COMMUNITY PROFILING FOR CROWDSOURCING QUERIES

Khalid Belhajjame ¹
Marco Brambilla ²
Daniela Grigori ¹
Andrea Mauri ²

¹ PSL, Paris-Dauphine University, LAMSADE, France
² Politecnico di Milano, Italy
Traditional vs Community Crowdsourcing

• General structure:
  • the requestor poses some questions
  • a wide set of responders are in charge of providing answers (typically unknown to the requestor)
  • the system organizes a response collection campaign

• Traditional Crowdsourcing
  • Cost – Quality Tradeoff
  • Complex results aggregation

• Community Crowdsourcing
  • Matching the task to the “correct” group of workers
Community

Set of people that share
• Interests
• Feature

..or belong to
• common entity
• social network
Leveraging communities

• Why?
  • Experts
  • More engaged

• How?
  • Determine the communities of performers
  • Target the correct community
  • Monitor them taking into the account the behavior of their members
The approach

- Models
  - Query Model
  - Community Model

- Matching strategies
  - Keyword based
  - Semantic based
Query Model

• Textual description of the task
  • Examples of query and responses

• Knowledge needed

• Prior knowledge (knowledge base) that can be used for partially answering or for identifying potential answers.

• Type of the task
  • Unary: tag, classify, like, …
  • N-ary: match, cluster, …

• Objects
  • Kind, description, text, metadata …

• Temporal
Community Model

- Textual description of the community
  - name, web page, ...

- Type of the community
  - Explicit: Statically existing and consolidated
  - Implicit: Dynamically built based on the need

- Definition
  - Intentional: defined by a property
  - Extensional: list of members
  - Both

- Grouping factor
  - Friendship, interest, location, expertise, affiliation
Community Model

• Content
  • Produced by the people of the community

• Members’ profiles
  • Explicit
  • Implicit

• Communication channel
  • Email, facebook, linkedin, twitter, blogs or web sites (reviews, expert sites), AMT
Relations between Communities

• Subsumption
  • A given community contains another community
  • i.e. Sport fans contains Soccer fans

• Similarity
  • Two communities refer to similar expertise or topic
  • i.e. Experts in classical music and experts on opera
Matching

- **Keyword Based**
  - Communities and query treated as bag of words
  - Requires indexing

- **Semantic Based**
  - Communities and query are mapped to concepts
  - Requires semantic annotation
Community Control

Community control consists in the adaptation of the crowdsourcing campaign according to the behavior of the community

- Task / Object allocation (granularity)
- Static / Dynamic
CrowdSearcher

A prototype that allows the definition, execution and control of a crowdsourcing campaign

http://crowdsearcher.search-computing.org/
Example (dynamic control)

e: AFTER UPDATE FOR \( \mu \)TObjExecution

c: CommunityControl[CommunityID== NEW.CommunityID].score<=0.5
CommunityControl[CommunityID== NEW.CommunityID].eval=10

a: SET CommunityControl[CommunityID == DB-Group].Enabled = true
Experiment

- 16 professors within two research groups in our department (DB and AI groups)
- The top 50 images returned by the Google Image API for each query
- Each experts have to evaluate 5 images at time
- Results are accepted when enough agreement on the class of the image is reached
- Evaluated objects are removed from new executions.
Communities

The communities:
- the research group of the professor,
- the research area containing the group (e.g. Computer Science)
- and the whole department (which accounts for more than 600 people in different areas)

Invitations are sent:
- **inside-out**: we started with invitations to experts, e.g. people in the same groups as the professor (DB and AI), and then expanded invitations to Computer Science, then to the whole Department, and finally to open social networks (Alumni and PhDs communities on Facebook and LinkedIn);
- **outside-in**: we proceeded in the opposite way, starting with the Department members, then restricting to Computer Scientists, and finally to the group's members.
Number of performers per community

- research group
- research area
- department
- social network
- total

9 / “a lot”
16%
24%
46%
Precision of performers per community
Precision of the evaluated objects

- Precision decreases with less expert communities
- Inside-out strategy (from expert to generic users) outperforms Outside-in strategy (from generic to expert users)
General observations

A given community of workers can be broken down into (possibly overlapping) sub-communities with different expertise

Experts from community feel more engaged with the task

- They are more demanding with respect to the quality of the application UI and the evaluated objects
- Provide feedbacks on the application, question and the objects evaluated
  - “How is it possible that this image is related to me?!”
Conclusions

- Communities can be effectively used for tasks that require domain expertise
- How to deal with tasks requiring **multiple** expertise
- How to build a knowledge base that allows profiling of both communities and queries in a **optimal way**
- How to cope with the **dynamics** over time of
  - Communities and task (changing needs)
  - Communities and worker expertise
Thanks for your attention

Any Question?

Contacts

Khalid Belhajjame
Marco Brambilla
Daniela Grigori
Andrea Mauri

Khalid.Belhajjame@dauphine.fr
Marco.Brambilla@polimi.it
Daniela.Grigori@dauphine.fr
Andrea.Mauri@polimi.it

http://crowdsearcher.search-computing.org/
References


