A Regularized Competition Model for Question Difficulty Estimation in Community Question Answering Services

Quan Wang† Jing Liu‡ Bin Wang† Li Guo†

† Institute of Information Engineering, CAS ‡ Harbin Institute of Technology

Introduction

Research Problem

- **Question difficulty estimation in community question answering**
- **Applications**
  - Question routing, incentive mechanism design, linguistic analysis

Previous Solutions

- **Competition-based methods**
  - Extract pairwise competitions from question answering threads
  - Estimate question difficulty based on extracted competitions
    - TrueSkill (Liu et al., 2013)
    - PageRank (Yang et al., 2008)
- **Drawbacks**
  - Data sparsity issue: each question gets only two competitions
  - Cold-start issue: cannot handle questions with no answers received

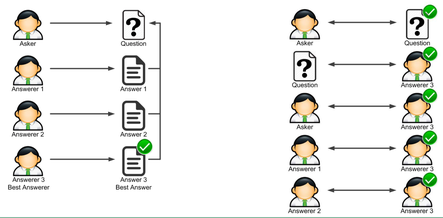
Our Solution

- **Competitions + textual descriptions**
  - For data sparsity issue: textual descriptions provide additional information
  - For cold-start issue: textual descriptions link cold-start questions to well-resolved ones

Regularized Competition Model

Assumption I: pairwise comparison assumption

- Question's difficulty > asker's skill
- Question's difficulty < best answerer's skill
- Best answerer's skill > all other answerers' skill



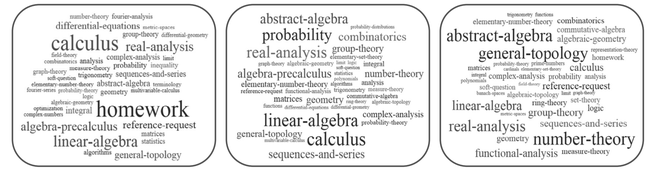
For assumption I: a margin-based loss

$$\ell\left(\bar{\theta}_i, \bar{\theta}_j\right) = \max\left(0, \delta - \left(\bar{\theta}_j - \bar{\theta}_i\right)\right)^p, \quad p = 1 \text{ or } 2$$

- Express question difficulty and user skill on the same scale
- If estimation is consistent with assumption, the loss is zero
- Otherwise, the loss is proportional to the violation

Assumption II: smoothness assumption

- Questions close to each other in textual descriptions have similar difficulty



(a) Low difficulty (b) Medium difficulty (c) High difficulty

Figure: Tag clouds of SO/Math questions with different difficulty levels

For assumption II: manifold regularization

$$\mathcal{R} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\bar{\theta}_i^{(q)} - \bar{\theta}_j^{(q)}\right)^2 w_{ij} = \bar{\boldsymbol{\theta}}_q^T \mathbf{L} \bar{\boldsymbol{\theta}}_q$$

- If textual descriptions are similar, difficulty gap will be small
- Can choose a variety of term weighting schemas
- Can choose a variety of similarity measures

$$\min_{\bar{\boldsymbol{\theta}}} \sum_{(i<j)\in C} \ell\left(\bar{\theta}_i, \bar{\theta}_j\right) + \frac{\lambda_1}{2}\bar{\boldsymbol{\theta}}^T\bar{\boldsymbol{\theta}} + \frac{\lambda_2}{2}\bar{\boldsymbol{\theta}}_q^T \mathbf{L}\bar{\boldsymbol{\theta}}_q$$

Experiments

Experimental Settings

- **Datasets**
  - SO/Math: 10528 questions and 6564 users
  - SO/CPP: 10164 questions and 14884 users
  - For evaluation
    - 539 annotated SO/Math question pairs
    - 521 annotated SO/CPP question pairs
    - Development/test/cold-start split
- **Baselines**
  - TrueSkill (TS), PageRank (PR), Competition Model (CM)
- **Evaluation metric**
  - Accuracy: proportion of question pairs that are correctly judged

Evaluation for Resolved Questions

- **Results**
  - RCM preforms significant better on both datasets
  - Improvements can be achieved by a variety of term weighting schemas and similarity measures
  - Improvements on SO/Math are greater than those on SO/CPP

	PR	TS	CM		RCM	
			H	Q	H	Q
SO/CPP	0.5876	0.6134	0.6340	0.6753	**0.7371**	0.7268
SO/Math	0.6067	0.6109	0.6527	0.6820	**0.7699**	**0.7699**

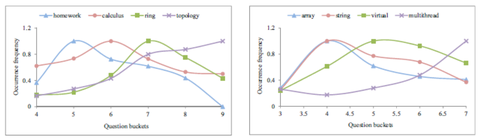
Evaluation for Cold-Start Questions

- **Procedures**
  - Select k well-resolved questions closest in textual descriptions as nearest neighbors
  - Calculate average difficulty of nearest neighbors
- **Results**
  - RCM performs consistently better on both datasets with different k values

	PR	TS	CM		RCM	
			H	Q	H	Q
SO/CPP	0.5870	0.5413	0.6120	0.6304	0.6380	**0.6609**
SO/Math	0.6411	0.6305	0.6653	0.7263	0.6958	**0.7442**

Difficulty Levels of Words

- **Procedures**
  - Split questions into buckets according to their difficulty
  - Calculate the frequency of a word in each bucket
- **Results**
  - RCM might provide an automatic way to measure difficulty levels of words



SO/Math SO/CPP

EMNLP 2014: Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, Oct 25--29, 2014

中国科学院信息工程研究所 INSTITUTE OF INFORMATION ENGINEERING,CAS
哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY