

Question Answering Using Structured and Semi-Structured User Generated Content

Doctoral thesis proposal

Denis Savenkov

Dept. of Math & Computer Science

Emory University

denis.savenkov@emory.edu

December, 2015

Abstract

Modern question answering (QA) systems employ a variety of different unstructured (text corpora), semi-structured (question answer pairs, infoboxes) and structured (knowledge bases - KB) data sources for candidate generation. However, typically each of the data sources is processed independently and the merging of the results happens only after the candidates are generated. I propose to enrich the expressiveness of data representation of each of the data sources with complementary information for joint reasoning. Text documents can be annotated with the information about properties and connections between mentioned entities. On the other hand for knowledge base question answering systems passages and text fragments retrieved for a candidate answer can help the system to find a good mapping between a natural language question and the corresponding structured query representation.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research Questions	1
1.3	Research Plan	1
1.3.1	Understanding Answer Contributions in CQA (Chapter 3)	1
1.3.2	Improving Search Quality with CQA Data (Chapter 5)	1
1.3.3	Research Timeline	1
1.4	Contributions and Implications	1
2	Related Work	2
2.1	Related Work on the CQA Side	2
2.1.1	Blah-Blah	2
3	Understanding Answer Contributions in CQA	3
3.1	Modeling Answerer Behavior in CQA	3
3.2	Summary	3
4	Improving Query Understanding with CQA Data	4
4.1	Understanding When Searchers Become Askers	4
4.2	Summary	4
5	Improving Search Quality with CQA Data	5
5.1	Predicting Searcher Satisfaction with Existing Answers	5
5.2	Summary	5

1 Introduction

1.1 Motivation

1.2 Research Questions

1.3 Research Plan

1.3.1 Understanding Answer Contributions in CQA (Chapter 3)

1.3.2 Improving Search Quality with CQA Data (Chapter 5)

1.3.3 Research Timeline

A tentative timeline for the work that needs to be done is shown below:

- Completing the work proposed in Section ?? (10/2013 - 11/2013): For detecting implicit question intent in web search, build a better evaluation set, develop more features for the classifier and compare with more baselines.
- Completing the work proposed in Section 1.3.2 (12/2013 - 04/2014): For improving question search for queries with question intent, develop models that build better query model and learn better word-to-word translation relationships based on the query-question click data and question category information.
- **If time permits**, extending the work proposed in Section 1.3.2 (03/2014 - 04/2014): Extend the methods developed above for question search to achieve better query expansion and results ranking for general web search.
- Thesis writing (05/2014 - 08/2014)
- Thesis defense (08/2014)

1.4 Contributions and Implications

2 Related Work

2.1 Related Work on the CQA Side

2.1.1 Blah-Blah

3 Understanding Answer Contributions in CQA

In this chapter, we summarize the proposed work and preliminary findings in understanding answer contributions in CQA systems (Research Question 1). In particular, we first focus on exploring the contextual factors that influence the answerer behavior in a large CQA system, and then study the effects of the information context of the answerer at the time a question is received on the answerers' reported ability, effort, and willingness to answer questions. The goal is to inform the construction of question routing and recommendation systems.

3.1 Modeling Answerer Behavior in CQA

3.2 Summary

4 Improving Query Understanding with CQA Data

In this chapter, we summarize the proposed work and preliminary findings in improving query understanding with CQA data (Research Question 2). In particular, we first focus on studying the types of web searches or information needs that lead to transition from searching to asking, and then try to identify implicit question intent in web search. The goal is to understand what types of queries would benefit most from CQA services and archives.

4.1 Understanding When Searchers Become Askers

4.2 Summary

5 Improving Search Quality with CQA Data

In this chapter, we summarize the proposed work and preliminary findings in improving search quality with CQA data (Research Question 3). In particular, we first focus on predicting searcher satisfaction with existing answers, and then try to improve retrieval of relevant questions for question-intent queries. The goal is to finally improve searcher satisfaction with CQA answers.

5.1 Predicting Searcher Satisfaction with Existing Answers

5.2 Summary