Graph Learning Methods in Session-based Recommendations and Legal Case Retrieval

Ruihong Qiu
r.qiu@uq.edu.au
The University of Queensland
Background
Structural Information in IR

• Interaction-level structure
  • User-item interaction in recommender systems;
  • legal case reference, etc.

https://medium.com/interacta/visualizing-american-case-law-5bd0243a7097
Structural Information in IR

• Content-level structure
  • User history; legal case text structure etc.

(a)  (b)

Rethinking the item order in session-based recommendation with graph neural networks
CaseGNN: Graph Neural Networks for Legal Case Retrieval with Text-Attributed Graphs
Challenge
An Informative Graph?

• The inductive bias in a graph:
  • Node:
    • Interaction-level: user, item, document, case
    • Content-level: item, word, sentence
  • Edge (structure):
    • Interaction-level: interaction, semantics relationship
    • Content-level: time order, knowledge relation
A Powerful Graph Learning Model?

• A general graph neural networks (GNN):
  • GCN, GAT, GraphSAGE, GIN, SGC, etc.

• Specific design:
  • Suitable for interaction data, item data, text data, etc.

Example I: Modelling User Session with Graph
My Own User Session

- From bottom to top indicates from old to new.

- What’s next?
Modelling General Sequential Patterns

• General sequences: **sentence** (semantic), **time series** (real number, periodic or trended)

• General tools: recurrent model, attention model ...

https://www.scylladb.com/glossary/time-series-data/
Development of PPTNet a Neural Network for the Rapid Prototyping of Pulsed Plasma Thrusters
https://www.tensorflow.org/text/tutorials/transformer
Direct Modelling of User Sessions

- Directly using **existing** models: GRU4Rec, SASRec, BERT4Rec ...

![Diagram](image)

**Encoding**

- RNN, GRU, LSTM, Attention ...

**Matching**

- Session representations
- Ranking prediction over all items, dot product ...

[Hidasi et al. 2015, Kang et al. 2018, Sun et al. 2019]
Issues in Modelling User Sessions

• The user session (shopping or watching history) is different from general sequences.
  • No grammar semantics;
  • No periodic feature;
  • No continuous trending etc.

• Need to identify the difference and develop proper models to learn sequential patterns in user sessions.
Structural Information

• A sequence of anonymous user history within a short time period.
• Items have **reappearance** in the sequence, such as re-click products; re-listen to songs; re-watch videos.
• They are **PIVOTAL**.
Converting to Graph

- Not like 1D sequence, graph has the topological structure with nodes and edges.
- Convert a sequence into a graph.
- Pivotal items become pivots.
FGNN Model

Graph construction

Weighted Graph Attentional Layer

Set2Set readout

Matching

Sequence $S$:

$\begin{array}{c}
v_3 \\
v_7 \\
v_5 \\
v_7 \\
v_6 \\
\end{array}$

$\begin{array}{c}
x_3 \\
x_7 \\
x_5 \\
x_6 \\
x_7^L \\
x_6^L \\
\end{array}$

$\mathcal{G}_S$

$7 \times \mathbf{W}$

Readout

$q_t$

Item set $\mathcal{V}$

$\hat{z}$

$\hat{y}$

Graph constraint

[Qiu et al. CIKM 2019]
Streaming Scenario: GAG Model

- Model trained on offline data
- Needs to update with online data
Cross-sequence Scenario

- Multiple sequence could contain same items
- Link multiple sequence together

[Qiu et al. TOIS 2020]
Positional Information

• An interaction at different positions will carry different meanings for user preference.

• Within a sequence, an early interaction would indicate the initial intention; A later interaction would indicate the latest intention.
Position in General Sequence

• Attention mechanism has a **positional encoding** (PE) (counting from left to right).

• Based on **sin/cos** function. Each position corresponds to one row.

https://kazemnejad.com/blog/transformer_architecture_positional_encoding/
Problems of PE In Interaction Sequence

- Original PE only tells how far away from the beginning (forward-awareness).
- Problematic example in following sessions for $v_6$:

  - Need to know how far away from both the beginning (forward-awareness) and the ending (backward-awareness).
Dual Positional Encoding

• PE for interaction sequence needs to be both forward-aware and backward-aware

• Dual Positional Encoding (PDE):

\[
\begin{align*}
p_{pos,2i}^l &= \sin(pos/f(i)), \\
p_{pos,2i+1}^l &= \cos(pos/f(i)), \\
p_{pos,2i+d/2}^l &= \sin((l-pos-1)/f(i)), \\
p_{pos,2i+1+d/2}^l &= \cos((l-pos-1)/f(i)),
\end{align*}
\]

Theorem 4.3.1. Dual positional encoding can represent the positional information of SBRS because it is both forward-aware and backward-aware.

