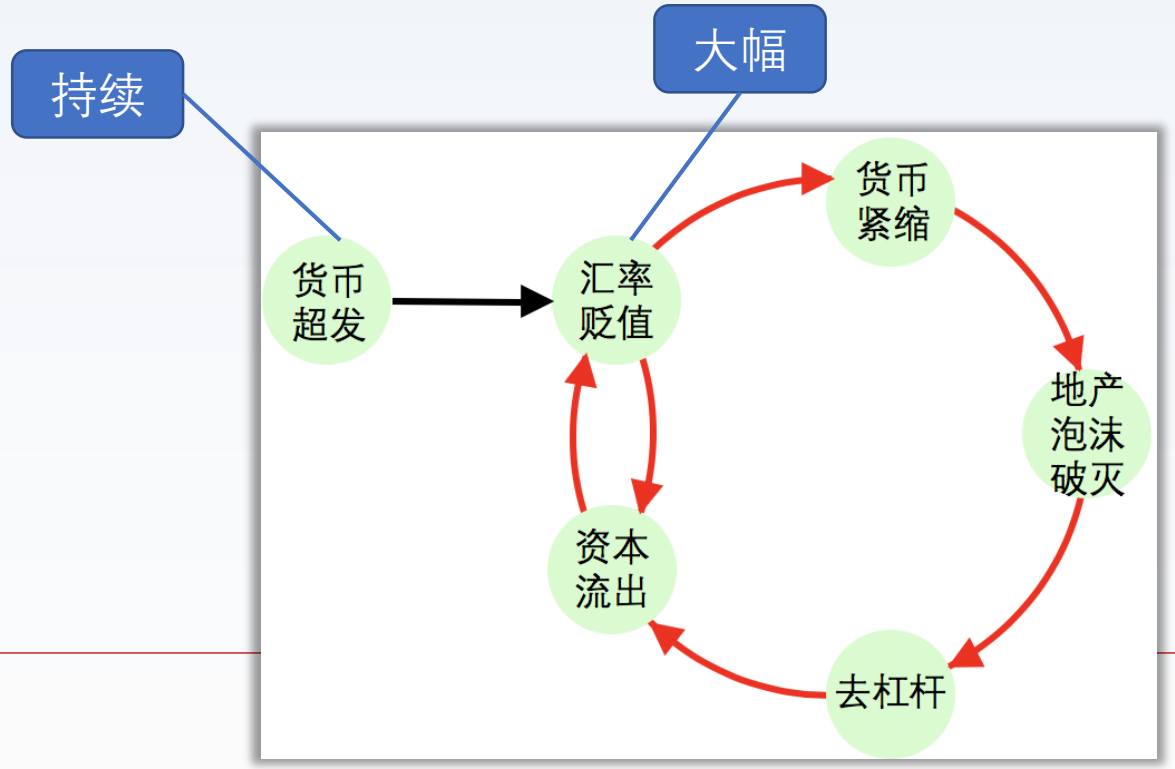
漩涡——汇率机制带来的贬值与资本流出恶性循环

经济和政策基本面冲击的力度大小、外汇市场所处的前定格局都会影响到净短期资本流动规模。汇率形成机制同样会影响净资本流动规模。不同的汇率形成机制下，冲击所带来的净短期资本流动变化大相径庭。浮动汇率下，内部或外部冲击所引发的资本流动压力会导致汇率调整，也因此改变国内外金融资产的相对价格，汇率价格浮动自发地平抑资本流入或者流出。固定汇率下，内部或者外部冲击所引发的资本流动对汇率形成压力，货币当局不得不调整利率，也因此改变国内外金融资产价格相对价格，平抑资本流动并维持固定汇率水平。

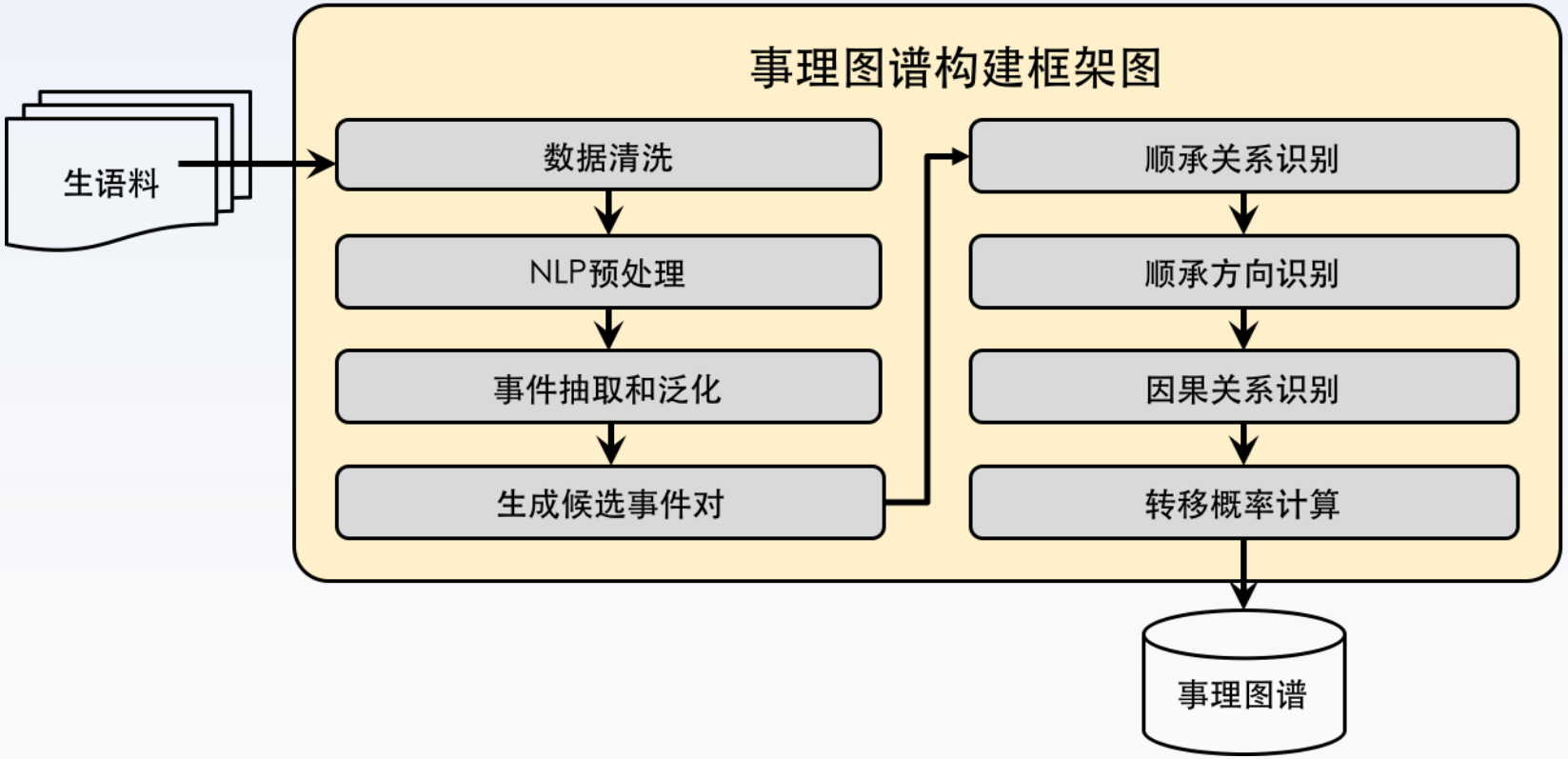
金融领域的**环状**事理图谱

事件属性

- 事理图谱除了关注事件之间的事理逻辑关系外，还关注事件自身的属性
- 事件属性用来描述事件发生的程度、持续时间等



金融事理图谱的构建


事理图谱构建框架图
生语料
数据清洗
NLP预处理
事件抽取和泛化
生成候选事件对
顺承关系识别
顺承方向识别
因果关系识别
转移概率计算
事理图谱

无监督“因果事件关系”抽取方法

- 规则模板+词性标注信息抽取具体因果事件
  - 使用规则<Pattern, Constraint, Priority>匹配因果事件因果提及文本
  - Pattern是抽因果的正则表达式，例如“after [sentence1], [sentence2] ”
  - Constraint 是句法结构约束，例如[sentence1]不能以数字开头
  - Priority表示如果多个Pattern匹配成功时的决策

