From Knowledge Graph to Event Evolutionary Graph

Ting Liu
Research Center for Social Computing and Information Retrieval

2017.11.05, China·Beijing
Outline

• Motivation of Event Evolutionary Graph
• Related Work
• Our Efforts on Event Evolutionary Graph
• Conclusion
Motivation of Event Evolutionary Graph

- Most existing knowledge bases focus on “concepts and their relations”, and failed to mine “event evolutionary logics”
- Event evolutionary logics (development principles and patterns between events) are valuable commonsense knowledge, mining this kind of knowledge is crucial for understanding human behaviour and social development
What is Event Evolutionary Graph?

- Event Evolutionary Graph (EEG): 事理图谱
  - EEG is a knowledge base of event evolutionary logics, which describes the event evolutionary principles and patterns
  - 事理图谱是一个事理逻辑知识库，描述事件之间的演化规律和模式

- Structurally: EEG is a Directed Cyclic Graph, whose nodes are events, and edges stand for the sequential and causal relations (顺承和因果) between events.

- Essentially: EEG is a knowledge base of event evolutionary logics, which describes the event evolutionary principles and patterns
Applications of Event Evolutionary Graph

• EEG can be applied to several downstream tasks, including event prediction, commonsense reasoning, consumption intention mining, dialogue generation, question answering, decision making system and so on.

• Large-scale EEG can have big application potentials as traditional Knowledge Graph.
## Differences and Relations between EEG and KG

<table>
<thead>
<tr>
<th>Research Target</th>
<th>Event Evolutionary Graph</th>
<th>Knowledge Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Target</td>
<td>Predicate-Events (谓词性事件) and their relations</td>
<td>Noun-Entities (名词性实体) and their relations</td>
</tr>
<tr>
<td>Organization Form</td>
<td>Directed Graph</td>
<td>Directed Graph</td>
</tr>
<tr>
<td>Main Knowledge Form</td>
<td>Event evolutionary logics and transition probability</td>
<td>Entities’ attributes and their relations</td>
</tr>
<tr>
<td>Determinary of Knowledge</td>
<td>Most event evolutionary logics are not deterministic</td>
<td>Most relations between entities are deterministic</td>
</tr>
</tbody>
</table>
• Events in ACE: An event is a specific occurrence involving participants. An event is something that happens. An event can frequently be described as a change of state
  – Traditional event extraction and classification tasks, ACE, KBP
  – Topic detection and tracking

• Events in EEG:
  – Not specific but abstract events
  – Represent as general, semantic complete predicate-phrases or segments
  – “have hot pot”, “watch movies”, “go to the airport” are reasonable event representations
  – “go to somewhere”, “do things”, “eat” are unreasonable or incomplete events representations
The sequential relation (顺承关系) between two events refers to their partial temporal orderings.

After having lunch, Tom paid the bill and left the restaurant.

吃过午饭后，汤姆到前台买单，然后离开了餐馆。
Causal Relation between Events

- Causal relation (因果关系) is the relation between one event (the cause) and a second event (the effect), where the second event happens as a consequence of the first.
- Causal relation is a subset of sequential relation
  - Satisfy the constraint of partial temporal order

The nuclear leak in Japan led to serious ocean pollution

日本核泄漏引起了严重的海洋污染。
Chain structured EEG under the scenario of “watch movies”.
Three Topology Structures of EEG

*Chain* structured EEG under the scenario of “watch movies”.
Tree structured EEG under the scenario of “plan a wedding”.

Three Topology Structures of EEG
Tree structured EEG under the scenario of “plan a wedding”.

Three Topology Structures of EEG
Cyclic structured EEG under the scenario of “fight”.

Three Topology Structures of EEG
Cyclic structured EEG under the scenario of “fight”.
Outline

• Motivation of Event Evolutionary Graph
• Related Work
• Our Efforts on Event Evolutionary Graph
• Conclusion
Two Relevant Research Fields

• Statistical Script Learning

• Event Relation Recognition
  – Temporal relation recognition
  – Causal relation recognition
Statistical Script Learning

- A very relevant research field to EEG

Development Stage

- In 1975, American researcher Schank proposed the concept of Script
- 2003: Japanese researchers proposed automatic acquisition of script knowledge
- 2008-2013: Pioneering work
- 2014~now: Recovery and development stage

![Trend of important script learning papers](image)
Statistical Script Learning (1)


Scripts [Schank & Abelson 1975] are an influential early encoding of situation-specific world event

- 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

(From [Schank & Abelson 1975])
Statistical Script Learning (2)

- [ACL 2008] Unsupervised learning of narrative event chains, Chambers, Jurafsky, Stanford University
• [AAAI 2014] Learning scripts as Hidden Markov Models, J. Walker Orr et al, Oregon State University

Figure 1: A portion of a learned “Answer the Doorbell” script
[ACL 2016] Using sentence-level LSTM language models for script inference, Pichotta and Mooney, University of Texas at Austin

Figure 3: Different system setups for modeling the two-sentence sequence “The dog chased the cat.” followed by “The cat ran away.” The gray components inside dotted boxes are only present in some systems.
Statistical Script Learning (5)

- Inferring the correct story ending according to story contexts
  - [NAACL 2016] A corpus and cloze evaluation for deeper understanding of commonsense stories, Mostafazadeh et al.
  - [EMNLP 2017] Story Comprehension for Predicting What Happens Next, Snigdha Chaturvedi, Haoruo Peng, Dan Roth, UIUC

Context: One day Wesley’s auntie came over to visit. He was happy to see her, because he liked to play with her. When she started to give his little sister attention, he got jealous. He got angry at his auntie and bit her hand when she wasn’t looking.

Incorrect Ending: She gave him a cookie for being so nice.
Correct Ending: He was scolded.
Temporal Relation Classification

- [TACL 2014] Dense event ordering with a multi-pass Architecture, Chambers et al, United States Naval Academy

**The TimeBank**

There were four or five people inside, and they just **started firing**

Ms. Sanders was **hit** several times and was **pronounced dead** at the scene.

The other customers **fled**, and the police **said** it did not **appear** that anyone else was **injured**.

**TimeBank-Dense**

There were four or five people inside, and they just **started firing**

Ms. Sanders was **hit** several times and was **pronounced dead** at the scene.

The other customers **fled**, and the police **said** it did not **appear** that anyone else was **injured**.
Event-Causality-Driven Stock Prediction

- [WSDM 2017] Constructing and embedding abstract event causality networks from text snippets, Sendong Zhao et al., HIT-SCIR

Experiments are carried out on long-term (one month) stock price movements using 12,482 events during this period.

Focus on predicting the increase or decrease of Standard & Poor’s 500 stock (S&P) index.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ET+SVM (Ding [10])</td>
<td>53.72</td>
</tr>
<tr>
<td>EoC\text{Causal-TransE}+SVM</td>
<td>55.41</td>
</tr>
<tr>
<td>EoCDual-CET+SVM</td>
<td>\textbf{56.76}</td>
</tr>
<tr>
<td>ET+DNN (Ding [10])</td>
<td>56.08</td>
</tr>
<tr>
<td>EoC\text{Causal-TransE}+DNN</td>
<td>57.77</td>
</tr>
<tr>
<td>EoCDual-CET+DNN</td>
<td>\textbf{59.80}</td>
</tr>
</tbody>
</table>
Outline

- Motivation of Event Evolutionary Graph
- Related Work
- **Our Efforts on Event Evolutionary Graph**
- Conclusion
HIT-SCIR: Our Exploration on EEG

