ZhihuRank: A Topic-Sensitive Expert Finding Algorithm in Community Question Answering Websites

Xuebo Liu, Shuang Ye, Xin Li, Yonghao Luo, and Yanghui Rao

ICWL 2015,Guangzhou,China

Outline

- Introduction
- ZhihuRank Algorithm
- Experiment
- Conclusion & Future work

Introduction

How to find expert for every questions in community question answering websites?

Community question answering websites such as Quora, Stack Overflow and Zhihu play an important role in our life. But there is a problem that many questions in these websites have no answer. One of the key points to solve this problem is to find out the users who can answer them, and then we can advise them to answer these questions.

来自子话题：推荐系统

移动端资讯类产品的核心竞争力是什么?

还没有回答 + 关注问题

NO ANSWER!

来自子话题：机器人

机器人这只股票的投资意见?

还没有回答 + 关注问题

来自子话题：数据挖掘

数据挖掘offer选择，网易还是腾讯?

2 个回答 + 关注问题

人类参加图灵测试是一种怎样的体验?

还没有回答 + 关注问题

来自子话题：数据挖掘

有必要自己写一套机器学习库吗?

还没有回答 + 关注问题

来自子话题：深度学习（Deep Learning）

安卓怎么调用基于C++实现的CNN前向传播算法? opencv是怎么实现的安卓接口呢

还没有回答 + 关注问题

Introduction

We created a method to find out experts in community question answering websites and experimental results show the effectiveness of our model when compared to other existing methods.

ZhihuRank Algorithm

Our model considers two aspects:

1. Relationship between users
2. Topical similarity

ZhihuRank Algorithm

Relationship between users

$$UR_z = \lambda T_z \times UR_z + (1 - \lambda) UZ''_z$$

$$T_z(i,j) = \frac{V_{j \to i}}{\sum_{for\ every\ user\ k} V_{j \to k}} \times sim_z(i,j),$$

We use PageRank to get user ranking in topic z.

ZhihuRank Algorithm

Topical similarity

We apply LDA to perform the topic extraction of users and questions, here are some topics:

TOPIC 1	英雄 (Hero)	Dota (Dota)	技能 (Skill)	比赛 (Contest)	游戏 (Game)
TOPIC 2	时间 (Time)	学习 (Study)	事情 (Matter)	很多 (Many)	工作 (Work)
TOPIC 3	情况 (Situation)	很多 (Many)	来说 (In terms of)	时间 (Time)	两个 (Two)
TOPIC 4	知乎 (Zhihu)	http (http)	用户 (User)	内容 (Content)	网站 (Website)
TOPIC 5	用户 (User)	产品 (Product)	互联网 (Internet)	微信 (WeChat)	需求 (Requirement)

ZhihuRank Algorithm

Topical similarity

We consider topical similarity between users when computing transition matrix *Tz*. And we consider topical similarity between users and questions when we compute user authority ranking of every questions *QR* using the formula below.

$$QR = QZ \times UR$$

Experiment

Dataset

We use the data from a Chinese community question answering website: Zhihu(知乎). There are more than 200k texts in our dataset.

- Traditional Chinese -> Simplified Chinese
- Delete stop words
- Chinese word segmentation

Experiment

Evaluation Metrics

Two evaluation metrics commonly used in information retrieval were chosen: MRR (Mean Reciprocal Rank) and NDCG (Normalized Discounted Cumulative Gain). This table is performance for expert finding for different methods.

Alogrithm	MRR	nDCG
IDF	0.75676	0.85459
IDV	0.74568	0.84730
PRF	0.78899	0.86280
PRV	0.79643	0.86455
TSPR	0.63710	0.77037
ZR	**0.84114**	**0.87893**

自然语言处理 修改

自然语言处理怎么最快入门? 修改

How to learn NLP?

刘知远, NLPer

researcher in Tsinghua university

北京 | 高等教育 | ♂

清华大学 | 助理研究员

清华大学 | 自然语言处理

Conclusion & Future work

We proposed an effective algorithm to find out experts in community question answering websites, this algorithm considers two aspects: **Relationship between users** and **Topical similarity .**

In the future we will try to reduce this algorithm’s computation complexity and parallelizing this algorithm, to make it’s possible to be used in real-world websites.

Questions?

Thanks!

lufo816@gmail.com