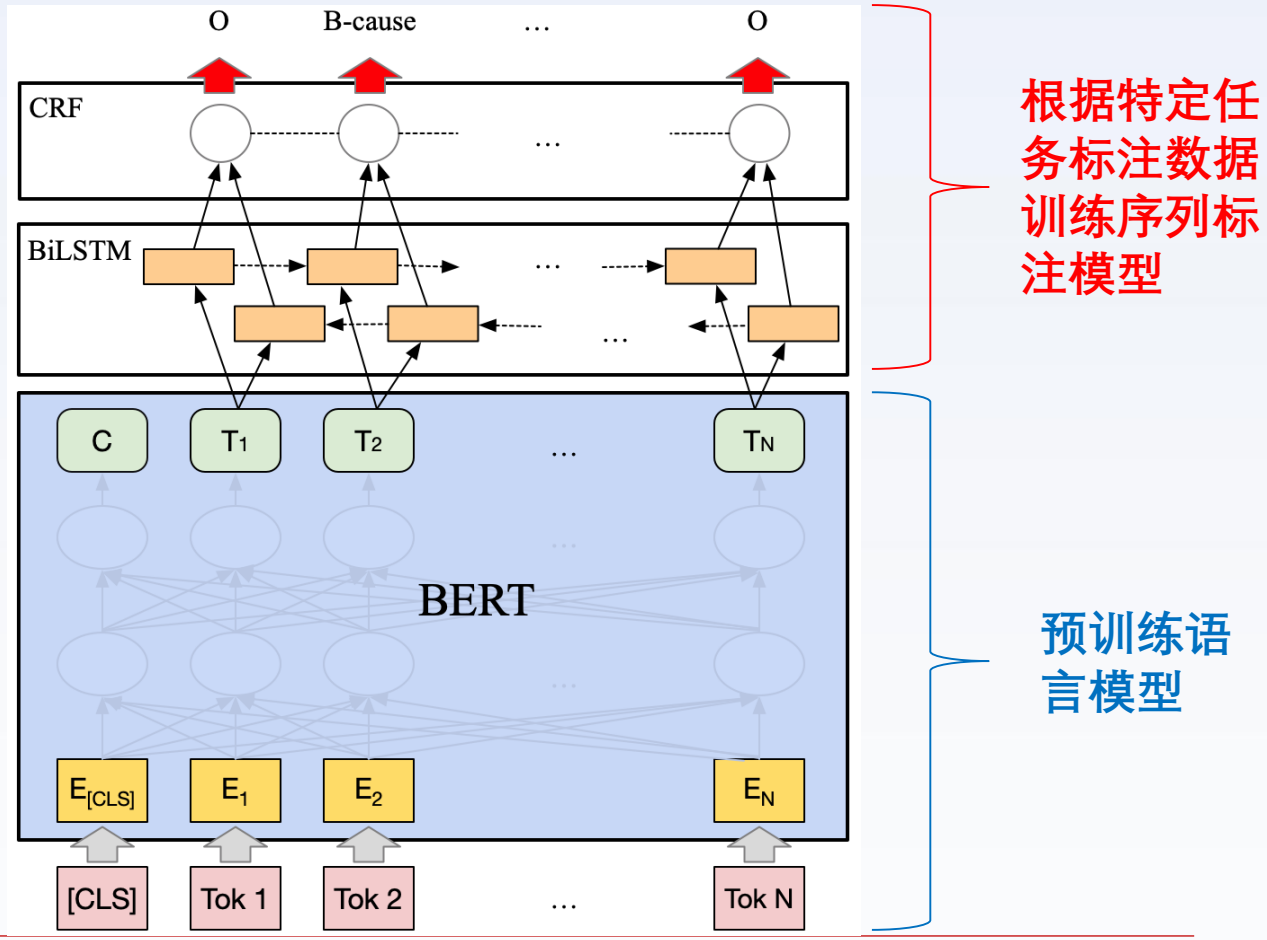
“after Iraq war, the oil price falls” → (Iraq war, the oil price falls)

“After 2 Years, US Capitol Restoration Project Is Complete” ~~→~~ (2 Years, US Capitol Restoration Project Is Complete)

Sendong Zhao, et al. Constructing and Embedding Abstract Event Causality Networks from Text Snippets. In Proc. of ACM WSDM 2017

有监督“因果事件关系”抽取方法

- 基于BERT的因果关系抽取模型
  - 使用BERT进行特征提取
  - 使用BiLSTM+CRF输出标签



Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2019

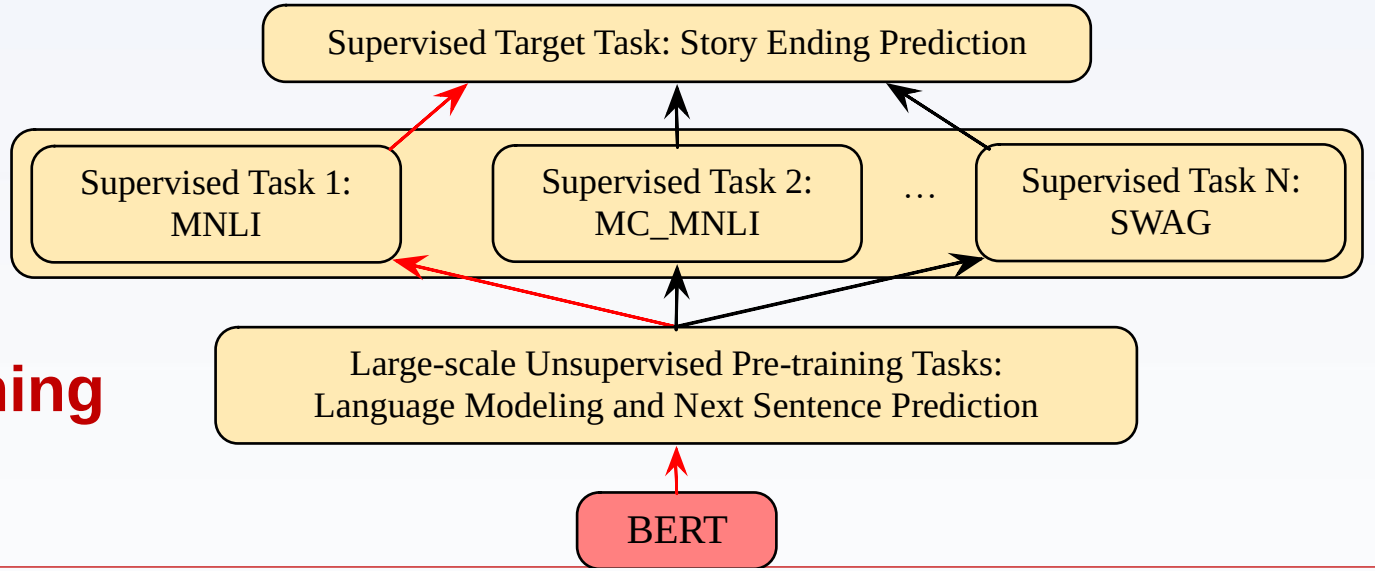
Transferable BERT

- 动机
  - 传统BERT在大规模无监督数据上进行预训练
  - Transferable BERT提出三层训练框架，在原有Pre-training和Fine-tuning框架下，增加在一个相关任务预训练层，学习相关任务的有监督信息

3. Fine-tuning

2. Supervised Pre-training

1. Unsupervised Pre-training



Zhongyang Li, Xiao Ding, Ting Liu. Story Ending Prediction by Transferable BERT. IJCAI, 2019.

有监督“因果事件关系”抽取方法

- 测试集实验结果
  - Token-level

模型	数据集大小	评价方法	Token-level accuracy
BERT	1642	自动评价	**0.86**
全部预测O标签	1642	自动评价	0.70

  - Mention-level

模型	数据集大小	评价方法	Precision	Recall	F1 score
BERT	178	人工评价	0.9060	0.7759	**0.85**
规则模板	178	人工评价	0.7472	0.7644	0.75

因果强度计算（1）

- 基于统计的因果强度计算方法

记$f(i_c, j_e)$为单词i出现在原因中且单词j出现在结果中的频数

使用最大似然估计得到i作为j必要条件的概率：

$$CS_{nec}(i_c, j_e) = \sqrt{\frac{f(i_c, j_e)}{\sum_w (w_c, j_e)}}$$

类似地得到i作为j充分条件的概率：

$$CS_{suf}(i_c, j_e) = \sqrt{\frac{f(i_c, j_e)}{\sum_w (i_c, w_e)}}$$

Luo et al. Commonsense Causal Reasoning between Short Texts. AAAI 2016 Knowledge Representation Workshop

因果强度计算（2）

$$CS_{nec}(i_c, j_e) = \sqrt{\frac{f(i_c, j_e)}{\sum_w (w_c, j_e)}} \qquad CS_{suf}(i_c, j_e) = \sqrt{\frac{f(i_c, j_e)}{\sum_w (i_c, w_e)}}$$

求两个概率的加权几何平均数，作为i与j间的因果强度

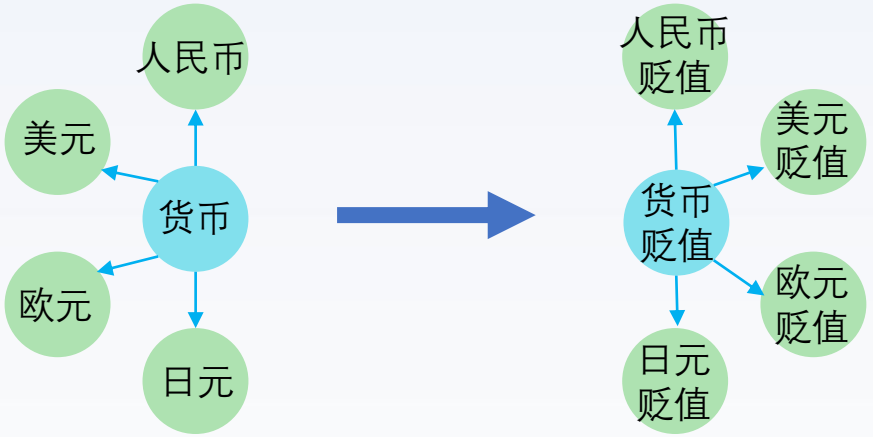
$$CS(i_c, j_e) = CS_{nec}(i_c, j_e)^{\lambda} CS_{suf}(i_c, j_e)^{1-\lambda}$$

给定事件T_1, T_2，将其中所有单词的因果强度组合作为事件的因果强度

$$CS_T(T_1, T_2) = \frac{1}{|T_1| + |T_2|} \sum_{i \in T_1} \sum_{j \in T_2} CS(i, j)$$

