Games for Games

Manipulating Contexts in Human Computation Games

Aneta Takhtamysheva
Est Creativity Rising
43 Rue de Richelieu
75001 Paris France
aneta9@yahoo.com

Robert Porzel [*]
University of Bremen
TZI, Digital Media Group
28359 Bremen Germany
porzel@tzi.de

Markus Krause [†]
University of Bremen
TZI, Digital Media Group
28359 Bremen Geermany
phateon@tzi.de

ABSTRACT

The present work and demonstration system aims at finding an efficient and cost-effective human computation method to expand the linguistic capabilities of interactive games that need it to respond appropriately to the language based input of their users. As a showcase scenario for the experiments conducted, we took interactive fiction applications and examined how the human computation game design and scoring approaches affects the quality of the data gathered. The ensuing analysis of the data confirms our initial hypothesis that game approaches can provide both the qualitative and quantitative data needed for the corresponding interactive games.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.13 [**Reusable Software**]: Domain engineering—*Domain engineering*

General Terms

Human Computation, Human-based Computation, Computer Games

Keywords

HBC, HC, Games, Data Engineering, Context

1. INTRODUCTION

Natural language data is required not only for natural language processing applications, such as dialog systems or information extraction systems, but also for several types of games, such as interactive fiction- or various types of role playing games. In either case, the data collection method applied needs to provide this required data in both quantitatively and qualitatively adequate ways. In the case of NLP applications tried and tested methods for collecting natural language data exist, e.g. in the form of the so-called Wizardof- Oz paradigm [2] or other experimental settings[3], next to new and emerging forms based on the human computation paradigm [8, 9]. The situation in the domain of games differs significantly. In this domain, it is still by and large the task of the engineer to design the language-based interactions. As always, such a state of affairs leaves the task of finding the right term of phrase to execute the intended action to the individual user. Burdening the user with this additional task - that is orthogonal to the task of figuring out the actions themselves - leads to usability problems, for example, in the showcase scenario of interactive fiction games examined herein [5].

Interactive fiction games are essentially text-based interactive narrations, at times supplemented with graphical media - ranging from ASCII-sketches to three dimensional environments [6]. Within these narrations, the user can interact with a digital environment by entering text at pre-specified game points. As many other fields in language-based interaction with digital systems a factor limiting the flexibility of the textual interaction originates in a data-acquisition bottleneck. In the approach implemented and described herein, we examine how this particular bottleneck can be widened by employing the aforementioned human computation paradigm. More specifically, we seek to examine how explicit design choices for the experimental human computation games[1] affects the nature of the data obtained via the experiments [1].

2. EXPERIMENTAL QUESTION&SET-UP

The demonstration system described herein is part of a larger undertaking to examine how one can tailor the hc-game design to suit the specific needs of the applications, e.g. an if-game or other application that requires a specific type of language data. A central parameter thereby is the level of detail of the linguistic expressions [7] or their granularity. As most games require from the user to denote specific actions in virtual *physical spaces*, we chose this to provide a

[*]Funded by the German Research Foundation(DFG) within the Collaborative Research Centre 637.
[†]Funded by the Klaus Tschira Foundation.

KDD-HCOMP'09 June 28, 2009, Paris, France.

[1]To avoid confusion regarding references to human computation games and interactive fiction games, the term *game* will be prefaced by the corresponding abbreviations *hc-* and *if-*.

setting for examining different descriptions of spatial states of affairs [4]. Consequently, our first question posed, was how specific design choices can serve to obtain more or less specific descriptions of actions executed in a spatial scene. One way of achieving this contrastive has been proposed [8], which suggests to manipulate the context in which an object - in our case a situated action depiction - is placed within the hc-game. In order to test this in an empirical demonstration, we implemented a contextually contrastive and a contextually non-contrastive setting for a set of 36 action depictions as follows:

- in the contrastive setting, each depiction was presented together with five different action depictions - showing an identical spatial scene;
- in the non-contrastive setting, each depiction was given individually.

