CSE 510 Spring '17 Project Milestone 2

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 the first three games (DirectAnnotation, Bingo and HangmanAdd). Furthermore, we designed the data-analysis method that will be used to analyze experiment result.

Remaining tasks include (1) implement the two games left, (2) deploy all five games to the crowds, and (3) analyze data collected from the experiments.

0.1 Implementation

We briefly describe our game implementation in this section.

- *Bingo*: sf

0.2 Evaluation

Our evaluation of the collected data includes two steps. First, we want to evaluate the precision of labels provided by users in the games. This quantitative evaluation directly reflects the causing facts of user performance (e.g., which document modification decreases the precision most in the game). Second, we plan to qualitatively analyze features collected by the users to explain why different types of document modification result in different performance. The qualitative analysis includes: (1) compare features extracted from the same document through different approaches, and (2) analyzing common properties of important features (e.g., part-of-speech, position, context).

1 Next Steps

Before the next milestone, we will quickly implement the sketches in Table. 2. We will iterate through a test-and-test procedure to (1) select 1-2 most suitable games for the real user study, and (2) fix potential problems in prototypes. Afterwards, we will deploy the system, collect and analyze the data.

2 Feedbacks Expecting

We are expecting feedbacks for the followings.

- Is our problem, "understanding human features", well motivated?
- Do our current game designs match our goal? We have vetoed a few designs previously since they may produce a different set data that cannot answer the question. For example, one of our previous design is a game that asking human compete labeling against the computer on documented processed by another player. We vetoed it because it may motivate the second play to choose words to fool the computer rather than naturally choosing decision making features.

GameSystem	Player	Mechanism	Pros&Cons	Difficulty	Compatible Incentive
Direct Annotation (Baseline)	1P	Directly ask users to choose features they used.	+ Access to the full doc – * Hindsight bias – * No checks	Easy	N/A (not a game)
Bingo	2P, Collab	Two users annotate one document. Score when they label the same feature.	+ * Less noisy + Access to the full doc – * Hindsight bias – * Low representation of individual users	Moderate	Reward Mutual agreement
Hangman, *Additive*	2P, Collab	**Guesser** asks for a word from annotator. **Annotator** returns a word to the guesser.	+ * Mediate hindsight bias + Engaging – Takes more time – Guesser cannot access document structure	Moderate	- Reward agreement - Penalize queries used
Censor	2P, Compete	**Censors** mask words in a document. **Identifiers** label the censored document.	+ * Keep doc structure + * Mediate hindsight bias + Engaging and easy – Can't reuse identifiers – Need to retain censors	**Censors**: Moderate **Identifiers**: Easy	**Censors**: Reward fooling users **Identifiers**: Reward for being correct
Hangman, *Subtractive*	2P, Compete	**Censors** remove a set of words from a document. **Guesser** queries one word at a time (can choose position).	+ * Keep doc structure + * Mediate hindsight bias + Engaging and easy – Can't reuse guessers – More mental load balancing prices	**Censors**: Moderate **Identifiers**: Easy	**Censors**: Reward fooling users **Identifiers**: Reward for being correct

Table 1: Designed games. Pros and Cons with "*" are the attributes affecting data quality mentioned in Section. **??**