Community Question Answering Improvement Techniques

Zhijun Zhang, Weikeng Chen and Xinyu Wang

Supervised by

Prof. Wilfred NG

Introduction

Community-based Question Answering (CQA) systems are becoming more and more popular today. They provide a new way of asking and answering questions for specific topics.

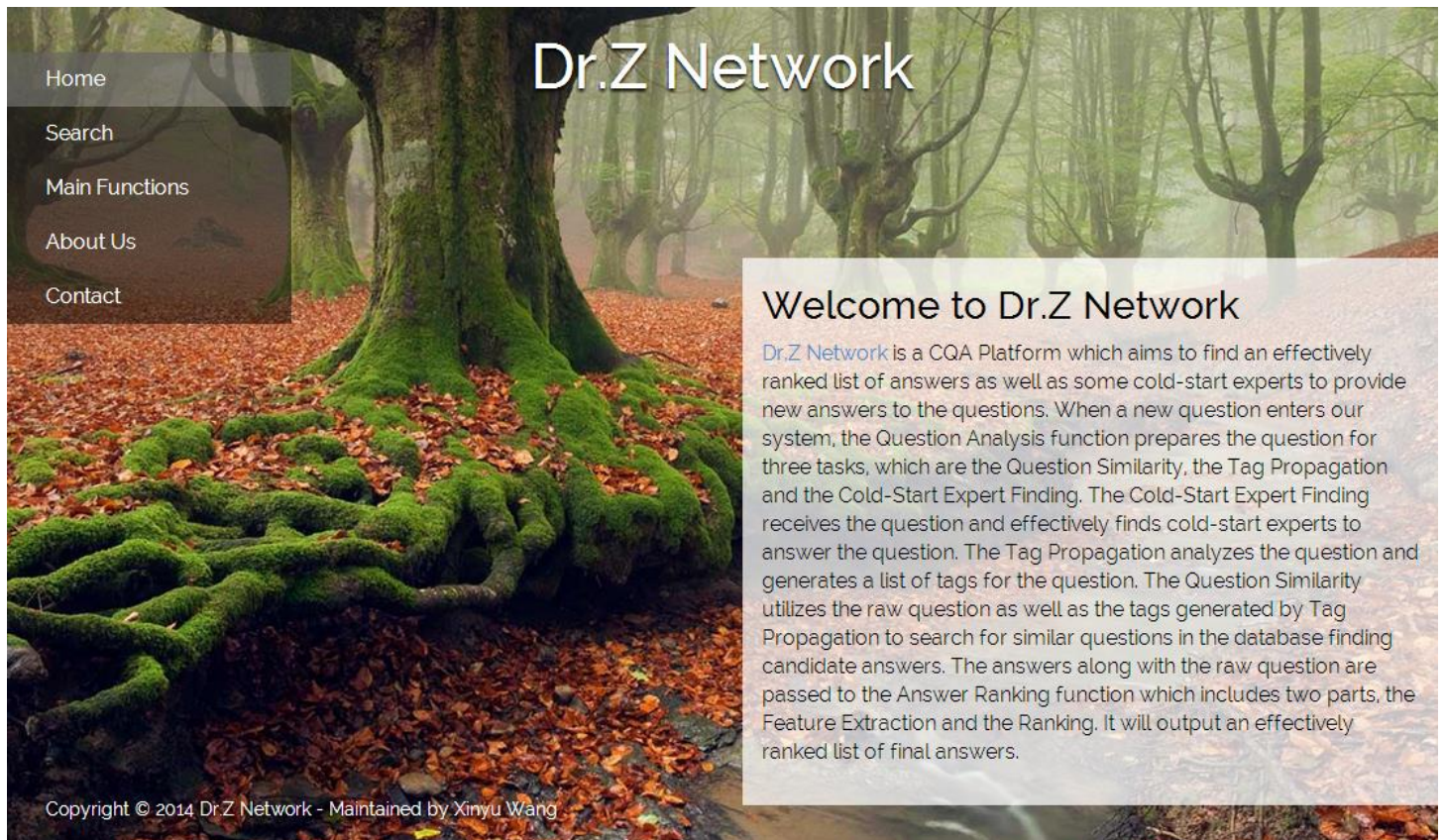There are three core functions of a CQA system. For a question posted in the system:

