Extractive Headline Generation Based on Learning to Rank on Community Question Answering

Tatsuru Higurashi*¹ Hayato Kobayashi*¹ *² Takeshi Masuyama*¹ Kazuma Murao*¹

*¹Yahoo Japan Corporation *² RIKEN AIP
Contents

• Motivation and Approach

• Proposed method

• Experiment and Results

• Conclusion
Contents

• Motivation and Approach

• Proposed method

• Experiment and Results

• Conclusion
Advice please.
A dog kept in the next house barks from morning to night.
...
How can I effectively manage this problem?

Answer of User A
No solution other than moving.

Answer of User B
Please contact the public health center.
はじめまして、よろしくお願いします、30代の男性です。アドバイスお願いします。
隣の家が飼っている犬が朝から晩まで吠え続けます。近所の人たちも飼い主に注意を促しているのですが、改善が見られません。
Snippet Extraction Makes Headline Informative

- **Approach**
  Extract a mid-substring as a snippet

**Posted Question**
Nice to meet you, thank you in advance. I'm a man in my thirties. Advice please.

A dog kept in the next house barks from morning to night. Neighbors have given the owner cautions against it.

**Snippet headline:** Informative

...A dog kept in the next house barks from morning...

**Prefix Headline:** Uninformative

Nice to meet you, thank you in advance. I'm a...
Contributions

1. Show empirical evidence that snippet headlines are more effective than prefix headlines

2. Propose extractive headline generation method based on learning to rank

3. Create Japanese dataset including headline candidates with "headline-ness" scores by crowdsourcing
Advantages of Snippet Headlines

• **Effective**

  ![Average Answer Rate](chart)

  - Prefix Headline: 0.31%
  - Snippet Headline: 0.75%
  - Increase: ×2.4

  A/B testing on Yahoo! Chiebukuro push notifications of smartphones

• **Practical**

  Snippet headlines never include generative errors

  • Commercial services (Yahoo! Chiebukuro) cannot accept generative errors
  • Abstractive methods cannot completely avoid generative errors
Our research is first attempt to address extractive headline generation for CQA service with substring of question based on learning to rank.
Contents

• Motivation and Approach

• Proposed Method

• Experiment and Results

• Conclusion
Posted Question
Nice to meet you, thank you in advance. I'm a man in my thirties. Advice please.
A dog kept in the next house barks from morning to night...

Candidate Generation
(candidate1, candidate2, candidate3, candidate4)

Candidate Ranking
(candidate1, candidate2, candidate3, candidate4)

Ranked Headline
...A dog kept in the next house barks from morning to night...
Nice to meet you, thank you in advance, I'm a...

... Advice please. A dog kept in the next house ...

... A dog kept in the next house barks from morning ...

- Make sentence which starts from beginning of each sentences of question.
- Cut subsequent sentences if it has over 20 Japanese characters.
- Put ellipsis at front and end of substring.
Candidate Ranking

- Pairwise Learning to Rank
  - L2-regularized L2-loss linear rankSVM

\[
\min_\omega \frac{1}{2} \omega^\top \omega + C \sum_{(i,j) \in P} \ell \left( \omega^\top \tilde{x}_i - \omega^\top \tilde{x}_j \right) \quad (\ell(d) = \max(0,1 - d)^2)
\]

\[
P = \{(i,j)|q_i = q_j, y_i > y_j, (x_i, y_i, q_i) \in D, (x_j, y_j, q_j) \in D \}
\]

- x: Candidate headline
- y: Vote score
- q: Question

[Lee 2014] Lee, Ching-Pei and Lin, Chih-Jen: Large-scale Linear RankSVM. Neural Computation, 26(4)
**Instruction for workers**
Select the best option from the list so that users can guess the content of the question and distinguish it from other ones.

<table>
<thead>
<tr>
<th>Randomly Sorted Headline Candidate</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  ... Neighbors have given the owner cautions against ...</td>
<td>2</td>
</tr>
<tr>
<td>2  ... This area has only private houses, not rented ...</td>
<td>0</td>
</tr>
<tr>
<td>3  ... How can I effectively manage this problem?</td>
<td>1</td>
</tr>
<tr>
<td>4  ... However, I will go crazy if I have to keep enduring ...</td>
<td>0</td>
</tr>
<tr>
<td>5  ... A dog kept in the next house barks from morning ...</td>
<td>6</td>
</tr>
<tr>
<td>6  Nice to meet you, thank you in advance, I'm a ...</td>
<td>0</td>
</tr>
<tr>
<td>7  ... Advice please. A dog kept in the next house ...</td>
<td>1</td>
</tr>
</tbody>
</table>

Number of votes by 10 workers per question
Crowdsourcing Results

- Ratio of questions whose prefix headlines were most voted candidates was **61.8%**
  
  = Room for improvement for prefix headline was up to **38.2%**.