Construction and Application of Travel Domain EEG
Travel Domain EEG Construction

Framework for Building EEG

- Data Cleaning
- NLP Processing
- Event Extraction and Generalization
- Event Pair Candidates Generation
- Sequential Relation Recognition
- Sequential Direction Recognition
- Causality Recognition
- Transition Probability Computation

Raw Corpus → EEG
Travel Domain EEG Construction

Framework for Building EEG

1. Raw Corpus
2. Data Cleaning
3. NLP Processing
4. Event Extraction and Generalization
5. Event Pair Candidates Generation

Sequential Relation Recognition

Sequential Direction Recognition

Causality Recognition

Transition Probability Computation

EEG

Remove html tags, special symbols and so on
Travel Domain EEG Construction

Framework for Building EEG

- Data Cleaning
- NLP Processing
- Event Extraction and Generalization
- Event Pair Candidates Generation
- Sequential Relation Recognition
- Sequential Direction Recognition
- Causality Recognition
- Transition Probability Computation

Raw Corpus

Segmentation, part-of-speech tagging, and dependency parsing
Travel Domain EEG Construction

Extract verb-object phrases from the dependency-parsed tree, filter the low-frequency phrases by a proper threshold.
Every two events from two consecutive sentences are considered as an event pair candidate.
Travel Domain EEG Construction

Regard the sequential relation and direction recognition as two separate supervised binary classification tasks.
Causality is rare in our experiment corpus, so causality recognition is not covered in this paper.
Use the following equation to approximate the transition probability from event A to event B:

\[ P(B|A) = \frac{\text{count}(A, B)}{\text{count}(A)} \]
Data Sets and Experiments Results

- **Experiment corpus**
  - 320,702 question-answering pairs crawled from travel topic on “Zhihu”

- Supervised classification of sequential relation and direction

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential Relation</td>
<td>2,173</td>
<td>1,563</td>
<td>610</td>
</tr>
<tr>
<td>Sequential Direction</td>
<td>1,563</td>
<td>1,349</td>
<td>214</td>
</tr>
</tbody>
</table>

F1 value of sequential relation classification: **85.7%**
F1 value of sequential direction classification: **92.9%**
Subgraph in our automatically constructed travel domain EEG under the scenario of “buy train tickets”.

Subgraph in our automatically constructed travel domain EEG under the scenario of “buy train tickets”.
Demonstration: Travel Domain EEG

Scale of travel domain EEG:
29,825 event nodes,
234,547 directed edges.
Travel Domain EEG: Potential Applications

• Consumption Intention Mining and Recommendation
  – Most events can join with ‘want to’, ‘plan to’, ‘will’ and become intention events
    • want to go to Beijing, want to watch movies, plan to climb Mountain Tai, will have hot pot
  – Certain events have notably consumption intention, and they can lead to following consumption events; it is of great value to find these events
    • watch movies, go on a tourist
Travel Domain EEG: Potential Applications

- **Dialogue Generation**
  - go to Beijing ➔ buy tickets
    - A: I plan to go to Beijing. B: Have you buy the tickets?
  - go to Taian ➔ climb Mountain Tai
    - A: I want to climb Mountain Tai. B: Then you need to go to Taian first.

- **Question Answering System**
  - Q: Is there any dos and don’ts if I want to climb Mountain Tai?
  - A: Remember to rent a coat, take some water and a flashlight.
HIT-SCIR: Our Exploration on EEG
Construction and Application of Financial Domain EEG
Sample Analysis

- Financial news contain lots of event-event causal relations
  - Explicit causality (with connectives): “Plasticizer incident led to liquor stocks plummeted.”
    - “塑化剂事件导致白酒股大跌。”
  - Implicit causality (no connectives): “Baidu Q2 earnings report: net profit increased by 82.9%, stock price rose by 7%.”
    - “百度Q2财报：净利同比增82.9%，股价盘后上涨7%”