1. To acquire a set of relevant answers
2. To rank the answers effectively
3. To find experts to provide answers

Objectives

In this project, we developed a CQA system. To improve the performance of the three core functions in our system, we applied three techniques:
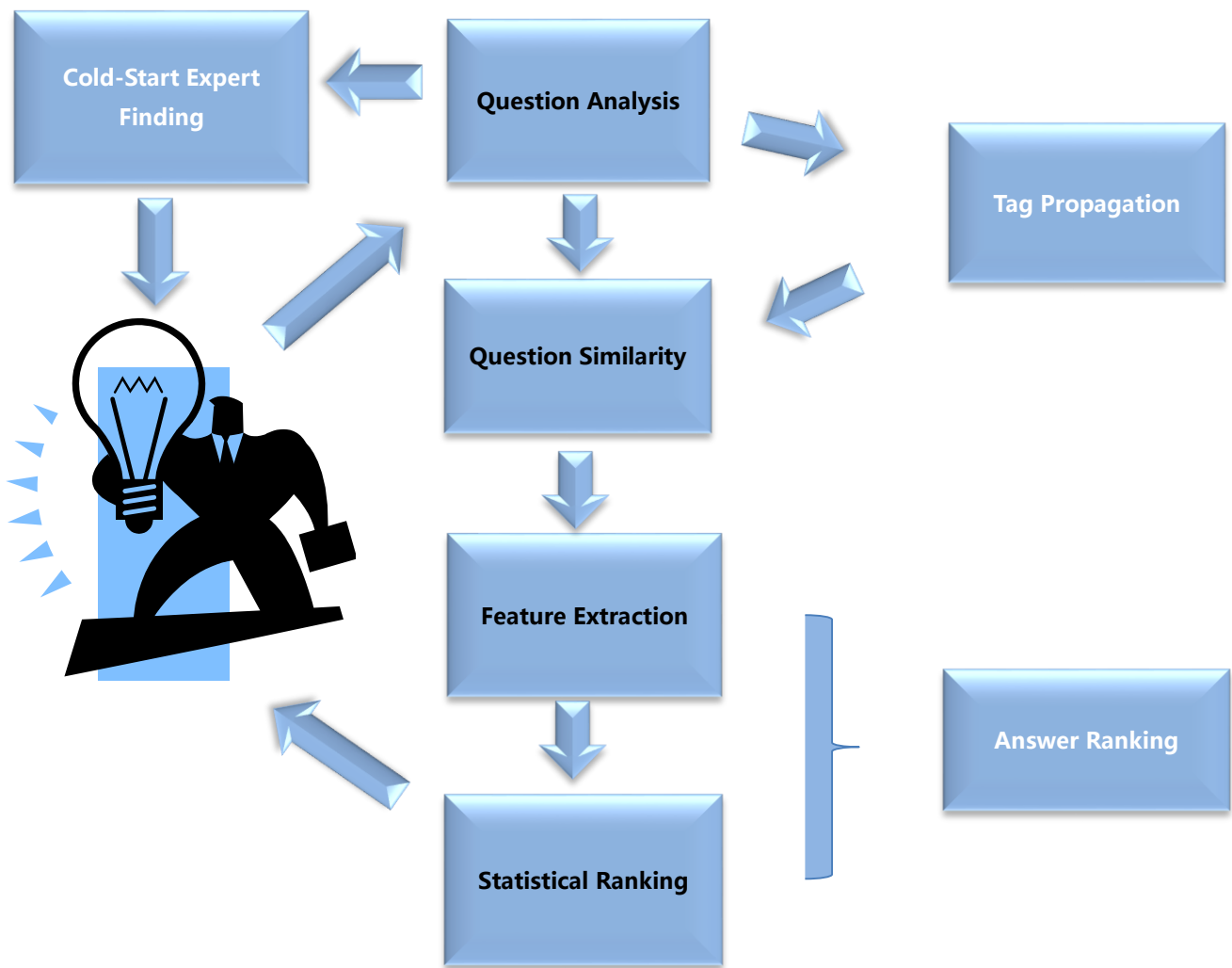
1. Tag Propagation
2. Answer Ranking
3. Cold-Start Expert Finding

We tested our system on Quora data and it performed competently.

