

事理图谱：事件演化的规律和模式

作者：李忠阳，赵森栋，丁效

引言

事件是人类社会的核心概念之一，人们的社会活动往往是事件驱动的。事件之间在时间上相继发生的演化规律和模式是一种十分有价值的知识。然而，当前无论是知识图谱还是语义网络等知识库的研究对象都不是事件。为了揭示事件的演化规律和发展逻辑，本文提出了事理图谱的概念，作为对人类行为活动的直接刻画。在图结构上，与马尔科夫逻辑网络（无向图）、贝叶斯网络（有向无环图）不同，事理图谱是一个有向有环图。现实世界中事件演化规律的复杂性决定了我们必须采用这种复杂的图结构。为了展示和验证事理图谱的研究价值和应用价值，我们从互联网非结构化数据中抽取、构建了一个出行领域事理图谱。初步结果表明，事理图谱可以为揭示和发现事件演化规律与人们的行为模式提供强有力的支持。

事理图谱的定义

首先，给出事件、事件间顺承和因果关系的定义。**事理图谱中的事件用抽象、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。**抽象和泛化指不关注事件的具体发生时间、地点和具体施事者，语义完备指人类能够理解该短语传达出的意义，不至于过度抽象而让人产生困惑。例如，“吃火锅”，“看电影”，“去机场”，是合理的事件表达；而“去地方”，“做事情”，“吃”，是不合理或不完整的事件表达。后面三个事件因为过度抽象而让人不知其具体含义是什么。**事件间顺承关系指两个事件在时间上先后发生的偏序关系；在英语体系研究中一般就叫做时序关系**

(Temporal Relation)，本文认为两者是等价的。例如，“小明吃过午饭后，付完账离开了餐馆。”吃饭、付账、离开餐馆，这三个事件构成了一个顺承关系链条。**事件间因果关系指在满足顺承关系时序约束的基础上，两个事件间有很强的因果性，强调前因后果。**例如，“日本核泄漏引起了严重的海洋污染”。“日本核泄漏”和“海洋污染”两个事件间就是因果关系，“日本核泄漏”是因，“海洋污染”是果，并且满足因在前，果在后的时序约束关系。事件顺承关系是比因果关系更广泛的存在。

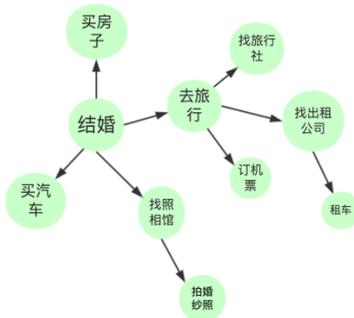


图 1：“结婚”场景下的树状事件演化图

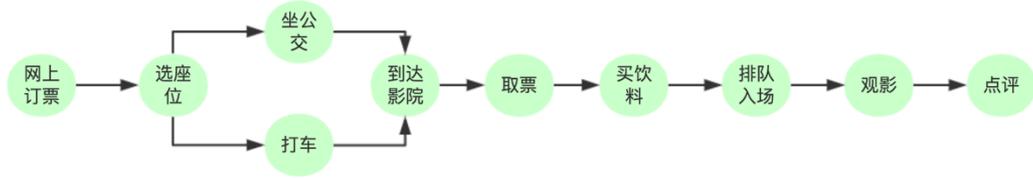


图 2：“看电影”场景下的链状事件演化图

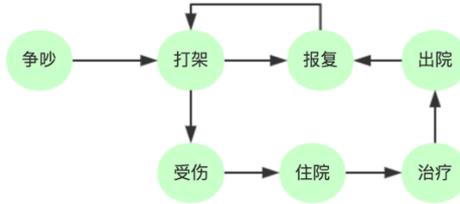


图 3：“打架”场景下的环状事件演化图

事理图谱 (Event Evolutionary Graph) 是一个描述事件之间顺承、因果关系的事理演化逻辑有向图。图中节点表示抽象、泛化的事件，有向边表示事件之间顺承、因果关系。边上还标注有事件间转移概率信息。图 1，图 2，图 3 分别展示了事理图谱中 3 个不同场景下，不同图结构的局部事件演化模式图。这种常识性事件演化规律往往隐藏在人们的日常行为模式中，或者用户生成的文本数据中，而没有显式地以知识库的形式存储起来。事理图谱旨在揭示事件间的逻辑演化规律与模式，并形成一个大类型常识事理知识库，作为对人类行为活动的直接刻画。