事件上下位关系获取

- 基于《大词林》的事件名词性上下位关系获取
  - 对事件中的主体/客体，在大词林中寻找其上位词
  - 若找到，将该词替换为上位词，替换后的事件作为上位事件



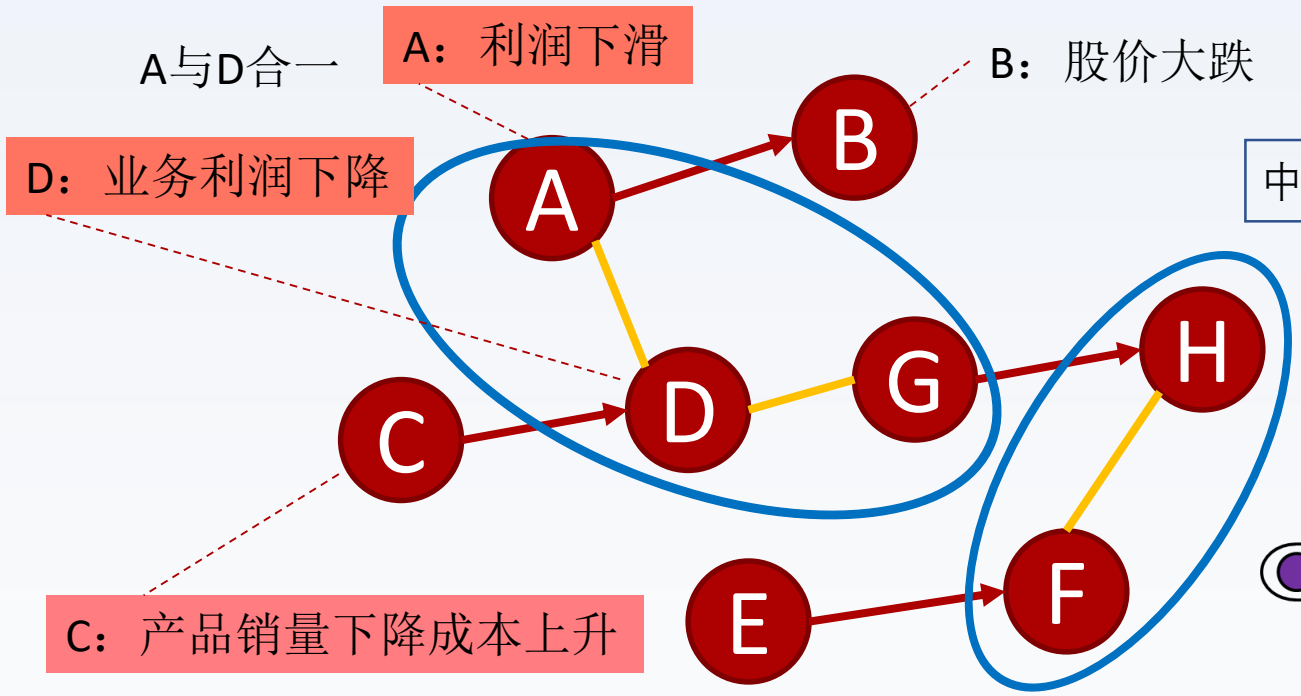
基于《大词林》获取到的上位事件数量： 1,883,792

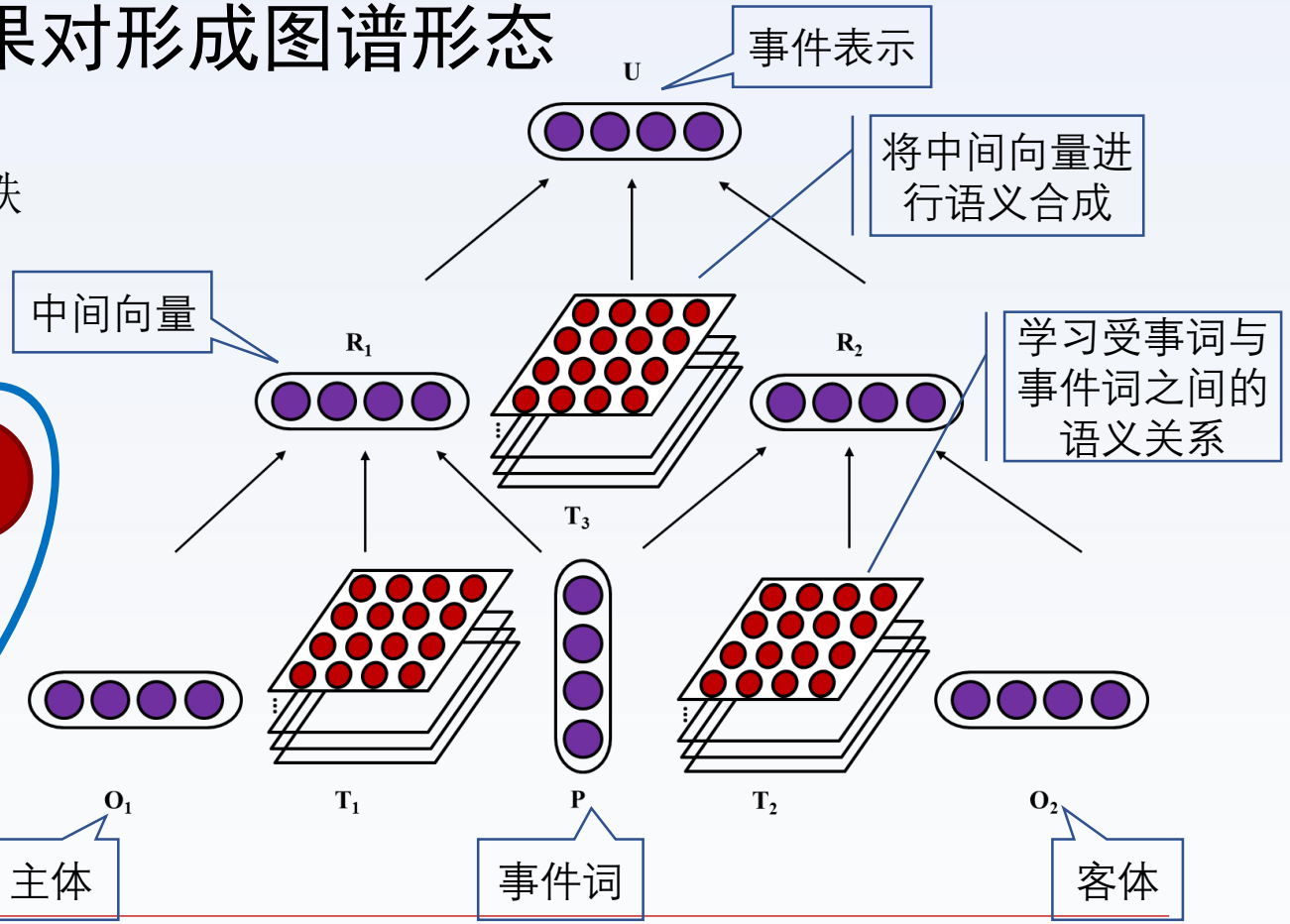
事件合一

$$R_1 = f(O_1^T T_1^{[1:k]} P + W \begin{bmatrix} O_1 \\ P \end{bmatrix} + b)$$

$$loss(E, E^r) = \max(0, 1 - f(E) + f(E^r)) + \lambda \|\Phi\|_2^2$$

- 通过对某些事件的合一，将因果对形成图谱形态

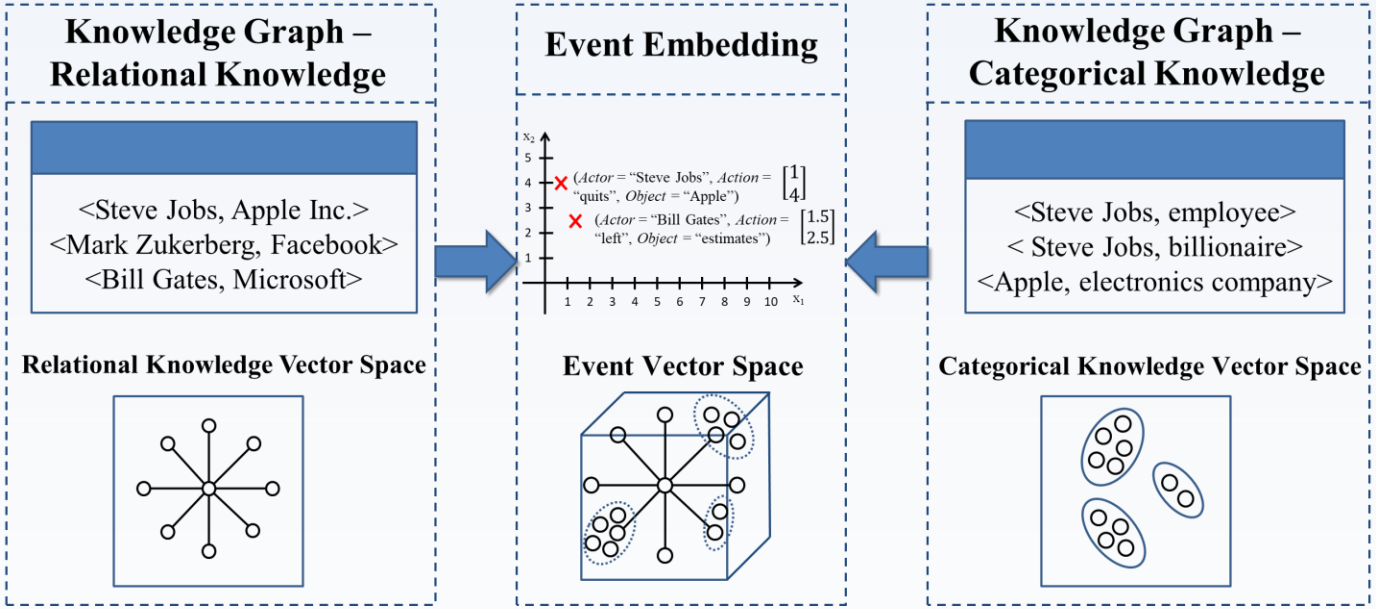




Xiao Ding, Yue Zhang, Ting Liu, Junwen Duan. Deep Learning for Event-Driven Stock Prediction. IJCAI, 2015.

知识驱动的事件表示学习

- 缺乏知识造成的歧义
  - 具有相同或相似的词但是语义不同的事件，学到相似的事件向量
  - 不具有相同或相似的词但是语义相同的事件，学到不相邻的事件向量
  - 举例
    - 乔布斯离开苹果　*vs*　约翰离开星巴克



Xiao Ding, Yue Zhang, Ting Liu, Junwen Duan. Knowledge-Driven Event Embedding for Stock Prediction. Coling, 2016.

知识驱动的事件表示学习模型

联合学习模型：

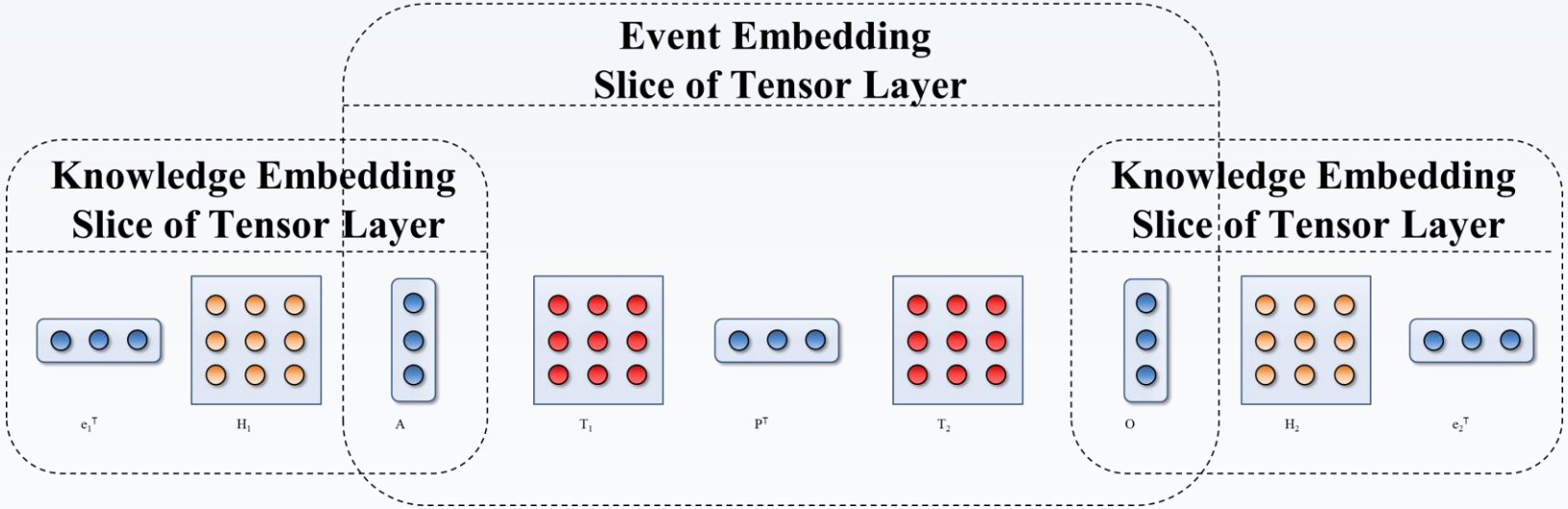
$$L = \alpha L_{\mathcal{E}} + (1-\alpha) L_{\mathcal{K}}$$

事件表示学习：

$$L_{\mathcal{E}} = loss(E, E^r) = \max(0, 1 - g(E) + g(E^r)) + \lambda \| \Phi \|_2^2,$$

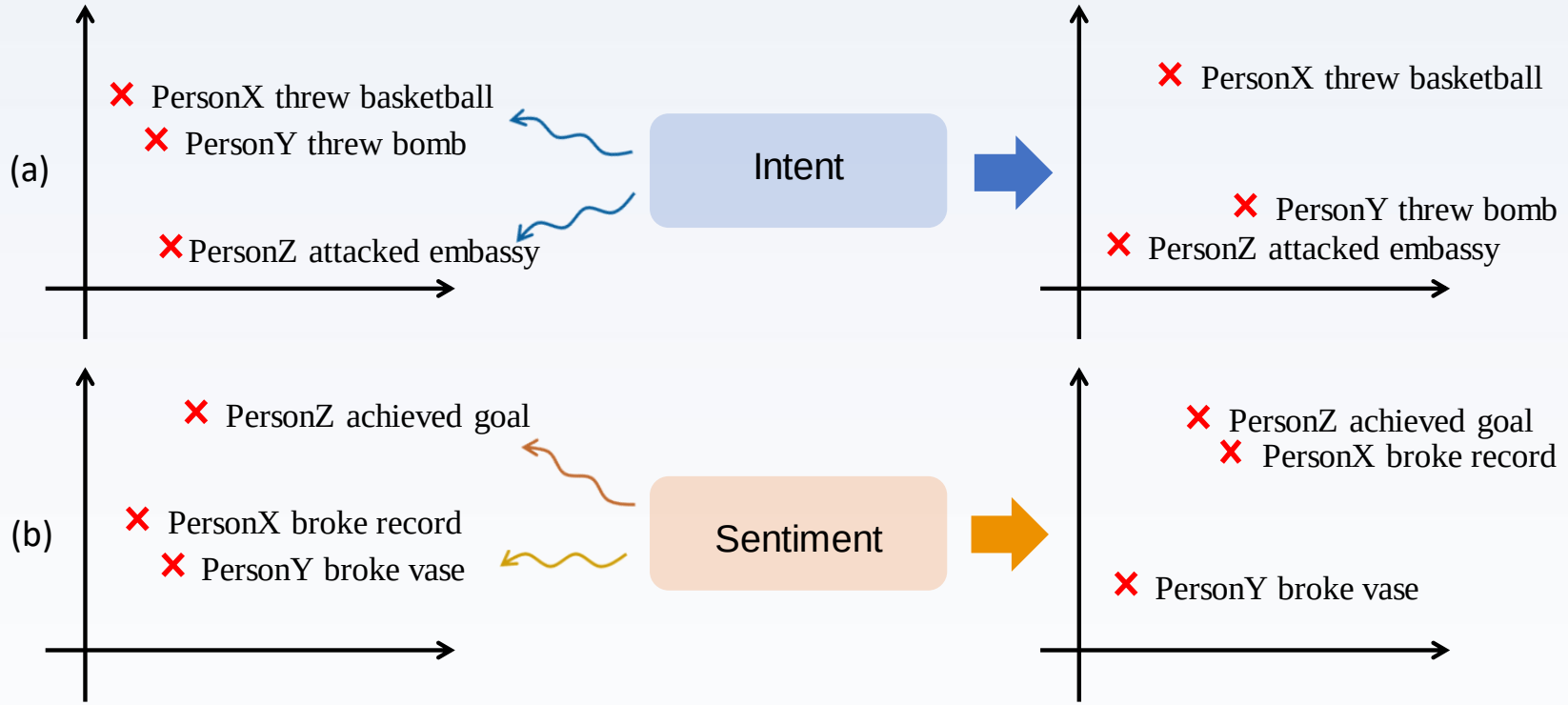
知识图谱表示学习：

$$L_{\mathcal{K}} = \sum_{i=i}^{N} \sum_{m=1}^{M} max\left(0, 1 - g(T^{(i)}) + g(T_c^{(i)})\right) + \lambda \|\Omega\|_2^2,$$



常识知识（情感、意图）增强的事件表示学习

- 事件是对客观事实的表达，然而客观事件的发生会对人类的主观情感产生影响，不同的事件其背后的意图也有所不同



Xiao Ding, Kuo Liao, Ting Liu, Junwen Duan, Zhongyang Li. Event Representation Learning Enhanced with External Commonsense Knowledge. EMNLP, 2019.

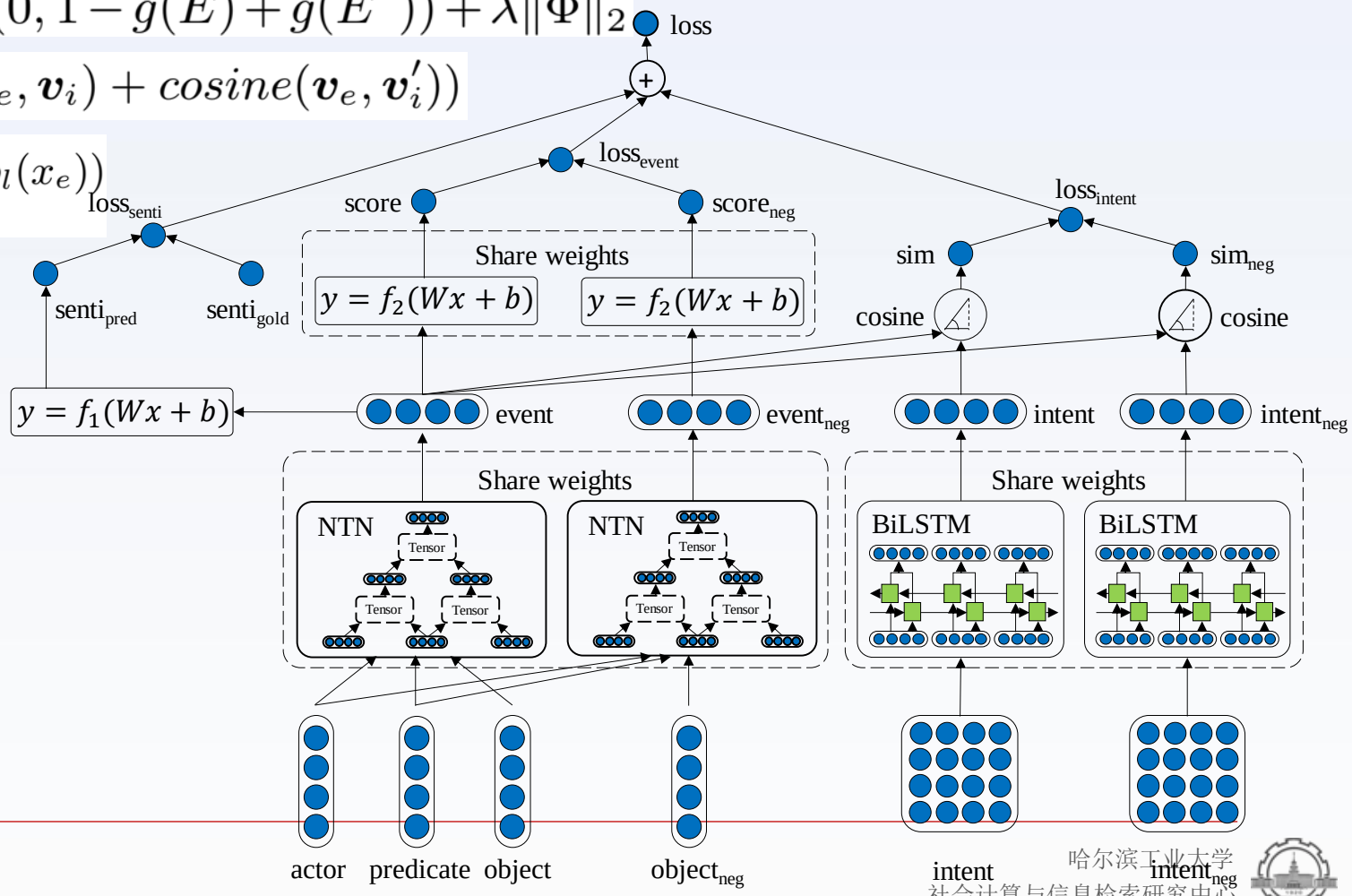
常识知识（情感、意图）增强的事件表示学习

多任务学习模型： $\mathcal{L} = \alpha\mathcal{L}_{\mathcal{E}} + \beta\mathcal{L}_{\mathcal{I}} + \gamma\mathcal{L}_{\mathcal{S}}$

