My first document

Duan Xueye

2013-09-01

1 INTRODUCTION

2 DATASET ANALYSIS

2.1 Data Collection

Zhubajie [1](ZBJ) is a famous online freelancing marketplace platform in China. Employers publish different kinds of human intelligent tasks(HITs) on ZBJ for freelancers to work on. There are three kinds of HITs in this platform: tender, pitch and piece-work. Tender is a kind of task that after the Employers publish a task with content and salary, freelancers submit description of themselves, and employer make a choice to work with one of the candidates. While a pitch task is a task that freelancers directly submit their works for employer to choose, and one or several of the candidates are picked. In a piece-work task, employer need different freelancers to fulfill a lot of micro-tasks. We crawl all the tasks and their corresponding freelancer submissions from ZBJ platform in 2015. We select eight categories of the tender and pitch HITs and tag them with five(need to be alternated) price bins according to the salary of the task.

2.2 Data Features

TO DO

3 Chosen work Prediction

In this section we will formulate the chosen work prediction problem in a tender or pitch HIT. We will then describe features of the freelancers, employers and submitted work.

3.1 Problem Formulation

Considering employer $e \in E$ (Employer set) publishes a HIT, t_{me} which is the m-th HIT from the whole HITs set T. And a freelancer $f \in F$ (Freelancer set) submit the n-th work w_{femn} to t_{me}. Our work is to predict the $top - k$ works which are going to be rewarded by e in t_{me}. If a work w_{femn} is chosen , then the result $r(w_{femn}) = 1$, otherwise $r(w_{femn}) = 0$.

3.2 Data features for prediction

Freelancer features Freelancer features describe the characters of the freelancer f to enable us to acquire to judge the ability of f to fulfill a HIT.

Category number Number of categories a freelancer f worked on.

---

[1]www.zbj.com

Number of works in related category   Describe the category-related interest of the freelancer f.

Fraction of works in related category   Also describe the category-related interest of the freelancer f. Fraction of work in related category is the number of works in related category divided by the total number of works

Number of works in related price bin   Describe the price-related interest of the freelancer f.

Fraction of work sin related price bin   Also describe the price-related interest of the freelancer f. Fraction of work in related price bin is the number of works in related price bin divided by the total number of works

System reputation   ZBJ platform gives a freelancer f a system reputation level when f fulfilled certain HITs. The more works, that are submitted by f, get reward by employers, the higher level the f will get from the system.

Efficiency   Efficiency is the number of works that f submits in a limited time period. We use three different time periods such as 7 days, 14 days and 30 days. Efficiency is a character that show us the work ability and passion of a freelancer f.

Category-related efficiency and Price-related efficiency   Similarly to efficiency, we calculate category-related efficiency and price-related efficiency in different categories or different price bins for a freelancer f.

Quality   We calculate the number of works rewarded and the number of works costs and a success rate of a freelancer. $n_{suc}(f)$ denote the number of works worked by f that get rewarded, while $n_{fail}(f)$ denote the number of works worked by f that do not get rewarded even when the task is finished. So the number of works rewarded is $n_{suc}(f)$; the number of works costs is $n_{suc}(f) - n_{fail}(f)$; success rate is $n_{suc}(f)/(n_{suc}(f) + n_{fail}(f))$.

Category-related Quality and Price-related Quality   Similarly to Quality, we calculate category-related Quality and price-related Quality in different categories or different price bins for a freelancer f.

Submitted Work Features

Relative reputation   The system reputation level of a work is the the system reputation level of the freelancer who submits the work. We use $rank(w_{femn}, t_{me})$ denote the system reputation level rank of work $repRank(w_{femn})$ compared with other works of the task t_{me}. $repRank(w_{femn}, t_{me})$ is the number of works whose system reputation level is bigger then the system reputation level of w_{femn}. And we use $num(t_{me})$ denote the total number of works submitted to the task t_{me}. Relative reputation is then denoted by $\frac{repRank(w_{femn}, t_{me})}{num(t_{me})}$.

Response time   Response time is the count of days after a task t_{me} is published.

Relative response time   Similar to relative reputation, we use $resposeRank(w_{femn})$ denote the time rank of a work w_{femn} submiteed to t_{me}. The earlier a work is submitted the lower its ranker is. And the relative response time is $\frac{resposeRank(w_{femn}, t_{me})}{num(t_{me})}$.

Employer Features   Employer features describe the preference of a employer e to make a choice.

Reputation preference   We calculate the median and mean system reputation level of the works selected by the employer e.

Relative reputation preference   We calculate the median and mean relative system reputation level of the works selected by the employer e.

Response time preference   We calculate the median and mean system response time works selected by the employer e.

Response time preference   This is a test line;

Response time preference   This is a test line;

Response time preference   This is a test line;

Response time preference   This is a test line;

Response time preference   This is a test line;

Response time preference   This is a test line;

Response time preference   This is a test line;

Response time preference   This is a test line;

4 METHODOLOGY