CQAVis: Visual Text Analytics for Community Question Answering

Enamul Hoque‡, Shafiq Joty†, Lluís Màrquez† and Giuseppe Carenini‡

†ALT group, Qatar Computing Research Institute, HBKU
‡Department of Computer Science, University of British Columbia
Community Question Answering

- Means for communities to share information and to collectively satisfy their information needs.

- Forums organize content in the form of question–comment threads

- A question posed by a user may be answered by a possibly long list of comments from other users.
Community Question Answering: Example

Which is the best bank in Qatar?

Finding answers to a new question is challenging:

- **Similar questions** are repeatedly posted
- **Long list** of comments
- Often **redundant and noisy** content
Our Approach

1) Combine **Search** and **NLP** techniques to:
   • Find **related question** threads in the forum
   • Rank the **comments** within each question thread

2) Introduce a novel **interactive visual interface** to:
   • Navigate through the **comments** to satisfy the **information needs**
Outline

- **Analyze** user requirements in CQA forums
  identify use cases, tasks and design needs

- **The system**
  - Combine **NLP** and **search** techniques
  - Design CQAVis using **user-centered** methodologies

- **Evaluate** CQAVis:
  - Online user study among hundreds of users in an **ecologically valid way**
  - **Lessons** learned from the study
User Requirements Analysis

- Analyzing existing HCI literature
  - What types of questions are asked?
    - Many questions are subjective in nature
  - Who answers and why?
    - Slower responses
  - What is the quality of answers?
    - Lots of variability
- Interviewing Qatar Living admins and users
  - Many naive users
  - Users want to find useful answers quickly
  - Difficulty in finding good answers
  - Screen space constraints

Summary of requirements:
- We need a comment classification technique to find good answers
- Interface should support users in identifying good answers quickly
- Interface should be simple and intuitive.
Outline

- **Analyze** User requirements in CQA forums
  identify use cases, tasks, and design needs

- **The system**
  - Combine **NLP** and **search** techniques
  - Design CQAVIs using **user-centered** methodologies

- **Evaluate** CQAVIs:
  - Online user study among hundreds of users in an **ecologically valid way**
  - **Lessons** learned from the study
System Overview

(a) Offline processing

Input question → Search engine module → Ranked question threads → Answer ranker module → Question threads with ranked list of answers

(b) Online processing

QatarLiving website → cQA Forum datasets → Pre-processing → Trained models → Presentation module
NLP Analysis: The Comment Classifier

• **The task:** Given a thread question and a list of comments associated with it, the task of the comment classifier is to assign a relevance score to each of the comments according to their goodness at answering the thread question.

• **SVM Classifier:**
  - Linear kernels over numeric and embedding features
  - Tree kernel over shallow syntactic trees of question and comment
Embedding Features

The sentence matrices contain the word embeddings along with other additional similar-features are passed through a hidden layer and on the convoluted feature maps, respectively, then combined to produce a similarity value using answer embedding (AE). These embeddings are then used to produce the question embedding (QE) and the kernel function is defined as a linear combination of numeric features, a linear kernel over embedding features, and four functions: a linear kernel over numeric features, a tree kernel over shallow syntactic trees, the partial tree kernel (Moschitti, 2006) for models, a linear kernel over numeric features, and a tree kernel over shallow syntactic trees. In addition to the numeric and embedding features, we also use tree kernels to measure the syntactic similarity between the question and the comment. In particular, we construct shallow syntactic trees for the question and for the answer, concatenate QE and AE, and use the resulting vector as features in our SVM model.

Figure 2: Question and answer embeddings

Table 1: Performance comparison of our classifier learned with a convolutional neural network.

<table>
<thead>
<tr>
<th>A</th>
<th>MAP</th>
<th>AvgRec</th>
<th>MRR</th>
<th>F1</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>59.53</td>
<td>72.60</td>
<td>67.83</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>73.54</td>
<td>84.61</td>
<td>81.54</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Our</td>
<td>77.66</td>
<td>88.05</td>
<td>84.93</td>
<td>66.16</td>
<td>75.54</td>
</tr>
<tr>
<td>Top</td>
<td>79.19</td>
<td>88.82</td>
<td>86.42</td>
<td>64.36</td>
<td>75.11</td>
</tr>
</tbody>
</table>

SemEval 2016 Dataset

- **Train:**
  - 2,669 questions
  - 17,900 comments
- **Test:**
  - 700 questions
  - 7,000 comments

SemEval 2016 Results
Outline

- **Analyze** User requirements in CQA forums
  identify use cases, tasks, and design needs

- **The system**
  - Combine **NLP** and **search** techniques
  - Design CQAVis using user-centered methodologies

- **Evaluate** CQAVis:
  - Online user study among hundreds of users in an **ecologically valid way**
  - **Lessons** learned from the study
Design of CQAVIs

- Guided by previous work on visualizing multiple conversations.
- Simplified and tailored to CQA data and tasks.