The objective therefore was to evaluate, if a contrastive context resulted in more specific descriptions of the situated actions then more general descriptions in the non-contrastive context. The specific game design implemented had to reflect the ensuing requirements for our controlled experimental setup, i.e.:

- in both cases, the task was to be constant - here to describe the depiction determined by the system
- in each case the context differed with regard to the contrastive feature described above
- in both cases the implemented hc-game provides sufficient motivation to play the hc-game via a scoring function afforded for each contextual setting

3. DEMONSTRATION SYSTEM

To describe the realization of the requirements specified above two so-called *LingoBox* games were implemented - one termed *Actionary* and the other *TwinMinds*. In the following, we will describe the game flow and denote where they differ respectively. Initially, two players enter the first round of the game as they log in. They proceed to a page describing the rules and allows them to select a free player or wait for one to come. Once the players are paired, they are automatically assigned roles of *Guesser* and *Describer*.

In the *Actionary* game, the player who is selected to be the *Describer* sees a set of six animated action depictions that are grouped using predefined sets. In each set, a different action is performed, e.g. taking a turn or walking back and forth within the same spatial scene. One of the action depictions is highlighted. The *Describer* can describe the highlighted animation using a text box. The entered text is passed to the *Guesser* which sees at the same time the same set of animations. The animation order, however, is random, to ensure that both players do not see the six animations in the same sequence. Receiving the description, the *Guesser's* task is to select the described animation as fast as possible. The scoring function afforded thereby is to calculated response-time graded points based for correct decisions and none for incorrect ones. Players then proceed to the next round, exchanging the roles of *Guesser* and *Describer*. Defining a maximum score of 1000 for the fastest correct response deemed possible and zero for an incorrect one, a four-fold partition was created employed to designate levels of expertise that were given to the players, e.g. top actionarists to players who reached more than 750 points.

In the *TwinMinds* game, both players are *Describer's* and see the same single animated action depiction - taken from the overall set of action depictions. Their task is to provide the same description as their partner. The resulting descriptions are scored by a special edit-distance metric, which ensures that the players get three points per matched word and one per unmatched words, which gives awards to descriptions that are more alike without punishing the effort to produce more than word. Again a four-fold level of expertise was devised as in the *Actionary* game.

4. RESULT

By the explicit design choices described above and implemented in the respective demonstration games two corpora of data were gathered. By comparing these corpora by frequency metrics based on average phrase length, the distribution of types of words and their corresponding amounts of tokens as well as by comparing the individual distribution of the parts of speech used, we found several notable results, which will be summarized as follows. The non-contrastive setting, indeed, produced more general descriptions, more tokens and fewer types in comparison to the contrastive setting. Note, that it still produced a larger variety, i.e. more types of words, then the *Actionary* game. There the contrastive context contributes to make the corpus of more detailed descriptions - less types and more tokens - also more homogeneous. Thereby, it can be demonstrated how one can start to tailor the data obtained by a hc-game, for example, with respect to the specific needs of a if-game. Lastly, in a post game study both games were ranked by the 45 players to be easy to understand and engaging to play.

5. REFERENCES

[1] Timothy Chklovski and Yolanda Gil. An analysis of knowledge collected from volunteer contributors. In *AAAI*, pages 564–571. AAAI Press / The MIT Press, 2005.

[2] Jean-Marc Francony, Evert Kuijpers, and Y. Polity. Towards a methodology for wizard of oz experiments. In *Third Conference on Applied Natural Language Processing*, March 1992.

[3] Iryna Gurevych and Robert Porzel. Empirical studies for intuitive interaction. pages 603–615. 2006.

[4] Ralf Klabunde and Robert Porzel. Tailoring spatial descriptions to the addressee: A constraint-based approach. *Linguistics*, 36(3):551–578, 1998.

[5] Michael Mateas and Andrew Stern. Procedural authorship: A case-study of the interactive drama facade. In *Digital Arts and Culture (DAC)*, 2005.

[6] Nick Montfort. *Twisty Little Passages : An Approach to Interactive Fiction*. The MIT Press, December 2003.

[7] Cécile L. Paris. Tailoring object descriptions to a user's level of expertise. *Comput. Linguist.*, 14(3):64–78, 1988.

[8] Luis van Ahn. *Human Computation*. PhD thesis, School of Computer Science, Pittsburgh, PA 15213, December 2005.

[9] Luis von Ahn. Human computation. In *K-CAP '07: Proceedings of the 4th international conference on Knowledge capture*, pages 5–6, New York, NY, USA, 2007. ACM.