- Except for causal relations, it also contains large quantities of sequential relations
  - “After nearly half a month’ suspension, IFLYTEK’ stock resumed trading limit.”
    - “停牌了近半个月的科大讯飞（002230.SZ）复牌，股价开盘即涨停。”
Target for Financial Domain EEG

• Target
  – Mine economic changes related (especially stock price movements) sequential and causal event relations from financial news articles, to construct the financial domain EEG

• Method
  – Causal and sequential event relation extraction
  – Transforming ‘cause-effect pairs’ into graph
Extraction Methods for Cause-Effect Pairs

- Exploit causal triggers to construct templates and match with regular expressions, to obtain the cause and effect mentions
  - For example: (.+)(result in|lead to|cause)(.+)(drop|rise)
  - (.+)(导致|引起|造成)(.+)(下跌|上涨)
- Carry out segmentation and POS tagging on the cause and effect mentions
- Finally, the sequence of verbs, nouns, and adjectives is regarded as the cause and effect after filtering with POS tags
Representative Extracted Cause-Effect Pairs

• Profits decline of ZTE lead to its stock price dropping sharply.
  – “中兴通讯利润下滑引发股价大跌”
• Astronomical compensation rumors of ZTE resulted in its stock dropping sharply.
  – “中兴通讯天价赔偿传闻导致股票大跌”
• The news of Shuanghui’ acquisition of Smith Field made Shuanghui development stock price rise.
  – “双汇并购史密斯菲尔德消息使得双汇发展股价大涨”
• Meat products production and sales decline, and cost raises caused the business profits to drop.
  – “肉制品产、销量下降，成本上升造成肉制品业务利润下降”
• Related competitors’ entering resulted in the market share of gross profit margin declined.
  – “相关竞争对手进入导致产品毛利率市场份额下降”
• Transforming ‘cause-effect event pairs’ into graph by merging certain events.

- **A**: Profit dropped
- **B**: Stock price drops sharply
- **C**: Production and sales decline, and cost raises
- **D**: Business profits decline

Merge A and D
Merge Events by Similarity Computation

- **Event representation**
  - Bags of verbs and nouns
  - Bags of verbs, nouns and adjectives
  - Average word embedding of all words
  - Average word embedding of verbs, nouns and adjectives

- **Similarity measure**
  - Jaccard similarity
  - Cosine similarity
Case Study
Data Sources and Scale of Financial EEG

- **Data sources:** 1,362,345 financial news articles
  - 716,391 individual stocks news articles from Tencent and Netease
  - 246,499 plate news articles from Netease
  - 399,455 articles filtered from 10 years’ newspaper articles

- **Scale of financial domain EEG:**
  - 247,926 event nodes
  - 154,233 cause-effect pairs/directed edges
  - 3,111,720 similar pairs/undirected edges
Outline

• Motivation of Event Evolutionary Graph
• Related Work
• Our Efforts on Event Evolutionary Graph
• Conclusion
Conclusion

- By farming in various domains for a long time, Knowledge Graph gradually shows its great value.
  – 知识图谱在各个领域精耕细作，逐渐显露价值
- Knowledge representation still needs breakthrough, and its inference ability needs to be improved.
  – 知识表示形式有待突破，推理能力有待提高
- Statistical script learning and event relation recognition attracted more and more research attentions.
  – 统计脚本学习和事件关系识别等事理图谱相关研究越来越吸引研究者关注
Conclusion

• Event evolutionary graph, whose nodes are predicate phrases and edges are event evolutionary logics, is in the ascendant.
  – 以“谓词性短语”为节点，以事件演化（顺承、因果）为边的事理图谱方兴未艾

• Event evolutionary graph will play an important role in event prediction and dialogue generation research fields, and improve the interpretability of artificial intelligence systems.
  – 事理图谱必将在预测、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
Collaborators

Bing Qin
Professor
Big Cilin

Ming Liu
Associate Professor
Big Cilin

Xiao Ding
Assistant Researcher
EEG
Thanks!