表 1. 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	事件间顺承、因果关系，以及转移概率信息	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

事理图谱与传统知识图谱有本质上的不同。如表 1 所示，事理图谱以事件为核心研究对象，有向边只表示两种事理关系，即顺承和因果；边上标注有概率信息说明事理图谱是一种事件间相继发生可能性的刻画，不是确定性关系。而知识图谱以实体为核心研究对象，实体属性以及实体间关系种类往往成千上万。知识图谱以客观真实性为目标，某一条属性或关系要么成立，要么不成立。

出行领域事理图谱

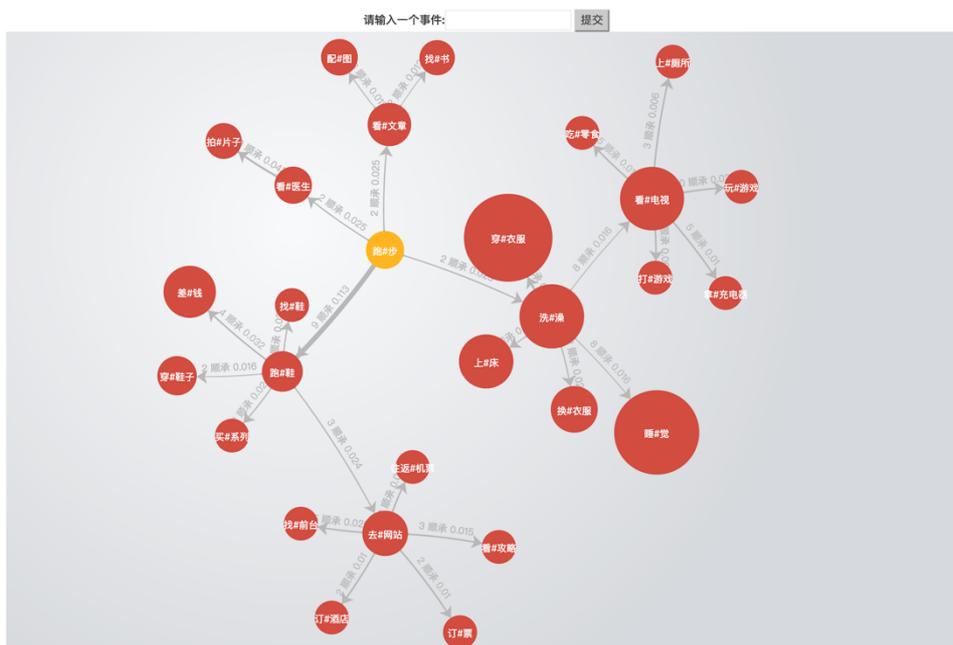


图 4：出行领域事理图谱 Demo

基于上文相关定义，我们从互联网无结构化数据构建了一个中文出行领域事理图谱。采用的语料是知乎“旅行”话题下的 32 万篇用户问答对。构建过程包括事件抽取、事件间顺承和因果关系识别、事件转移概率计算等步骤。图 4 是该事理图谱的 Demo 展示。以“跑步”作为输入事件，我们采用广度优先搜索向外扩展，形成了图 4 中以“跑步”为核心事件所扩展出来的局部事理关系图。从该图中，我们至少可以发现 3 个有趣的事件演化链条。“跑步、看医生、拍片子”的分支属于“运动受伤”场景下的事件演化模式，“跑步、洗澡、睡觉”分支属于“运动休闲”场景下的事件演化模式，“跑步、(买)跑鞋、去网站”分支属于“运动消费”场景下的事件演化模式。这个例子揭示了事理图谱对事件演化规律刻画的准确性与多样性。图中边上还标有事件转移概率等信息。