事件表示学习： $\mathcal{L}_{\mathcal{E}} = loss(E, E^r) = \max(0, 1 - g(E) + g(E^r)) + \lambda\|\Phi\|_2^2$

意图表示学习： $\mathcal{L}_{\mathcal{I}} = max(0, 1 - cosine(\boldsymbol{v}_e, \boldsymbol{v}_i) + cosine(\boldsymbol{v}_e, \boldsymbol{v}_i'))$

情感表示学习： $\mathcal{L}_{\mathcal{S}} = -\sum_{x_e \in C}\sum_{l \in L} p_l^g(x_e) \cdot log(p_l(x_e))$



事件泛化

- 基于事件聚类的泛化方法
  - 相似事件聚成一类，取类别中所有相似事件中公共成分作为泛化事件
  - 同时“具体因果边”也被泛化成“抽象因果边”

Sendong Zhao, et al. Constructing and Embedding Abstract Event Causality Networks from Text Snippets. In Proc. of ACM WSDM 2017

事理图谱应用——脚本事件预测

- Script（脚本） in artificial intelligence (Schank and Abelson, 1977)
  - 描述特定场景/上下文的事件序列

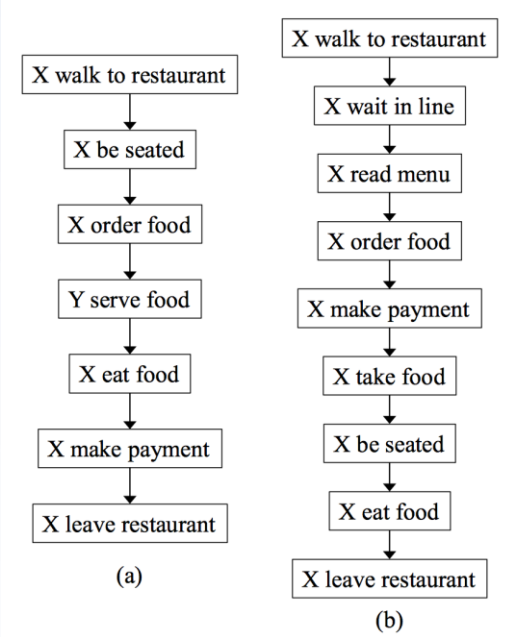
- 例子：餐厅脚本

  场景1：进入餐馆

  场景2：点单

  场景3：吃饭

  场景4：离开餐馆



Zhongyang Li, Xiao Ding, Ting Liu. Constructing Narrative Event Evolutionary Graph for Script Event Prediction. In Proc. of IJCAI, 2018.

事理图谱应用——脚本事件预测

- 任务
  - 给定事件上文，从候选事件列表中选出接下来最有可能发生的事件

角色

X=顾客 Y=服务员

事件上下文

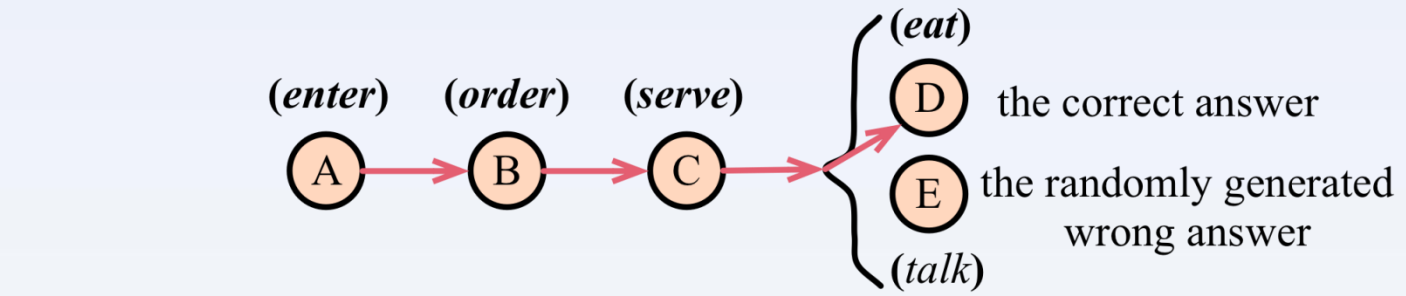
去(X, 餐厅)， 坐下(X)， 读(X, 菜单)， 点(X, 菜)，
上(Y, X, 菜)， 吃(X, 菜)， 付款(X)， _____

c_1:获得(X,奖学金)
c_2:开车(X)
c_3:花费(X,时间)
c_4:联系(X,设计师)
c_5:离开(X,餐厅)

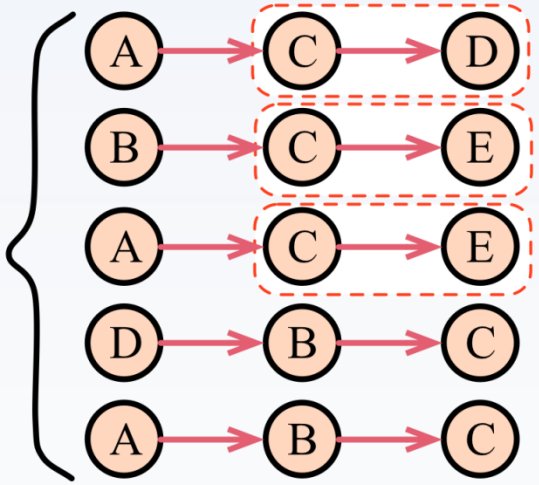
?

基于事理图谱的脚本事件预测研究动机

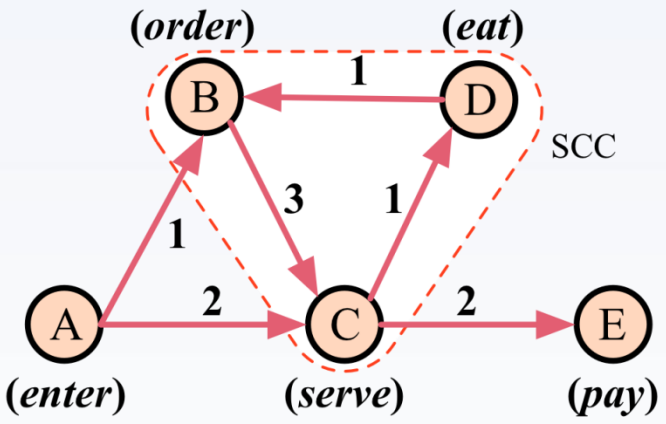
- 前人工作探索了如何利用event pair和event chain进行预测，我们的工作探索如何利用event graph的稠密连接信息来帮助事件预测
- 图结构学习到的事件表示更加利于预测
- 图结构能克服事件之间的不连通性，学习更好的事件关联



(a) Given an event context (A, B, C), choose the subsequent event from D and E.



(b) Training event chains.

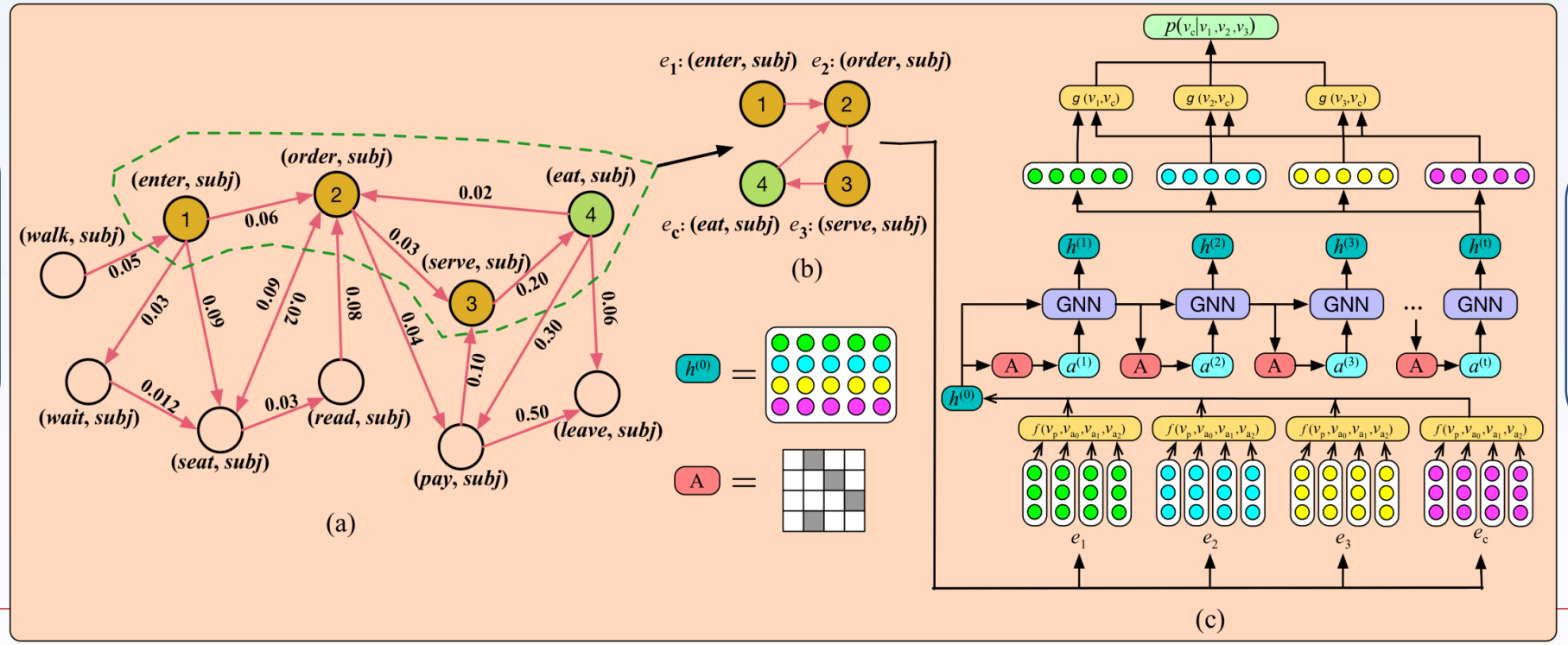


(c) Narrative event graph based on event chains in (b).

