Pricing Crowdsourcing-based Software Development Tasks
(ICSE'13-NIER)

Ke Mao
dreamict[at]gmail.com

Feb. 11th, 2013
@ CREST, UCL
Institute of Software, Chinese Academy of Sciences
Pricing Crowdsourcing-based Software Development Tasks

Ke Mao*,†, Ye Yang*, Mingshu Li*
*Institute of Software, Chinese Academy of Sciences
†University of Chinese Academy of Sciences
Beijing, China
{maoke, yangye, mingshu}@nfs.iscas.ac.cn

Mark Harman
Dept. of Computer Science,
University of College London
London, UK
mark.harman@ucl.ac.uk

Ke Mao
Master Student at Institute of Software, Chinese Academy of Sciences
China | Computer Software

http://www.linkedin.com/in/kemao
Overview

- **Background**
  - Crowdsourcing: Micro task VS. Complex task
  - The TopCoder Platform

- **Motivation**
  - New Phenomenon
  - The Pricing Issue

- **Methodology**

- **Experiments & Insights**

- **Conclusion**
Overview

- Background
  - Crowdsourcing: Micro task VS. Complex task
  - The TopCoder Platform
A Recent News...

- Typical work day of one star developer:
  - 09:00 a.m. – Arrive and surf Reddit, watch cat videos
  - 11:30 a.m. – Take lunch
  - 01:00 p.m. – Ebay time
  - 02:00 p.m. – Facebook updates – LinkedIn
  - 04:30 p.m. – End of day update e-mail to management
  - 05:00 p.m. – Go home
Introduction to Crowdsourcing

- A proper way...
  - Labor of the Internet
  - Low cost
  - Suprising deliverable

- Wisdom of the Crowd
What is Crowdsourcing?

- "Crowdsourcing" defined by Jeff Howe:
  - The act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call.

- Crowdsourcing VS. Outsourcing:
  - The crucial prerequisite is the use of the open call format and the large network of potential laborers.
Micro Task VS. Complex Task

Crowdfunding
Finanzielle Hilfe von Online-Darlehen, Sponsoren oder Spendern für Non-Profit Initiativen oder Unternehmen.

Tools
Anwendungen, Plattformen und Tools, die die Zusammenarbeit, Kommunikation und die Abwicklung von Gruppen von Menschen fördern.

Cloud Collaboration
Nutzung eines dispersen, virtuellen Pools von Arbeitnehmern, der zur Verfügung steht, um Aufgaben verschiedener Komplexität zu erfüllen.

Civic Engagement
Kollektive Aktionen im öffentlichem Interesse.

Collective Knowledge
Sammlung von Wissen und Information aus einem Pool von Mitarbeitern.

Collective Creativity
Creative Talent Pools, in denen Kunst- oder Medien-Inhalte entwickelt werden.

Community Building
Aufbau von Communities durch aktivtes Engagement von Einzelpersonen, die gemeinsame Leidenschaften, Überzeugungen und Interessen teilen.

Open Innovation
Nutzung von Ressourcen außerhalb des Unternehmens um gemeinsam Ideen zu generieren, zu entdecken und umzusetzen.
Micro Task VS. Complex Task

GROWTH of workers by category

Workers doubled in number 2010-2011

in 2011
Number of workers based on 26 CSPs

+151%
+131%

FREELANCE IDEATION EXPERTISE MICRO-TASKS SOFTWARE

+165%
+103%

2009 2011
ALL CATEGORIES

Credit: http://sandfishdesign.co.uk, © 2012, Crowdsourcing, LLC
What is TopCoder?

- The world's largest competitive community for crowdsourced software development

- The TopCoder Community is 425,993 strong

- Membership
  - China
  - India
  - U.S.
What is TopCoder?

- What kinds of projects can I do with TopCoder?
  - Mobile Applications
  - Analytics and Optimization
  - Scientific Algorithm Development
  - Online Communities
  - Open Platforms
  - Digital Media
  - Business Systems
  - ...

How Does It Work?