Dr.Z Network

Home

Search

Main Functions

About Us

Contact

Welcome to Dr.Z Network

Dr.Z Network is a CQA Platform which aims to find an effectively ranked list of answers as well as some cold-start experts to provide new answers to the questions. When a new question enters our system, the Question Analysis function prepares the question for three tasks, which are the Question Similarity, the Tag Propagation and the Cold-Start Expert Finding. The Cold-Start Expert Finding receives the question and effectively finds cold-start experts to answer the question. The Tag Propagation analyzes the question and generates a list of tags for the question. The Question Similarity utilizes the raw question as well as the tags generated by Tag Propagation to search for similar questions in the database finding candidate answers. The answers along with the raw question are passed to the Answer Ranking function which includes two parts, the Feature Extraction and the Ranking. It will output an effectively ranked list of final answers.

A screenshot of our website.

Tag Propagation

In CQA systems the number of tags will directly affect the quantity of answers to a question. Thus we propose a semi-supervised tag recommendation schema, label propagation dynamically, which learns the correlation between labels based on the social network in a dynamic way.

With this algorithm we discovered the relationship between tags to improve the multi-label methods as well as solve the tag incompleteness problem.

Set Primary Topic: Social Networks (online) ✖
What is lacking in today's social networking services? Edit
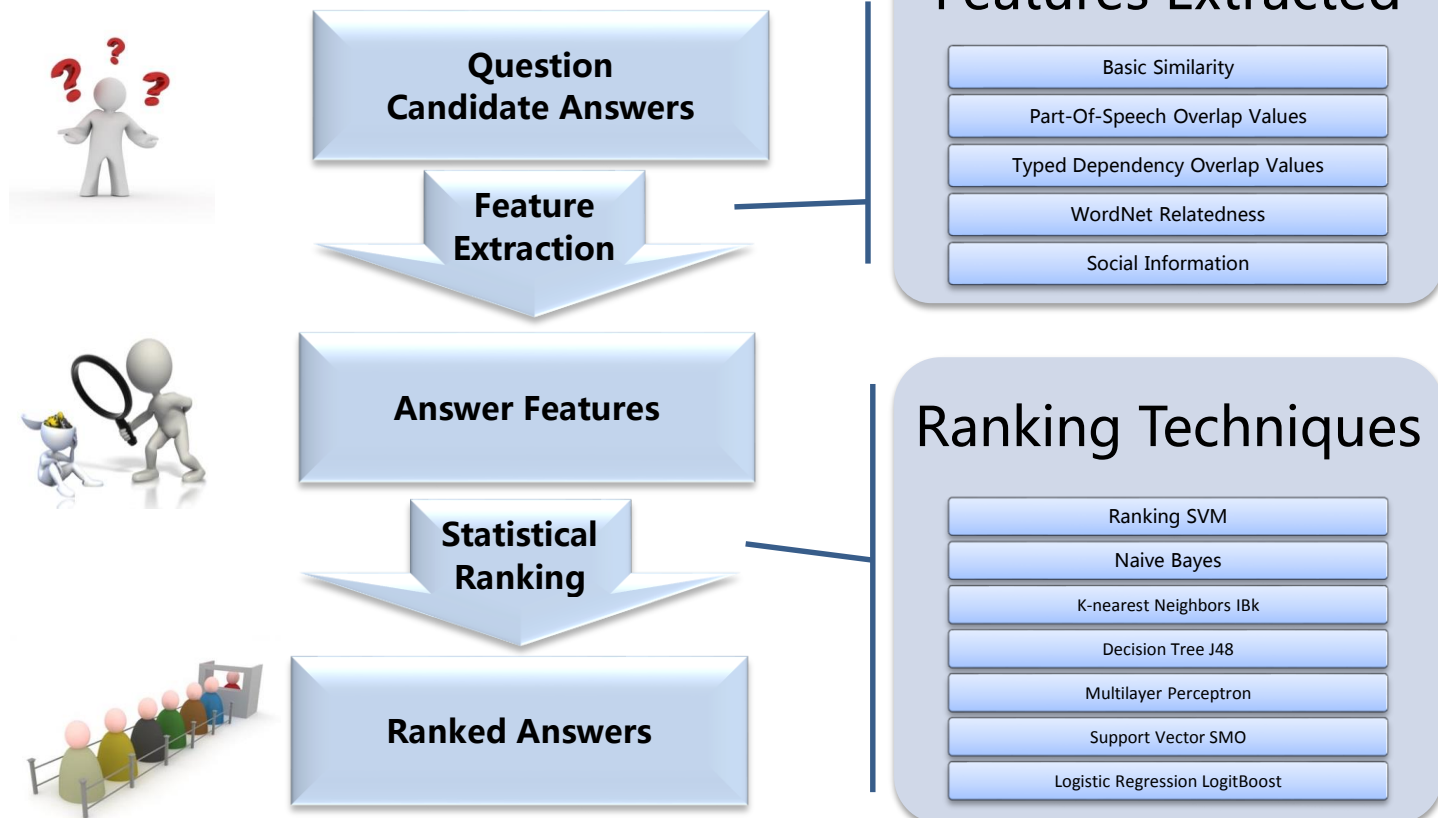
7 Answers   Ask to Answer