事理图谱的理论基础与应用

理论上，事理图谱是一种概率有向图。它与概率图模型中的贝叶斯网络、马尔科夫逻辑网络既有不同又有联系。贝叶斯网络采用有向无环图来表达变量节点之间的条件依赖与独立性关系，马尔科夫随机场则采用无向图来表达变量间的相互作用关系。从这个层面上讲，事理图谱用有向有环图表达常量（事件）之间的演化关系。贝叶斯网络和马尔科夫逻辑网络的本质是研究多个随机变量的联合概率分布。而事理图谱是研究事件的链式依赖和表征事件发展方向的可能性。因此与贝叶斯网络和马尔科夫逻辑网络相比，事理图谱既有其结构特殊性，研究对象也有所不同。我们可以借鉴贝叶斯网络和马尔科夫逻辑网络中的研究成果，但研究对象的不同又对我们的理论和工具提出了新的要求。具体来说有环特性增加了事理图谱图结构的复杂性，因此传统概率图模型上的信念传播算法就面临了挑战。事理图谱这种复杂网络中的事件链条和链式依赖的挖掘与确定性评价也是一个全新的课题。

本质上，事理图谱提供了这样一种功能：给出一个抽象事件（“看电影”），它能告诉你围绕该抽象事件在时间演化顺序上先后可能发生什么样的事情（“订票”，“选座位”，“打车”，“取票”，“买饮料”，“排队入场”等）。据此，事理图谱将至少可以在以下两方面起到重要作用。

1. 智能对话系统。如果在对话中出现了 A 事件，可以在回复中提及 A 的前提事件或后继事件，构

造语义上更加智能与合理的回复。

2. 消费意图识别与推荐系统。在事理图谱中，某些事件会成为消费意图显著事件（“去旅行”，“逛街”，“爬泰山”等），能够触发一系列后续消费事件。我们把这类事件节点找出来，有助于隐式消费意图的识别，进而做出合理的商品推荐。

事理图谱的应用不仅限于此。在人工智能应用中，常识推理是一种十分重要且必需的能力。作为一种常识性事件演化逻辑知识库，事理图谱的应用潜力十分巨大。

相关工作

在调研过程中，我们发现跟事理图谱密切相关的两个研究方向是统计脚本学习和事件间时序因果关系识别。前者关注事件链条的抽取，事件预测以及事件间转移概率的建模。给定文本中两个事件，后者关注如何识别它们之间的时序、因果关系以及关系方向。借鉴这两个领域的研究成果，有助于事理图谱构建。

```
script: restaurant
roles: customer, waitress, chef, cashier
reason: to get food so as to go up in pleasure
        and down in hunger

scene 1: entering
PTRANS self into restaurant
ATTEND eyes to where empty tables are
MBUILD where to sit
PTRANS self to table
MOVE sit down

scene 2: ordering
ATRANS receive menu
MTRANS read menu
MBUILD decide what self wants
MTRANS order to waitress

scene 3: eating
ATRANS receive food
INGEST food

scene 4: exiting
MTRANS ask for check
ATRANS receive check
ATRANS tip to waitress
PTRANS self to cashier
ATRANS money to cashier
PTRANS self out of restaurant
```

图 5：饭店用餐脚本

脚本学习研究可以追溯到上世纪 70 年代。如图 5 所示，Schank & Abelson [1]提出用脚本 (Script) 来建模特定场景的事件常识知识。那时的脚本通常是学者手工编码、人工构造的。Chambers & Jurafsky [2]采用简化的事件表示形式 (Verb+dependency)，从新闻数据中自动抽取统计脚本，并提出用“挖词填空”方法来评估脚本事件预测模型。该工作成为近年来脚本学习研究的开创性工作。此后，陆续有多篇研究工作跟随 Chambers & Jurafsky 的研究框架，从事件表示、脚本事件预测模型等方面提出了很多改进方案。Pichotta & Mooney [3]首次提出用循环神经网络的方法直接对句子级事件进行操作，生成下一个句子级事件。2016 年在 ACL、EMNLP、AAAI、Coling、LREC 等会议上有多篇脚本学习的研究工作出现，并且发布了多个公开数据集，表明这项研究逐渐成为 NLP 领域一个新的研究热点。

