Diversified Query Generation Guided by Knowledge Graph

Xinyao Shen, Jiangjie Chen, Jiaze Chen, Chun Zeng, Yanghua Xiao
Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University
ByteDance AI Lab
Fudan-Aishu Cognitive Intelligence Joint Research Center
Outline

• Query Generation
• Knowledge-Enhanced Diversified Query Generator (KEDY)
• Experiments
• Summary
Query Generation

- Input: article and title
- Output: queries
- Query
  - Related
  - Fluency
  - Diversity
  - Popularity

<table>
<thead>
<tr>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appearance</strong> comparison photos of <strong>Hollywood</strong> stars</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the past ten years, movie companies headed by Marvel and DC have almost maintained a tempo of 2-4 movies a year. They have also brought us such box office and good word-of-mouth double-harvest works as “Iron Man”, “Avengers”, “Aquaman” and “Spider-Man&quot; further sweeping the American comics super hero craze to every corner of the world. In today’s issue, I will bring you the appearance comparison photos of the actors starring in the American comics super hero movies so that you can understand the connotation of talent excellence. The appearance of Hollywood stars changes such as Jason Momoa who starred in “Aquaman”. Jason’s sturdy figure does not need to be said, and his performance in “Aquaman&quot; has really shone the audience...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) European and American star appearance</td>
</tr>
<tr>
<td>(2) Hollywood star appearance ranking</td>
</tr>
<tr>
<td>(3) Top 10 Marvel beauties</td>
</tr>
<tr>
<td>(4) Spider-Man actor Tom Holland’s new romance</td>
</tr>
</tbody>
</table>
Query Generation

• Click-through data
  – Long-tail
  – Diversity

• Seq2Seq model

• Graph model
  – Long article

• KG Enhanced Model
  – Spider-Man actor Tom Holland’s new romance
In the past ten years, film companies headed by Marvel and DC have released a series of hit movies. The collaboration between these two companies has been crucial in expanding the scope of superhero stories. Marvel, through its iconic characters like Spider-Man and Captain America, has been particularly influential. Spider-Man actor Tom Holland’s new romance has added to the buzz surrounding the Marvel Cinematic Universe.

Graph Construction

Knowledge Graph

Central Graph $G_C$

Multi-hop Graph $G_M$

Graph Representation Learning

GCN

Article Graph

Central Graph $G_C$

Multi-hop Graph $G_M$

GraphNet

Popularity-guided Graph Attention

Article Graph Representations

Central Graph Representations

Multi-hop Graph Representations

Diversified Generation

Control Gate $\mu^*$

$\{0,1,2\}$

Vocab

Central

Multi-hop

Output

Output Queries
1. European and American star appearance
2. Hollywood star appearance ranking
3. Top 10 Marvel Beauties
4. Spider-Man actor Tom Holland’s new romance...

In this diagram, we see the integration of various graph representations through GCN and GraphNet. The output queries illustrate the types of queries that can be generated using the output queries.
Graph Construction

- Article Graph
  - Node: keyword + sentences
- Central Graph
- Multi-hop Graph

**Algorithm 1: Construct Entity Interaction Graph**

**Input:** Title \( T \) and Article \( A \)

**Output:** Entity Interaction Graph

1. Do word segmentation of Title \( T \) and Article \( A \);
2. Do Named Entity Recognition (NER) and keywords extraction algorithm of Article \( A \) and get the entity set \( E \);
3. **while not at end of this article do**
   4. read current sentence \( s \);
   5. **if** contains \( e \in E \) **then**
      6. Add \( s \) to node \( n_e \);
   7. **else**
      8. Add \( s \) to node \( n_{empty} \);
   9. **end**
10. **end**
11. Assign Title \( T \) as node \( n_t \);
12. **for** node \( n_i \) and \( n_j \) **do**
    13. Edge Weight \( w_{ij} = \) number of shared sentences of \( n_i \) and \( n_j \)
14. **end**
Graph Representation Learning

- Article Graph Encoding
- Knowledge Sub-graph Encoding
  - Central Graph Encoding
  - Popularity-guided Graph Attention

\[
\eta_{eq}^r = \sigma(P^T \cdot \tanh(W_p \cdot e_p + W_q \cdot e_q))
\]

\[
P = W_r \cdot \lambda_{eq} \cdot r
\]

\[
\lambda_{eq} = \frac{e_{vt} - s_k}{n \cdot \tau} + \frac{k}{n}
\]

\[
p_{eq} = \sum_{e_q} \eta_{eq}^r \cdot [e_p \circ e_q]
\]
Diversified Generation

• Context Representation
  – Attention

\[
\begin{align*}
  c_{t-1}^A &= \sum_{i=1}^{n} \alpha_{t-1}^i \cdot h_i \\
  c_{t-1}^C &= \sum_{e_i \in C} \beta_{t-1}^{e_i} \cdot g_{e_i} \\
  c_{t-1}^M &= \sum_{e_p \in M \cap V^i} \gamma_{t-1}^{p} \cdot p_{e_p}
\end{align*}
\]

• Diversified Token Generation
  – Control Gate

\[
q_t = \begin{cases} 
  \sigma(s_t \cdot w), & \mu^* = 0 \\
  \sigma(s_t \cdot g_{e_i}), & \mu^* = 1 \\
  \sigma(s_t \cdot e_q), & \mu^* = 2
\end{cases}
\]
## Experiments