Typical Duration is 19 - 24 Weeks

Build “Factory”
Virtual Global Work Force
TopCoder Community
Competition-based

Credit: www.topcoder.com, © 2007, TopCoder, Inc
Overview

- **Motivation**
  - New Phenomenon
  - The Pricing Issue
Motivation - New Phenomenon

- New Paradigm
  - Crowdsourced development
    - 1. **Open** call format
    - 2. Large networked potential labor

Fig. 1 Illustration of crowdsourcing-based software development process.
Motivation - New Phenomenon

- New Phenomenon In SE activity
  - 2 examples that challenge traditional law
    - Parkinson's Law
    - COCOMO Model
Motivation - New Phenomenon

- Parkinson's Law

("Work expands so as to fill the time available for its completion. ")

Fig. 2 Correlation between the time allocated and the actual time consumed
**Motivation - New Phenomenon**

- **Basic COCOMO Model**

\[ \text{EFFORT} = a \times \text{SIZE}^b \]

Fig. 3 The effort estimated by COCOMO model, compared to the actual effort.
Motivation - The Pricing Issue

Inappropriate price often lead to low capital efficiency and task starvation

How to build empirical pricing models?
Overview

- Methodology
## Methodology

- **Price Drivers**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Meaning</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Regression Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>TECH</td>
<td>number of technologies which will be used</td>
<td>1.0</td>
<td>7.0</td>
<td>2.1</td>
<td>2.0</td>
<td>1.3</td>
<td>-6.265</td>
</tr>
<tr>
<td>DEPE</td>
<td>number of component dependencies</td>
<td>1.9</td>
<td>3.9</td>
<td>2.0</td>
<td>3.0</td>
<td>2.555</td>
<td>0.606</td>
</tr>
<tr>
<td>REQU</td>
<td>number of pages of requirement specification</td>
<td>1.7</td>
<td>3.7</td>
<td>1.8</td>
<td>2.0</td>
<td>-9.422</td>
<td>-1.332</td>
</tr>
<tr>
<td>COMP</td>
<td>number of pages of component specification</td>
<td>3.0</td>
<td>10.0</td>
<td>5.7</td>
<td>8.480</td>
<td>3.414</td>
<td>0.001</td>
</tr>
<tr>
<td>SEQU</td>
<td>number of sequence diagrams of the design</td>
<td>3.0</td>
<td>5.5</td>
<td>3.0</td>
<td>6.5</td>
<td>-3.032</td>
<td>-1.360</td>
</tr>
<tr>
<td>SCOR</td>
<td>score of winner’s submission of design phase</td>
<td>92.0</td>
<td>92.0</td>
<td>6.5</td>
<td>6.5</td>
<td>-3.032</td>
<td>-1.360</td>
</tr>
<tr>
<td>AWRD</td>
<td>winner’s award of design phase, measured in dollars</td>
<td>0.0</td>
<td>4200.0</td>
<td>920.1</td>
<td>900.0</td>
<td>457.6</td>
<td>0.503</td>
</tr>
<tr>
<td>EFRT</td>
<td>winner’s effort in design phase, measured in days</td>
<td>0.0</td>
<td>468.0</td>
<td>29.0</td>
<td>5.0</td>
<td>77.2</td>
<td>0.273</td>
</tr>
<tr>
<td>SUML</td>
<td>size of UML design file, measured in KB</td>
<td>142.0</td>
<td>142.0</td>
<td>85.5</td>
<td>85.5</td>
<td>-0.293</td>
<td>-1.386</td>
</tr>
<tr>
<td>WRAT</td>
<td>rating of the winner in design phase</td>
<td>1632.0</td>
<td>1632.0</td>
<td>629.3</td>
<td>629.3</td>
<td>-0.021</td>
<td>-0.994</td>
</tr>
<tr>
<td>REGI</td>
<td>number of registrants in design phase</td>
<td>8.0</td>
<td>5.2</td>
<td>2.692</td>
<td>2.692</td>
<td>0.914</td>
<td>0.361</td>
</tr>
<tr>
<td>SUBM</td>
<td>number of submissions in design phase</td>
<td>2.0</td>
<td>2.4</td>
<td>-10.416</td>
<td>-10.416</td>
<td>1.601</td>
<td>0.110</td>
</tr>
<tr>
<td>ISUP</td>
<td>indicating if the task aim at component update</td>
<td>0.0</td>
<td>1.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.4</td>
<td>-71.352</td>
</tr>
<tr>
<td>ISJA</td>
<td>indicating if the development language is Java</td>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>ISCS</td>
<td>indicating if the development language is C#</td>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.5</td>
<td>23.986</td>
</tr>
<tr>
<td>SIZE</td>
<td>estimated size of the component, measured in KSLOC</td>
<td>0.4</td>
<td>21.9</td>
<td>3.0</td>
<td>2.3</td>
<td>2.3</td>
<td>29.772</td>
</tr>
<tr>
<td>const</td>
<td>the constant term</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