Enamul Hoque and Giuseppe Carenini (IUI 2016).
Designing CQAVIs: What Information Should be Presented?

**Related question**
- **relatedness score** w.r.t. new question [0,1]

**Comment**
- **classifier score** normalized to a value between [0,1]

**Top answers**
- combine **relatedness** and **classifier** scores

**Qatar Living forum dump (from March 2016):**
- Total conversations: 202,304
- Total comments: 2,043,022
- Avg. comments per conversations: **10.21**
Design of CQAVis: Iterative Prototyping

V1 (based on training corpus)

V2 (first live demo)

V3 (Qatar Living beta version)

Qatarliving.com/betasearch
Final Design of CQAVis
Live Demo
Outline

- **Analyze** User requirements in CQA forums
  
  identify use cases, tasks, and design needs

- **The system**
  
  - Combine **NLP** and **search** techniques
  
  - Design CQAVis using **user-centered** methodologies

- **Evaluate** CQAVis:
  
  - Online user study among hundreds of users in an **ecologically valid way**
  
  - **Lessons** learned from the study
User Evaluation of CQAVIs

Online User Study

Real users, real data and real tasks
- To enhance ecological validity
- Uncontrolled, less incentives for users

Pilot (lab) study
- To revise the study

http://www.QatarLiving.com/betasearch
Online User Study: Results Analysis

• **User statistics:**
  • 768 participants

• **Data Collection:**
  • Query log
  • Interface actions
  • Questionnaires (measured on Likert scale)
  • Open-ended comments
User Evaluation of CQAVis: Questionnaires

• Subjective ratings
  From 56 users

Thanks for using our new tool for querying QatarLiving! Your feedback would be greatly appreciated.

Please rate how strongly you agree or disagree with each of the following statements with respect to the tool you have used (more stars means higher agreement).

I found this tool to be useful ❤❤❤❤❤
I found this tool easy to use ❤❤❤❤❤
I found this tool enjoyable to use ❤❤❤❤❤
This tool enabled me to find answers relevant to my questions. ❤❤❤❤❤

Would you prefer this tool over the regular one? *
  ○ Yes
  ○ No
  ○ Can’t tell

Would you prefer this tool over the regular one? *
Yes: 75%, No: 6.25, Can’t tell: 18.75%
User Evaluation of CQAVis: Interaction Statistics

- **Interactions in Conversation view**
  - Filter Comments: 9.01%
  - Hover Comment: 74.48%
  - Sort Comments: 6.51%
  - Click Thread Overview: 39.40%
  - Hover Thread Overview: 54.01%
  - Click Conversation: 87.21%
  - Sort Conversation: 11.98%
  - Hover Conversation: 97.40%
  - Search: 97.98%

- **Interactions in Question list view**

Time spent: 155 sec (median), 2483 sec (average)
Lessons Learned

Design
• Less is more
• Enhance learnability
• Introduce familiar visualizations

Evaluation
• Challenges in collecting feedback from users
Conclusions

• We successfully integrate **NLP** and **InfoVis** for CQA
  • **NLP**: retrieve and rank comments given a new questions
  • **InfoVis**: helps the user in rapidly navigating the comments

• Used by a **population** with possibly **low visualization literacy**

• **Lessons** learned from the study

• **Future work:**
  • Further improve the **comment classifier**
  • Apply CQAVis in other forum conversations
Thank you!

Check out the live demo: iyas.qcri.org
Supplementary slides
User requirements analysis

• Analyzing Existing HCI literature
  • What types of questions are asked?
    • Many questions are subjective in nature
  • Who answers and why?
    • Slower responses
  • What is the quality of answers?
    • Lots of variability
• Interviewing Qatar Living admins and users
  • Many naive users
  • Users want to find useful answers quickly
  • Difficulty in finding good answers
  • Screen space constraints

• We need a comment classification technique to find good answers
• Interface should be simple and intuitive.
• Interface should support users in identifying good answers quickly
User Evaluation of CQAVis

• Some qualitative feedback?
• About limitations?
Literature Review

• Conversation visualization in general
  • Convis and others

• Specific for CQA domain
Future Work

• Further improve the comment classifier
• Apply CQAVis in other forum conversations