基于可扩展的图神经网络脚本事件预测方法

- 我们提出了可扩展的图神经网络(Scaled Graph Neural Network-SGNN)来解决图上的预测、推理问题
  - 重点解决了GNN在大规模图结构上无法进行推理的问题

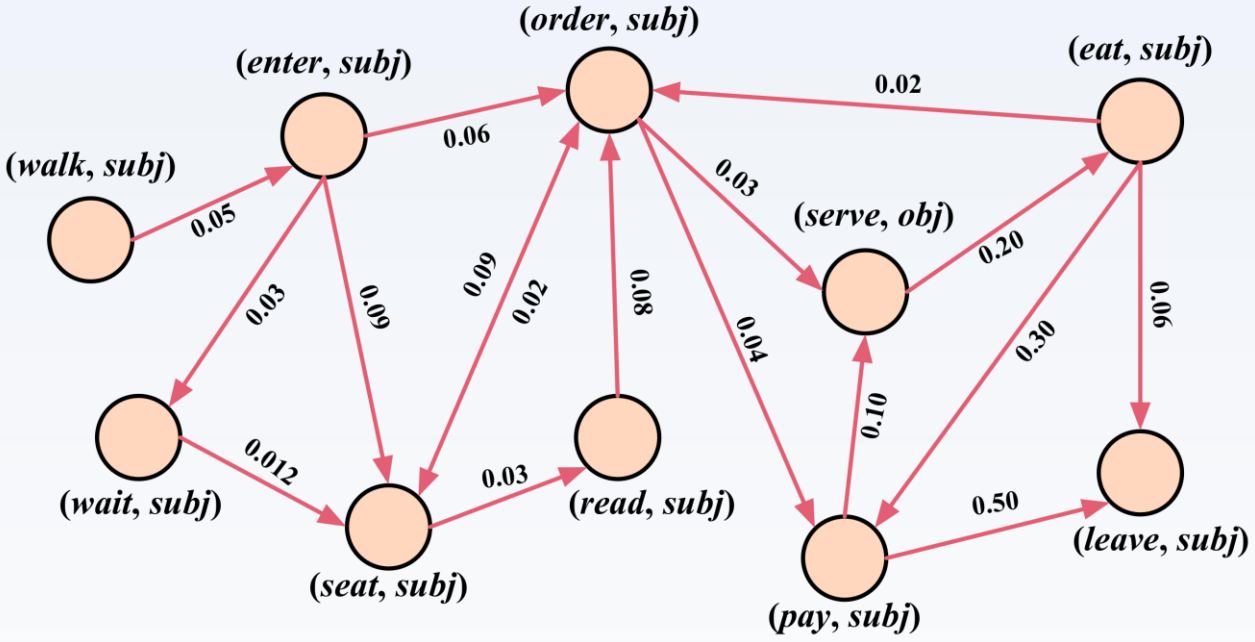
步骤一：基于自动抽取的叙事事件链条构建叙事事理图谱



步骤二：提出可扩展图神经网络进行事理图谱网络表示学习，用以解决大规模图上的推理问题

步骤一：构建完成的叙事事理图谱

- 构建完成的叙事事理图谱包含
  - 10万个节点
  - 618万条有向边



步骤二：网络表示学习—学习初始事件表示

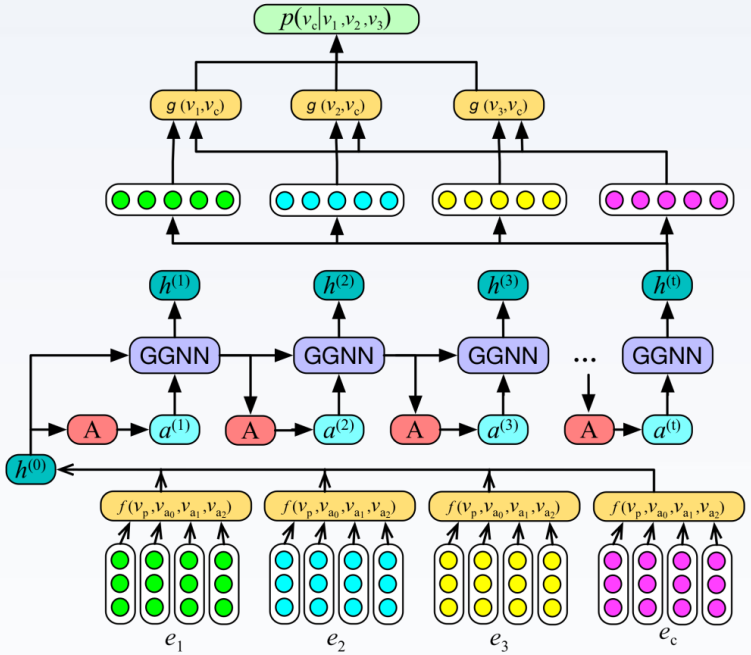
- 学习初始事件表示
  - 给出一个事件 $e_i = \{p(a_0, a_2, a_2)\}$ 以及动词和各事件元素的词向量：$v_p, v_{a_0}, v_{a_1}, v_{a_2} \in \mathbb{R}^d$，得到整个事件的表示：$v_{e_i} = f(v_p, v_{a_0}, v_{a_1}, v_{a_2})$

- **Average:** Use the mean value of the verb and all arguments vectors as the representation of the whole event.
- **Nonlinear Transformation** [Wang *et al.*, 2017]:

$$v_e = \tanh(W_p \cdot v_p + W_0 \cdot v_{a_0} + W_1 \cdot v_{a_1} + W_2 \cdot v_{a_2} + b) \quad (2)$$

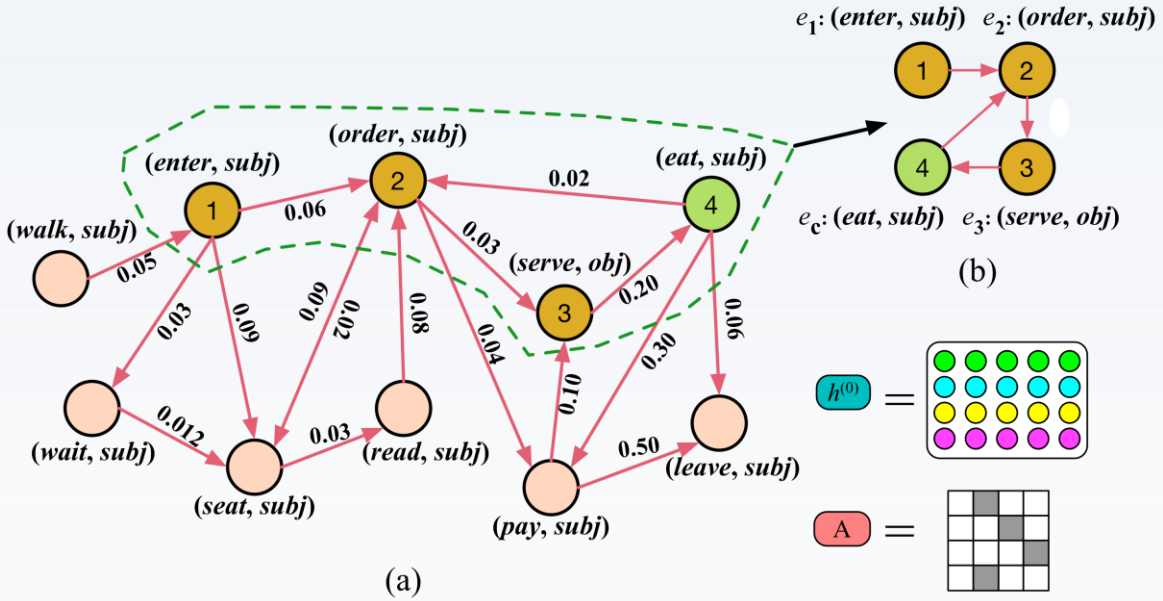
where W_p, W_0, W_1, W_2, b are model parameters.

- **Concatenation** [Granroth-Wilding and Clark, 2016]: Concatenate the verb and all argument vectors as the representation of the whole event.



步骤二：网络表示学习—输入

- 通过SGNN更新事件表示（基于GGNN, Li et al. 2015改进得到）
  - SGNN的输入包括两个矩阵$h^{(0)}$ 和 A



$h^{(0)}$表示事件上文以及所有候选事件的向量表。

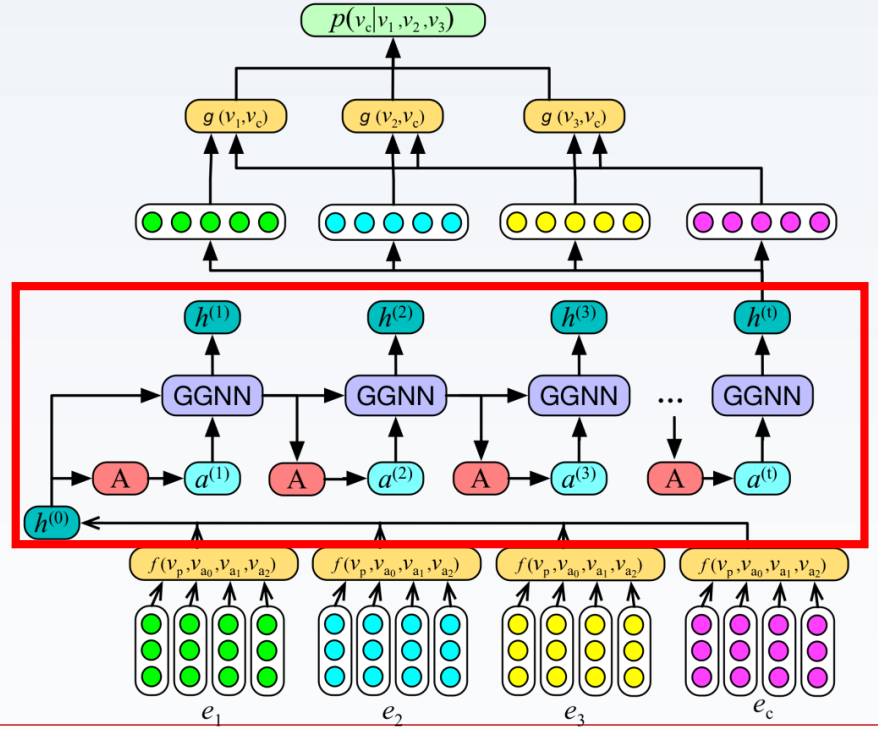
A是相应的子图邻接矩阵：

$$\mathbf{A}[i,j] = \begin{cases} w(\mathbf{v}_j|\mathbf{v}_i), & \text{if } \mathbf{v}_i \rightarrow \mathbf{v}_j \in E, \\ 0, & \text{others.} \end{cases}$$