TABLE I. DESCRIPTIVE STATISTICS AND REGRESSION COEFFICIENTS OF PROPOSED FACTORS
Methodology

- Predictive Models
  - Multiple Linear Regression Model:

\[
PRICE = \beta_1 \text{TECH} + \beta_2 \text{DEPE} + \beta_3 \text{REQU} + \beta_4 \text{COMP} + \beta_5 \text{SEQU} + \beta_6 \text{SCOR}
+ \beta_7 \text{AWRD} + \beta_8 \text{EFRT} + \beta_9 \text{SUML} + \beta_{10} \text{WRAT} + \beta_{11} \text{REGI} + \beta_{12} \text{SUBM}
+ \beta_{13} \text{ISUP} + \beta_{14} \text{ISJA} + \beta_{15} \text{ISCS} + \beta_{16} \text{SIZE} + \beta_0 + \epsilon
\]  

- 8 other Machine Learning & Statistical models

<table>
<thead>
<tr>
<th>3 Decision Tree based learners</th>
<th>2 Instance based learners</th>
<th>1 Neural Net</th>
<th>1 Support Vector Machine</th>
<th>1 Logistic Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5.0, CART, QUEST</td>
<td>KNN-1, KNN-k ∈ [3, 7]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Overview

- Experiments & Insights
Experiments

- **Aim:**
  - To answer the following RQs.

- **RQs:**
  - *Baseline Comparison*
    - How much better?
  - *Performance Assessment*
    - Which is the best?
  - *Actionable Insights*
    - What guidance can we offer?
Experiments

- Dataset
  - Sep 29th 2003 to Sep 2nd 2012
  - 2,895 design and 3,015 development tasks
  - 490 successful sw dev projects from TopCoder

- Validation method
  - LOOCV
Experiments

- **Performance Measures:**

  \[
  MRE_i = \left| \text{actual}_i - \text{estimated}_i \right| / \text{actual}_i
  \]

  \[
  MMRE = \frac{\sum_i MRE_i}{T}
  \]

  \[
  MdMRE = \text{median}(MRE_1, MRE_2, ..., MRE_i, ..., MRE_T)
  \]

  \[
  StdMRE = \frac{1}{T - 1} \sqrt{\sum_i (MRE_i - MMRE)^2}
  \]

  \[
  Pred(N) = \frac{1}{T} \sum_i \left\{ \begin{array}{ll} 
  1 & \text{if } MRE_i \leq N / 100 \\
  0 & \text{otherwise}
  \end{array} \right.
  \]

  (where \( i \in \{1...T\} \))
Experimental Results

- **Answer to RQ1:**
  - Outperformed by all 9 predictive models, according to Pred(30) measure

![Performance of pricing models learned by each approach](image)

Fig. 5 Performance of pricing models learned by each approach
Experimental Results