事理图谱跟脚本学习研究虽然概念上有相似之处，但两种研究的出发点是不同的。脚本学习研究往往不关注事件的具体表示形式，经常以一个十分抽象的动词或者依存关系表示一个事件，人类一般无法理解这些事件具体含义是什么。脚本学习重点关注事件预测，给出一些已经发生的事件，预测接下来会发生什么事件。脚本学习研究目标不是构造事理知识库，并不显式地将抽取出的事件组织成有向图结构，而以链状的事件链条为事件组织形式，而且一般以事件在文中出现的顺序作为事件先后发生的顺序。事理图谱的

研究目标是将事件演化规律和模式组织成一个大型事理常识知识库，这个知识库是人类行为活动的真实刻画。

事件间时序关系识别是一个评测驱动的研究领域。该研究的标准数据集是基于 TimeML 标准标注的 TimeBank 语料。SemEval2007, SemEval2010, SemEval2013 相继开展了该任务的评测。参加者需给出文本中指定事件对之间的时序关系属于哪一种。历届评测以英文语料为主，后来又加入了西班牙语、意大利语、法语的评测。SemEval2010 发布了一个小规模中文数据集。

在 NLP 领域，已有大量工作对事件间因果关系抽取与识别进行了深入研究。这里仅列举部分有代表性的工作。Zhao 等人[5]通过提取多种类型的特征，对单个句子中的两个事件进行因果关系识别，发现因果关联词类别特征可以有效降低过拟合。Radinsky 等人[6]通过因果模板匹配的方法从新闻标题中自动抽取因果对，并将这种因果知识用于新闻事件预测。Zhao[7]采用了与[6]中类似的方法来进行因果对的抽取，并利用 WordNet 和 VerbNet 对事件进行泛化，构建了一个抽象因果网络。在此抽象因果网络上，他们进一步将事件进行 Embedding 表示，以便于事件预测、股票预测等任务。Mirza 等人[8]通过采用多分类器级联的方法，在小规模标注语料上同时进行事件时序关系识别与因果关系识别，取得了不错的实验结果。

前人在事件间时序和因果关系识别上的研究已有大量技术积累，在构建事理图谱过程中，这些技术可以为我们所借鉴。

结束语

人们每天的活动在不知不觉间进行。实际上，每个人的头脑里都有一个隐形的“事理图谱”，人们据此展开每天的行为活动。本文提出了事理图谱的概念并给出了详细定义。事理图谱是事件演化规律和模式的“知识库”，是人类行为活动的真实刻画。我们尝试从大规模无结构化数据中构建事理图谱，并且搭建了一个 Demo，让人们可以看到事理图谱的雏形。事理图谱的构建涉及多项自然语言处理核心技术，未来我们将精心打磨各个技术点，旨在构建准确、全面的常识事理知识库。这是一个非常有潜力的研究方向，也一定会吸引更多的学者投入到这个研究领域中来。

参考文献

- [1] Schank R C, Abelson R P. Scripts, plans, and knowledge[M]. New Haven, CT: Yale University, 1975.
- [2] Chambers N, Jurafsky D. Unsupervised Learning of Narrative Event Chains[C]//ACL, 2008, 94305: 789-797.
- [3] Pichotta K, Mooney R J. Using sentence-level LSTM language models for script inference[J]. ACL, 2016.
- [4] Li P, Zhu Q, Zhou G, et al. Global Inference to Chinese Temporal Relation Extraction[J]//Coling, 2016.
- [5] Zhao S, Liu T, Zhao S, et al. Event causality extraction based on connectives analysis[J]. Neurocomputing, 2016, 173: 1943-1950.
- [6] Radinsky K, Davidovich S, Markovitch S. Learning causality for news events prediction[C]//Proceedings of the 21st international conference on World Wide Web. ACM, 2012: 909-918.

[7] Zhao S, Wang Q, Massung S, et al. Constructing and embedding abstract event causality networks from text snippets[C]//WSDM. ACM, 2017: 335-344.

[8] Mirza P, Tonelli S. CATENA: CAusal and TEmporal relation extraction from NATural language texts[C]//Coling. 2016: 64-75.