

Clash of Crowds: Competitive Crowdsourcing

Tongshuang Wu, Quan Ze Chen, Chenglong Wang

In the last milestone, we refined our research questions and sketched our game mockups. Till now, we have implemented our game platform (for coordinating players to start coop games) and four games (DirectAnnotation, Bingo, HangmanAdd and HangmanSubtract). Furthermore, we designed the data-analysis method that will be used to analyze experiment result.

Remaining tasks include deploy all five games to the crowds and analyze data collected from the experiments.

1 Implementation

We briefly describe our game implementation in this section.

- *DirectAnnotate*. This is the base game where the player directly labels document sentiment and selects words that influence decision making (informative words).
- *Bingo*. This is the game where two players coop on labeling document sentiment and extract informative words. The game requires labels collected from the team are the same, and the game is finished when two players reach an agreement on four informative words.
- *HangmanAdd*. Players in this game are assigned as annotator and guesser. The annotator label the document and provides a list of ranked informative words to assist the guesser to guess the document sentiment. Game result counts on how many words are used and whether the guesser result agree with the annotator's label.
- *HangmanSubtract*. This is a competitive game where the two players are assigned as roles masker and guesser. The masker is asked to label the document and mask informative words that are important to his decision. The masker's goal is to hinder the guess to figure out the correct the document sentiment and the guesser is asked to guess document based on the masked document. (The guesser can unmask words with the reduction on the payment bonus and the masker earn more if labeled less but more effective.)

2 Data Analysis Method

In these games, the following data can be collected (we use • to denote data can be collected and ○ to denote that can be partially collected). Given these collected data, our inter-game analysis plan is listed below.

Data	Annotation	Bingo	HangmanAdd	HangmanSubtract
Labels-full	•	•	•	•
Labels-verify		○	•	•
Features-init	•	•	•	•
Features-rank		•	•	•
Features-refine				•
Features-verify		•	•	•

Table 1: *Labels-full* refers to whether we can extract sentiment label from the game, *labels-verify* refers to whether the game supports different players to verify their label result, *feature-init* refers to whether the game obtains features provided by the gamer, *feature-rank* refers to whether obtained features are ranked, *feature-refine* refers to whether the player can refine initially collected feature, and *feature-verify* refers to whether features are verified though multiple players.

- (All games-Machine Algorithm): which approach best extract label and features and how they differ from ones extract by machine.
- (Annotation-Bingo): Whether the incentive to reach agreement improves the labeling and feature extraction result.
- (HangmanAdd-HangmanSubtract): How effective are the informative words in affecting sentiment decision making (from a collaborative view and and competitive view). Also, how important is the document structure information in contributing to sentiment labeling (the former has no structure information).

3 Next Steps

We are in our process of deploying more games to collect data. The rest of task is to collect and data and perform data analysis in detail.

4 Feedbacks Expecting

We are expecting feedbacks for the followings.

- We notice that many user may not pay attention to the task training instruction, what can we do to improve its effectiveness?
- Besides comparing game results in terms of labeling accuracy (quantitatively) and extract features (qualitatively), is there any other analysis we can do to obtain more interesting result?