<table>
<thead>
<tr>
<th>Model</th>
<th>Entertainment</th>
<th></th>
<th></th>
<th>Sport</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-1</td>
<td>B-2</td>
<td>B-4</td>
<td>R-1</td>
<td>R-L</td>
<td>B-1</td>
</tr>
<tr>
<td>TextRank [22]</td>
<td>22.6</td>
<td>11.6</td>
<td>0.9</td>
<td>26.3</td>
<td>22.5</td>
<td>22.5</td>
</tr>
<tr>
<td>PtrGen [21]</td>
<td>49.5</td>
<td>38.4</td>
<td>18.1</td>
<td>42.5</td>
<td>41.3</td>
<td>49.4</td>
</tr>
<tr>
<td>Transformer [28]</td>
<td>50.6</td>
<td>39.7</td>
<td>19.0</td>
<td>44.3</td>
<td>42.9</td>
<td>50.7</td>
</tr>
<tr>
<td>Transformer+KG [28]</td>
<td>50.9</td>
<td>39.9</td>
<td>19.2</td>
<td>44.3</td>
<td>43.0</td>
<td>50.8</td>
</tr>
<tr>
<td>CVAE [39]</td>
<td>50.7</td>
<td>39.7</td>
<td>19.1</td>
<td>44.1</td>
<td>42.9</td>
<td>50.7</td>
</tr>
<tr>
<td>DP-GAN [34]</td>
<td>51.0</td>
<td>39.9</td>
<td>19.0</td>
<td>44.2</td>
<td>42.9</td>
<td>50.9</td>
</tr>
<tr>
<td>BART [13]</td>
<td>51.7</td>
<td>40.6</td>
<td>20.8</td>
<td>46.5</td>
<td>44.2</td>
<td>51.7</td>
</tr>
<tr>
<td>BART+KG [13]</td>
<td>52.2</td>
<td>40.9</td>
<td>21.0</td>
<td>46.8</td>
<td>44.5</td>
<td>52.1</td>
</tr>
<tr>
<td>M-CNTRL [35]</td>
<td>52.7</td>
<td>41.2</td>
<td>20.9</td>
<td>47.1</td>
<td>44.8</td>
<td>52.9</td>
</tr>
<tr>
<td>Graph2Seq [17]</td>
<td>52.8</td>
<td>41.2</td>
<td>20.9</td>
<td>47.2</td>
<td>45.7</td>
<td>52.7</td>
</tr>
<tr>
<td>G-S2A [6]</td>
<td>53.1</td>
<td>41.3</td>
<td>20.5</td>
<td>47.5</td>
<td>46.1</td>
<td>53.0</td>
</tr>
<tr>
<td>G-S2A+KG [6]</td>
<td>53.8</td>
<td>41.6</td>
<td>20.8</td>
<td>47.8</td>
<td>46.2</td>
<td>53.6</td>
</tr>
<tr>
<td><strong>KEDY (Ours)</strong></td>
<td><strong>56.9</strong></td>
<td><strong>44.7</strong></td>
<td><strong>23.9</strong></td>
<td><strong>50.2</strong></td>
<td><strong>48.6</strong></td>
<td><strong>56.6</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Entertainment</th>
<th></th>
<th></th>
<th>Sport</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cor</td>
<td>Div</td>
<td>Info</td>
<td>Flu</td>
<td>Nov</td>
<td>Avg</td>
</tr>
<tr>
<td>PterGen [31]</td>
<td>4.76</td>
<td>2.63</td>
<td>3.82</td>
<td>4.16</td>
<td>3.65</td>
<td>3.80</td>
</tr>
<tr>
<td>Transformer+KG [28]</td>
<td><strong>4.83</strong></td>
<td>2.65</td>
<td>3.87</td>
<td>4.04</td>
<td>3.71</td>
<td>3.83</td>
</tr>
<tr>
<td>CVAE [39]</td>
<td>4.75</td>
<td>3.02</td>
<td>3.95</td>
<td>4.08</td>
<td>3.73</td>
<td>3.91</td>
</tr>
<tr>
<td>DP-GAN [34]</td>
<td>4.76</td>
<td>3.01</td>
<td>3.92</td>
<td>4.11</td>
<td>3.71</td>
<td>3.90</td>
</tr>
<tr>
<td>Graph2Seq [17]</td>
<td>4.81</td>
<td>3.01</td>
<td>4.12</td>
<td>4.60</td>
<td>3.80</td>
<td>4.07</td>
</tr>
<tr>
<td><strong>KEDY (Ours)</strong></td>
<td><strong>4.82</strong></td>
<td><strong>4.03</strong></td>
<td><strong>4.31</strong></td>
<td><strong>4.65</strong></td>
<td><strong>4.08</strong></td>
<td><strong>4.35</strong></td>
</tr>
</tbody>
</table>
Experiments

- Diversity Evaluation
  - Correlation between Popularity and Query Clicks
  - Effectiveness of Popularity Knowledge Incorporation

<table>
<thead>
<tr>
<th>Model</th>
<th>Self-BLEU-2</th>
<th>Dist-1</th>
<th>Dist-2</th>
<th>Ent-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer+KG</td>
<td>35.8</td>
<td>0.027</td>
<td>0.125</td>
<td>6.26</td>
</tr>
<tr>
<td>M-CNTRL</td>
<td>28.2</td>
<td>0.056</td>
<td>0.312</td>
<td>7.52</td>
</tr>
<tr>
<td>G-S2A+KG</td>
<td>27.5</td>
<td>0.067</td>
<td>0.321</td>
<td>7.23</td>
</tr>
<tr>
<td>KEDY</td>
<td>21.7</td>
<td>0.186</td>
<td>0.521</td>
<td>8.68</td>
</tr>
</tbody>
</table>

(a) Real data.  (b) Generated data.

(a) The popularity score.  (b) “Unique words”.
Summary

- **Diversity**: knowledge graphs
- **Popular**: popularity-guided graph attention
- **Future work**
  - User preferences
  - Semantic feature
Thank you