步骤二：网络表示学习—更新参数

- 通过SGNN更新事件表示（基于GGNN, Li et al. 2015改进得到）
  - SGNN的更新部分与GGNN和GRU的门控制类似

$$
\begin{aligned}
a^{(t)} &= \mathbf{A}^\top h^{(t-1)} + b \\
z^{(t)} &= \sigma(W^z a^{(t)} + U^z h^{(t-1)}) \\
r^{(t)} &= \sigma(W^r a^{(t)} + U^r h^{(t-1)}) \\
c^{(t)} &= \tanh(W a^{(t)} + U(r^{(t)} \odot h^{(t-1)})) \\
h^{(t)} &= (1 - z^{(t)}) \odot h^{(t-1)} + z^{(t)} \odot c^{(t)}
\end{aligned}
$$



步骤二：预测正确事件结尾

- 挑选正确的后续事件

第i个上文事件与第j个候选事件的相关性得分计算：

$$s_{ij} = g(h_i^{(t)}, h_{c_j}^{(t)})$$

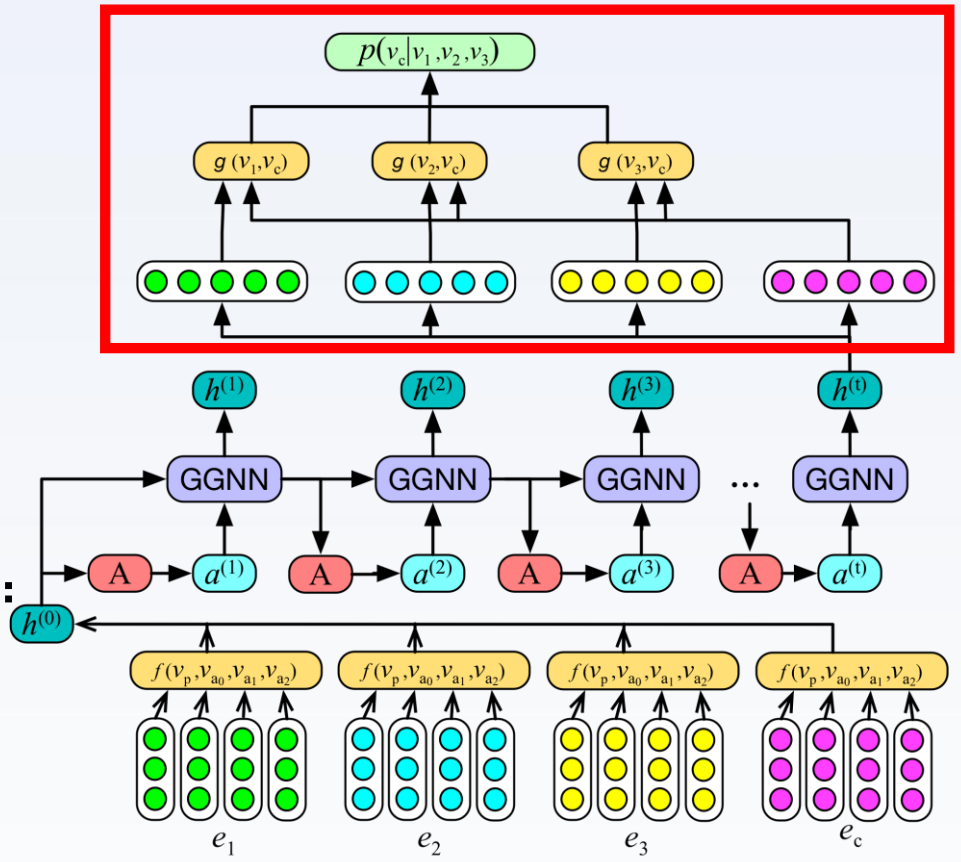
给出上文 $e_1, e_2, \ldots, e_n$，候选事件 e_{c_j} 的可能性打分为：

$$s_j = \frac{1}{n}\sum_{i=1}^{n} s_{ij}$$

然后通过选择得分最高的候选事件作为正确答案：

$$c = \max_j s_j$$

损失函数：$L(\Theta) = \sum_{I=1}^{N}\sum_{j=1}^{k}(\max(0, margin - s_{Iy} + s_{Ij})) + \frac{\lambda}{2}||\Theta||^2$

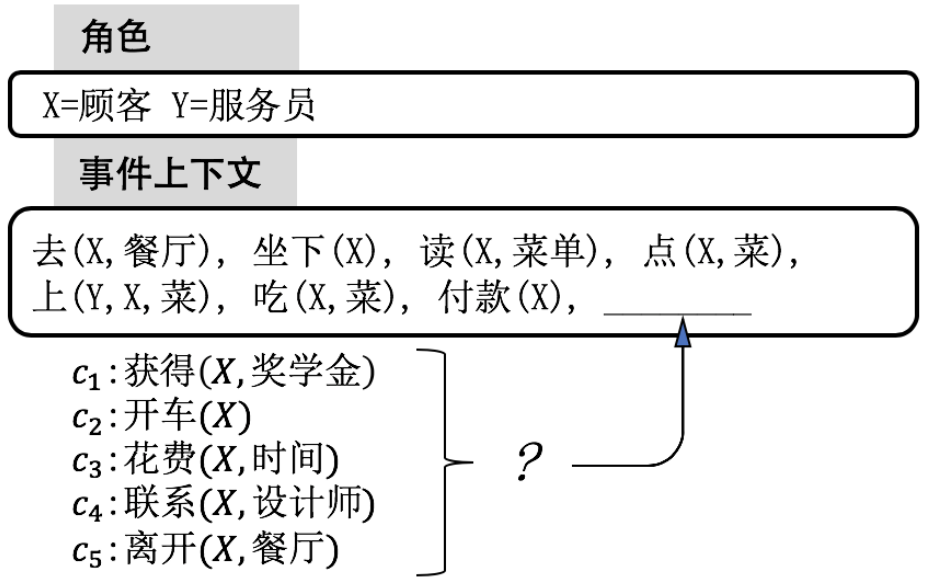


实验数据

- 数据集: Gigaward语料NYT部分

	Training	Development	Test
#Documents	830,643	103,583	103,805
#Chains for NEEG	5,997,385	-	-
#Chains for SGNN	140,331	10,000	10,000

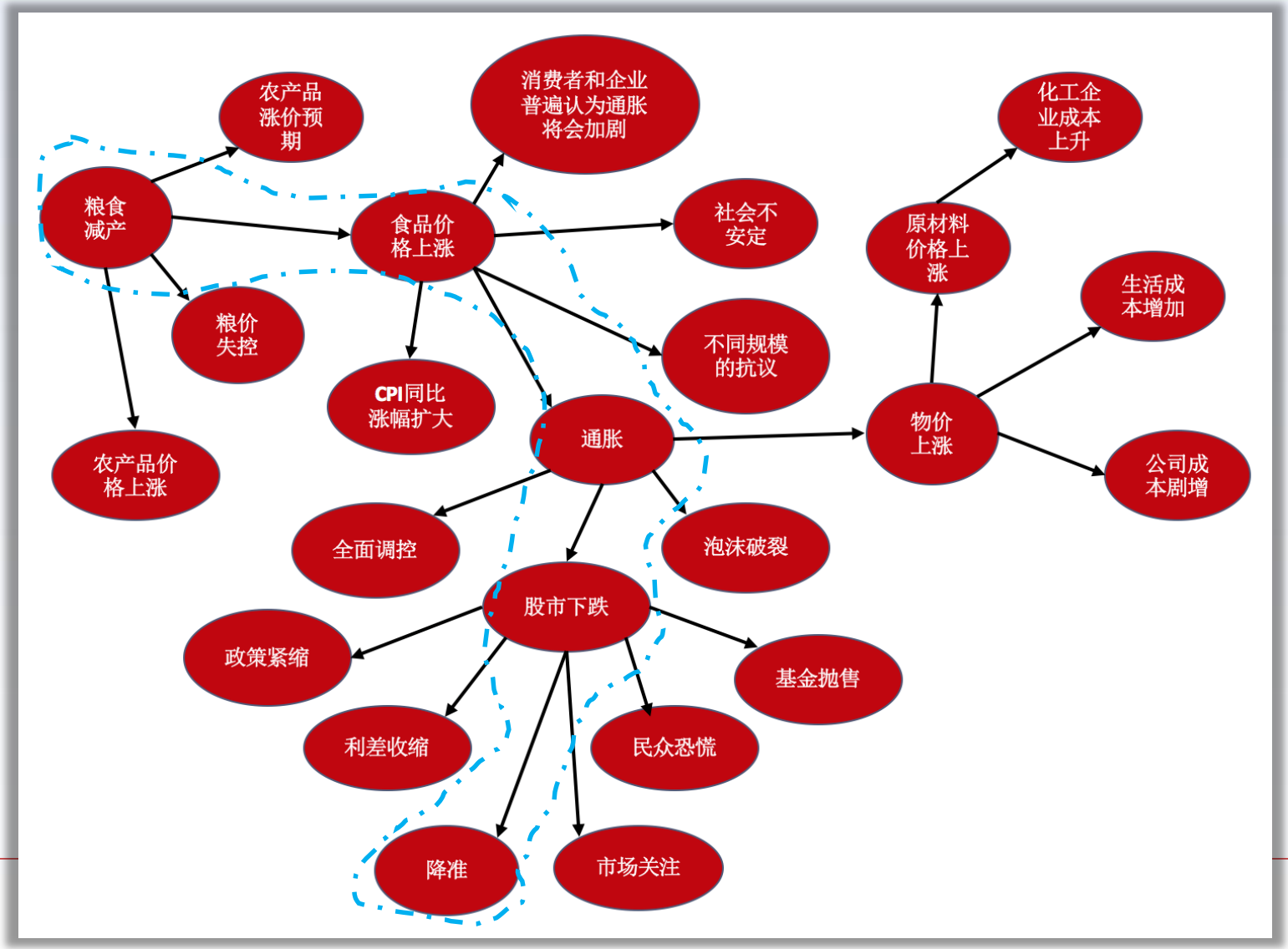
- 评估: Multiple Choice Narrative Cloze (MCNC)