• Also a Learnable Dual Positional Encoding (LDPE)
PosRec Model

Session $S$

Session graph $G_s$

PGGNN

L/DPE

Transformer

$\nu_3 \ \nu_7 \ \nu_5 \ \nu_7 \ \nu_6$

$\nu_3' \ \nu_7' \ \nu_5' \ \nu_7' \ \nu_6'$

[Qiu et al. TOIS 2021] 22
Visualisation of PEs

• One sequence has 10 items, and the other one has 20.
• **Forward-awareness** for the beginning position.
• Only Dual and Learnable Dual PE have **backward-awareness**.
• Learnable one can identify **different levels of importance**.

(a) General  (b) Dual  (c) Learnable Dual
Example II: Legal Case Retrieval with Graph
Legal Case Retrieval Workflow

- **Query case**
- **Case pool**
- **Retriever**
- **Relevant cases**
Legal Cases

Lafond v. Muskeg Lake Cree Nation (2008), 330 F.T.R. 60 (FC)

Summary:
Lafond was elected as a councillor to the Muskeg Lake Cree Nation Band Council. After ...

In the recent decision of FRAGMENT_SUPPRESSED, the Federal Court of Appeal: ...

For these reasons, the application for judicial review of Chief Ledoux's decision will be allowed.

Parties: plaintiff & defendant

Case Summaries

Citation

Judgment
Related Work in Legal Case Retrieval

- Bert-based model
- BERT-PLI
- Encode paragraphs with BERT
- Paragraph-level interaction

[Shao et al. 2020] 27
Related Work in Legal Case Retrieval

- Bert-based model
- SAILER
- Generative pretraining

[Li et al. 2023]
Related Methods’ Characteristics

• **Pros**
  • Better accuracy with semantics by *legal corpus pre-training*
  • Dividing case text for *lengthy* problem

• **Cons**
  • Case text dividing -> loss of legal *context* information & case *global* view
Challenges

• Legal structural information:
  • **High-order interactions** of elements in a case: parties, crime activities and evidences

• Lengthy legal text limitation:

<table>
<thead>
<tr>
<th>Datasets</th>
<th>LeCaRD</th>
<th>COLIEE2023</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Chinese</td>
<td>English</td>
</tr>
<tr>
<td>Avg. length/case</td>
<td>8,275</td>
<td>5,566</td>
</tr>
<tr>
<td>Largest length of cases</td>
<td>99,163</td>
<td>61,965</td>
</tr>
<tr>
<td>Avg. relevant cases/query</td>
<td>10.33</td>
<td>2.69</td>
</tr>
</tbody>
</table>
Solution

• Graph data is an effective data structure to incorporate the **abundant structural information in legal cases.**

• **Transform** a legal **case** into a Text-Attributed Case **Graph** (TACG).

• An Edge Graph Attention Layer (EdgeGAT) and a readout function are proposed to obtain a graph level case representation.
TACG

• Extract the **entities** and the **relations** using Information Extraction
• Separate legal **fact** and legal **issue**
• Create text-attributed **case graph**, with a **virtual** node
CaseGNN Framework

[Tang, Qiu et al. ECIR 2024] 33
## Overall Performance

<table>
<thead>
<tr>
<th>Methods</th>
<th>COLIEE2022</th>
<th>COLIEE2023</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P@5</td>
<td>R@5</td>
</tr>
<tr>
<td><strong>One-stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM25</td>
<td>17.9</td>
<td>21.2</td>
</tr>
<tr>
<td>LEGAL-BERT</td>
<td>4.47</td>
<td>5.30</td>
</tr>
<tr>
<td>MonoT5</td>
<td>0.71</td>
<td>0.65</td>
</tr>
<tr>
<td>SAILER</td>
<td>16.6</td>
<td>15.2</td>
</tr>
<tr>
<td>PromptCase</td>
<td>17.1</td>
<td>20.3</td>
</tr>
<tr>
<td>CaseGNN (Ours)</td>
<td><strong>35.5±0.2</strong></td>
<td><strong>42.1±0.2</strong></td>
</tr>
<tr>
<td><strong>Two-stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAILER</td>
<td>23.8</td>
<td>25.7</td>
</tr>
<tr>
<td>PromptCase</td>
<td>23.5</td>
<td>25.3</td>
</tr>
<tr>
<td>CaseGNN (Ours)</td>
<td>22.9±0.1</td>
<td><strong>27.2±0.1</strong></td>
</tr>
</tbody>
</table>

- CaseGNN is better in most situations.
CaseGNN Case Study

• Successful retrieval by CaseGNN but not by PromptCase.

Original text: entities and relationships are far from each other. Language models are not good at long dependency.

TACG: brings multiple entities together.
Takeaway and Conclusions
Takeaway and Conclusions

• There are **abundant structural information** in different IR scenarios

• The **graph construction determines** how good a graph-based method can be

• The graph learning module is effective. But a **dedicated design** can further improve performance.
Thank You!

Ruihong Qiu
r.qiu@uq.edu.au