

Data-driven Answer Selection in Community QA Systems

ABSTRACT

Community question answering system (CQA), one of the fastest-growing user-generated-content (UGC) portals, has raised as an enormous market, so to speak, for the fulfillment of complex information needs. CQA enables users to ask/answer questions and search through the archived historical question-answer (QA) pairs.

EXISTING SYSTEM

Finding similar questions from historical archives has been applied to question answering, with well theoretical underpinnings and great practical success. Nevertheless, each question in the returned candidate pool often associates with multiple answers, and hence users have to painstakingly browse a lot before finding the correct one.

DRAWBACKS

- Information seekers usually have to wait a long time before getting answers to their questions.
- A large proportion of questions do not get any response even within a relatively long period.

PROPOSED SYSTEM

We present a novel scheme to rank answer candidates via pair wise comparisons. In particular, it consists of one offline learning component and one online search component. In the offline learning component, we first automatically establish the positive, negative, and neutral training samples in terms of preference pairs guided by our data-driven observations. We then present a novel model to jointly incorporate these three types of training samples. The closed-form solution of this model is derived. In the online search component, we first collect a pool of answer candidates for the given question via finding its similar questions. We then sort the answer candidates by leveraging the offline trained model to judge the preference orders.

ADVANTAGES

- Our model can achieve better performance than several state-of-the-art answer selection baselines
- Our model is non-sensitive to its parameters
- Our model is robust to the noises caused by enlarging the number of returned similar questions
- The pair wise learning to rank models including our proposed PLANE are very sensitive to the error training samples.

SYSTEM REQUIREMENTS

H/W System Configuration:-

Processor	-	Pentium –III
RAM	-	256 MB (min)
Hard Disk	-	20 GB
Key Board	-	Standard Windows Keyboard
Mouse	-	Two or Three Button Mouse
Monitor	-	SVGA

S/W System Configuration:-

Operating System	:	Windows95/98/2000/XP
Application Server	:	Tomcat5.0/6.X
Front End	:	HTML, Jsp
Scripts	:	JavaScript.
Server side Script	:	Java Server Pages.
Database	:	MySQL 5.0

Database Connectivity : JDBC

www.takeoffprojects.com