- Answer to RQ2:
  - Decision tree based learners
  - C5.0, QUEST, CART

Fig. 5 Performance of pricing models learned by each approach
Insights

- Answer to RQ3:
  - Significance Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Meaning</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>S.Dev</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>TECH</td>
<td>number of technologies which will be used</td>
<td>1.0</td>
<td>7.0</td>
<td>2.1</td>
<td>2.0</td>
<td>1.3</td>
<td>-6.265</td>
<td>-0.569</td>
<td>0.570</td>
</tr>
<tr>
<td>DEPE</td>
<td>number of component dependencies</td>
<td>0.0</td>
<td>16.0</td>
<td>2.9</td>
<td>2.0</td>
<td>3.0</td>
<td>2.555</td>
<td>0.060</td>
<td>0.546</td>
</tr>
<tr>
<td>REQU</td>
<td>number of pages of requirement specification</td>
<td>2.0</td>
<td>23.0</td>
<td>3.7</td>
<td>3.0</td>
<td>1.8</td>
<td>-9.422</td>
<td>-1.332</td>
<td>0.183</td>
</tr>
<tr>
<td>COMP</td>
<td>number of pages of component specification</td>
<td>3.0</td>
<td>56.0</td>
<td>11.3</td>
<td>10.0</td>
<td>5.7</td>
<td>8.480</td>
<td>3.414</td>
<td>0.001</td>
</tr>
<tr>
<td>SEQU</td>
<td>number of sequence diagrams of the design</td>
<td>0.0</td>
<td>45.0</td>
<td>5.3</td>
<td>3.0</td>
<td>5.5</td>
<td>7.666</td>
<td>2.922</td>
<td>0.004</td>
</tr>
<tr>
<td>SCOR</td>
<td>score of winner’s submission of design phase</td>
<td>70.2</td>
<td>99.7</td>
<td>90.4</td>
<td>92.0</td>
<td>6.5</td>
<td>-3.032</td>
<td>-1.360</td>
<td>0.175</td>
</tr>
<tr>
<td>AWRD</td>
<td>winner’s award of design phase, measured in dollars</td>
<td>0.0</td>
<td>4200.0</td>
<td>920.1</td>
<td>900.0</td>
<td>457.6</td>
<td>0.503</td>
<td>17.912</td>
<td>0.000</td>
</tr>
<tr>
<td>EFRP</td>
<td>winner’s effort in design phase, measured in days</td>
<td>0.0</td>
<td>468.0</td>
<td>29.0</td>
<td>5.0</td>
<td>77.2</td>
<td>0.273</td>
<td>1.607</td>
<td>0.109</td>
</tr>
<tr>
<td>SUML</td>
<td>size of UML design file, measured in KB</td>
<td>24.0</td>
<td>644.0</td>
<td>158.1</td>
<td>142.0</td>
<td>85.5</td>
<td>-0.293</td>
<td>-1.386</td>
<td>0.166</td>
</tr>
<tr>
<td>WRAT</td>
<td>rating of the winner in design phase</td>
<td>481.0</td>
<td>3251.0</td>
<td>1688.2</td>
<td>1632.0</td>
<td>629.3</td>
<td>-0.021</td>
<td>-0.994</td>
<td>0.321</td>
</tr>
<tr>
<td>REGI</td>
<td>number of registrants in design phase</td>
<td>1.0</td>
<td>38.0</td>
<td>9.3</td>
<td>8.0</td>
<td>5.2</td>
<td>2.692</td>
<td>0.914</td>
<td>0.361</td>
</tr>
<tr>
<td>SUBM</td>
<td>number of submissions in design phase</td>
<td>1.0</td>
<td>26.0</td>
<td>2.8</td>
<td>2.0</td>
<td>2.4</td>
<td>-10.416</td>
<td>-1.601</td>
<td>0.110</td>
</tr>
<tr>
<td>ISUP</td>
<td>indicating if the task aim at component update</td>
<td>0.0</td>
<td>1.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.4</td>
<td>-71.352</td>
<td>-1.998</td>
<td>0.046</td>
</tr>
<tr>
<td>ISJA</td>
<td>indicating if the development language is Java</td>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ISCQ</td>
<td>indicating if the development language is C#</td>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.5</td>
<td>23.986</td>
<td>0.808</td>
<td>0.419</td>
</tr>
<tr>
<td>SIZE</td>
<td>estimated size of the component, measured in KSLOC</td>
<td>0.4</td>
<td>21.9</td>
<td>3.0</td>
<td>2.3</td>
<td>2.3</td>
<td>29.772</td>
<td>4.189</td>
<td>0.000</td>
</tr>
<tr>
<td>const</td>
<td>the constant term</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>463.194</td>
<td>2.437</td>
<td>0.015</td>
</tr>
</tbody>
</table>

TABLE I. DESCRIPTIVE STATISTICS AND REGRESSION COEFFICIENTS OF PROPOSED FACTORS
Insights

Answer to RQ3:

- Rules of Thumb
  - ISUP $\Rightarrow$ $70$
  - COMP(4 pages) $\Rightarrow$ $30$
  - SEQU(4 diagrams) $\Rightarrow$ $30$
  - SIZE(1 KSLOC) $\Rightarrow$ $30$

- May not always be right
  - But "Why am I bucking the trend?"
Overview

- Conclusion
Conclusion

- Analyzed 5,910 sw dev tasks on TopCoder
- Proposed 16 price drivers
- Assessed 12 empirical pricing models
- Useful prediction quality is achievable \((\text{Pred}(30) > 0.8)\)
- Actionable advice can be extracted from our models to assist the developers on TopCoder
Future Work

- Quality & Risk Factors

- Price / Quality Trade off
  - Assessing task complexity via UML design

- Multi-objective Optimization
  - Price / Quality / Risk
Thanks!