Social Networks (online): Discovery   Consumer Internet   Product Design (software)   Design   Twitter   Questions That Contain Assumptions   Edit
What is currently lacking in social networks? Edit

115 Answers   Ask to Answer
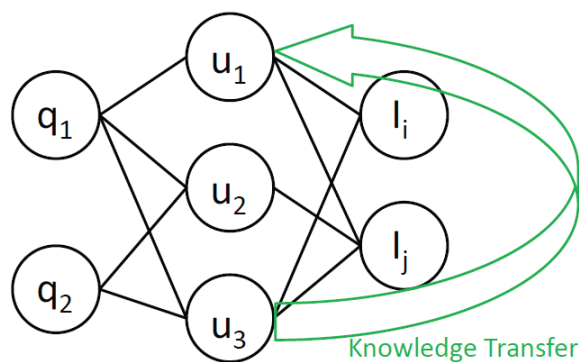
Answer Ranking

Question Candidate Answers

↓

Feature Extraction

Answer Features

↓

Statistical Ranking

Ranked Answers

Features Extracted

- Basic Similarity
- Part-Of-Speech Overlap Values
- Typed Dependency Overlap Values
- WordNet Relatedness
- Social Information

Ranking Techniques

- Ranking SVM
- Naive Bayes
- K-nearest Neighbors IBk
- Decision Tree J48
- Multilayer Perceptron
- Support Vector SMO
- Logistic Regression LogitBoost

Cold-Start Expert Finding

When a new user first joins a CQA system, no previous activity is available and thus the user is said to be cold-start. We can see that majority of the users only answered less than ten questions. We can also see that a significant number of cold-start users obtain high scores in much questions. Thus by transferring the knowledge information of hot-start users on Twitter to cold-start users through their common interest connections, we could also find the experts with

Knowledge Transfer

S	q_1	q_2
u_1	1	?
u_2	-1	1
u_3	1	-1

Conclusions

In this project, we built an effective Question Answering system that can find for each newly inputted question the best experts to answer it using Cold-Start Expert Finding, the tags to label it using Tag Propagation and a ranked list of existing answers to solve it using Answer Ranking.