实验结果

Methods	Accuracy
Random	20.00
PMI [Chambers and Jurafsky, 2008]	30.52
Bigram [Jans *et al.*, 2012]	29.67
Word2vec [Mikolov *et al.*, 2013]	42.23
DeepWalk [Perozzi *et al.*, 2014]	43.01
EventComp [Granroth-Wilding and Clark, 2016]	49.57
PairLSTM [Wang *et al.*, 2017]	50.83
SGNN-attention (without attention)	51.56
SGNN (ours)	**52.45**
SGNN+PairLSTM	52.71
SGNN+EventComp	54.15
SGNN+EventComp+PairLSTM	**54.93**

金融事理图谱样例


农产品涨价预期
消费者和企业普遍认为通胀将会加剧
化工企业成本上升
粮食减产
食品价格上涨
社会不安定
原材料价格上涨
生活成本增加
粮价失控
不同规模的抗议
CPI同比涨幅扩大
通胀
物价上涨
公司成本剧增
农产品价格上涨
全面调控
泡沫破裂
股市下跌
政策紧缩
基金抛售
利差收缩
民众恐慌
降准
市场关注

2018年《金融事理图谱V1.0版本》正式发布

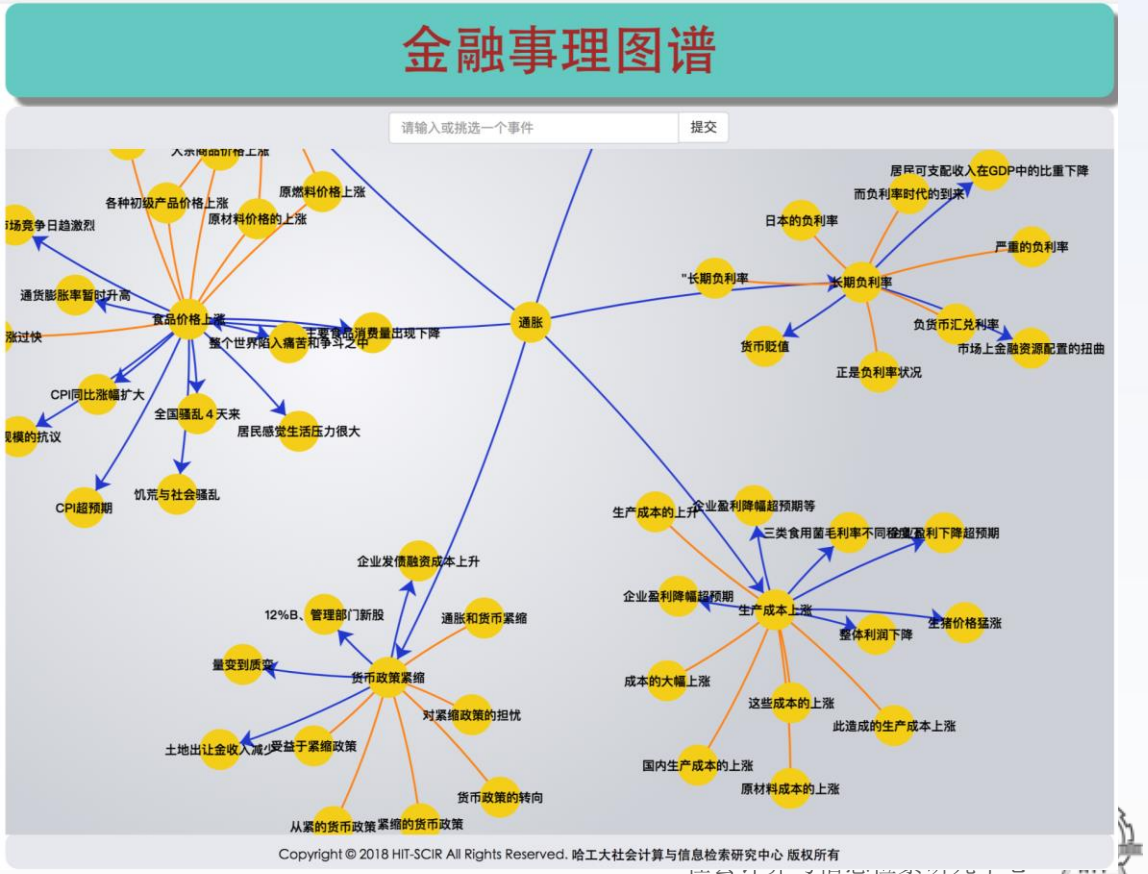
- 2018年9月10日， HIT-SCIR正式发布《金融事理图谱V1.0版本》
- 《金融事理图谱》是在金融领域自动构建的基于因果关系的大规模中文事理逻辑知识库

哈工大SCIR
18-9-10 来自专业版微博

哈工大社会计算与信息检索研究中心提出了事理图谱的概念，并基于大规模财经新闻文本构建了一个金融领域事理图谱。本文中我实验室正式对外发布该金融事理图谱Demo（网页链接）。链接：

赛尔新闻 | 哈工大SCIR正式对外发布金融事理图谱Demo V1.0

知识图谱等现有知识库普遍以名词性实体为核...



金融事理图谱V2. 0版本

	金融事理图谱v1.0	金融事理图谱v2.0
URL	http://eeg.8wss.com/	http://elg.8wss.com/
数据源	北京语言大学新闻语料、腾讯、网易、股吧、和讯、Resset等网站的金融新闻	北京语言大学新闻语料、腾讯、网易、股吧、和讯、Resset、新浪、中财网、中金在线、证券之星等网站的金融新闻、金融领域微信公众号、研报
数据规模（文档数）	11,653,062	12,121,279（增加100余万篇文档）
具体事件数	1,348,459	2,187,086（增加80余万个事件）
具体因果关系数	1,410,642	1,607,638（增加近20万个因果关系）
抽象事件数	-	2,021,289
抽象因果关系数	-	414,777
上下位关系数	-	1,883,792
因果关系抽取F值	59.54%	85.12%（提升近16%）

事理图谱的潜在应用

- **消费意图识别与推荐**
  - 几乎所有事件都可以和“想”、“打算”、“计划”连起来形成一个意图事件
    - 想去北京，想看电影，计划去爬泰山，想吃火锅
  - 部分节点是消费意图显著节点，可以触发一系列后续消费事件，找到这类节点是很有价值的
    - 逛街，看电影，去旅游等等

事理图谱的潜在应用

- **辅助对话生成**
  - 去北京→买票
  - A:我打算去北京玩。 B:那你买票了吗?
  - 去泰安→爬泰山
  - A:我想去爬泰山。 B:那你得先到泰安。
- **问答系统**
  - 爬泰山有什么注意事项?
  - 租大衣、带水、带手电等等

事理图谱研究基础

- Zhongyang Li, Xiao Ding, Ting Liu. Story Ending Prediction by Transferable BERT. In Proc. of **IJCAI**, 2019.
- Xiao Ding, Kuo Liao, Ting Liu, Junwen Duan, Zhongyang Li. Event Representation Learning Enhanced with External Commonsense Knowledge. **EMNLP**, 2019.
- Li Du, Xiao Ding, Ting Liu, Zhongyang Li. Modeling Event Background for If-Then Commonsense Reasoning Using Context-aware Variational Autoencoder. **EMNLP**, 2019.
- Zhongyang Li, Xiao Ding, Ting Liu. Constructing Narrative Event Evolutionary Graph for Script Event Prediction. In Proc. of **IJCAI**, 2018: 4201-4207.
- Xiaocheng Feng, Lifu Huang, Duyu Tang, Bing Qin, Heng Ji, Ting Liu. A Language-Independent Neural Network for Event Detection. **ACL** 2016.
- Sendong Zhao, Quan Wang, Sean Massung, Bing Qin, Ting Liu, Chengxiang Zhai. Constructing and Embedding Abstract Event Causality Networks from Text Snippets. In Proc. of ACM **WSDM** 2017.
- Zhongyang Li, Sendong Zhao, Xiao Ding, Ting Liu. EEG: Knowledge Base for Event Evolutionary Principles and Patterns[C]. Chinese National Conference on Social Media Processing. Springer, Singapore, 2017: 40-52.
- Xiao Ding, Bing Qin, Ting Liu. BUEES: a Bottom-Up Event Extraction System. Frontiers of Information Technology & Electronic Engineering, 2015, 16(7): 541-552.
- Xiao Ding, Bing Qin, Ting Liu. *Building Chinese Event Type Paradigm Based on Trigger Clustering*. In Proc. of IJCNLP, 2013: 311-319.
- Sendong Zhao, Ting Liu, Sicheng Zhao, Yiheng Chen, Jian-Yun Nie. Event causality extraction based on connectives analysis. Neurocomputing (2016)173: 1943-1950.

总结

- 知识图谱在各个领域精耕细作，逐渐显露价值
- 知识表示形式有待突破，推理能力有待提高
- 以“谓词性短语”为节点，以事件演化（顺承、因果、上下位等）为边的事理图谱方兴未艾
- 事理图谱必将在预测、推荐、对话等领域发挥重要作用，有力地提升人工智能系统的可解释性
- 期待大家共同推动事理图谱的相关研究和产业应用

谢 谢！