Our Goal

- Improve uninformative headlines of 38.2%
- Remaining 61.8% unchanged
Contents

• Motivation and Approach

• Proposed Method

• Experiment and Results

• Conclusion
Features for Ranking Model

• Features
  • Bag-of-Words: 30,820 dimension sparse vector based on tf-idf
  • Embedding: 100 dimension dense vector based on doc2vec
  • Position: 10 dimension binary vector representing candidate position
Compared Methods

- **Simple Baselines**
  - *Prefix*: Select first candidate
  - *DictDel*: Delete uninformative sentence with rule (Used in A/B testing)
  - *Random*: Select randomly

- **Unsupervised Baselines**
  - *ImpTfidf*: Select most important candidate with highest tf-idf value
  - *SimTfidf*: Select most similar candidate to original question with cosine similarity
  - *LexRank*: Select candidate with highest score based on LexRank (Erkan&Radev 2004)

- **Supervised Baselines**
  - *SVM*: Select candidate with highest confidence learned as classification task
  - *SVR*: Select candidate with highest predicted votes learned as regression task
Evaluation Metrics

• **Average Votes**
  - Measures how appropriate candidates selected by each method
  - Determines the overall performance of each method
  - Corresponds to DCG@1
  
  \[
  \text{Average votes} = \frac{\text{Sum of votes for the best candidates for all questions}}{\text{No. of questions}}.
  \]

• **Change Rate**
  - Measures how much each method changed the default prefix headline
  - Determines the impact of application to actual CQA service.
  
  \[
  \text{Change rate} = \frac{\text{No. of questions where the best candidate is not the prefix headline}}{\text{No. of all questions}}.
  \]
## Results: Quantitative Analysis

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Votes</th>
<th>Change Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal</td>
<td>5.56</td>
<td>38.2</td>
</tr>
<tr>
<td><strong>MLRank(ours)</strong></td>
<td><strong>4.28</strong></td>
<td><strong>9.9</strong></td>
</tr>
<tr>
<td>Prefix</td>
<td>4.19</td>
<td>0</td>
</tr>
<tr>
<td>DictDel</td>
<td>4.23</td>
<td>2.2</td>
</tr>
<tr>
<td>Random</td>
<td>1.39</td>
<td>85.9</td>
</tr>
<tr>
<td>ImpTfidf</td>
<td>1.68</td>
<td>81.1</td>
</tr>
<tr>
<td>SimTfidf</td>
<td>2.27</td>
<td>79.3</td>
</tr>
<tr>
<td>LexRank</td>
<td>2.27</td>
<td>55.7</td>
</tr>
<tr>
<td>SVM</td>
<td>4.09</td>
<td>16.7</td>
</tr>
<tr>
<td>SVR</td>
<td>3.00</td>
<td>52.5</td>
</tr>
</tbody>
</table>

1. **MLRank(ours)** performed the best among all methods.
2. **Prefix** (First sentence) can be a good summary.
3. **DictDel** (Rule-Based) was more useful than Prefix.
4. Change rate of **DictDel** was small, which means small impact on service.
5. Change rates of unsupervised methods were high, but the overall performances were low.
### Results: Qualitative Analysis

#### Examples of prefix headline and snippet headline

<table>
<thead>
<tr>
<th>Prefix Headline</th>
<th>Snippet Headline</th>
</tr>
</thead>
<tbody>
<tr>
<td>カテ違いならゴメンナサイ。今、財布が破れています。</td>
<td>…今、財布が破れてツライ状況です。新しから抜けて…</td>
</tr>
<tr>
<td>I am sorry if the category is wrong. Now, my wallet is torn…</td>
<td>…Now, my wallet is torn, and I’m having a hard time. A new one ...</td>
</tr>
<tr>
<td>27歳女です。環境的になかなか新しい出会いがなく…</td>
<td>…環境的になかなか新しい出会いがなく…</td>
</tr>
<tr>
<td>I am a 27-year-old woman. Owing to my environment, there is little chance of new ...</td>
<td>…Owing to my environment, there is little chance of new encounters with men ...</td>
</tr>
</tbody>
</table>

Uninformative expressions are successfully excluded, and informative expressions are added.
Contents

• Motivation and Approach

• Proposed Method

• Experiment and Results

• Conclusion
Conclusion

Addressed a snippet headline generation task for push notifications of CQA

• Showed empirical evidence that snippet headlines are more effective than prefix headlines 2.4 times in average answer rate
• Proposed extractive headline generation method based on learning to rank
• Created dataset including headline candidates with "headline-ness" scores by crowdsourcing

Future Work

• Investigate effectiveness in practical situations on web service.
• Make the dataset publicly available.
Thank you