Effects of Game on User Engagement with Spoken Dialogue System

Hayato Kobayashi, Kaori Tanio, Manabu Sassano, Yahoo Japan Corporation

Introduction

Background
- Making users actively utter queries is important in a spoken dialogue system.
  - There have been several studies based on gamification for addressing this problem (Gustafson et al., 2004; Bell et al., 2005; Hjalmarsson et al., 2007; Rayner et al., 2010; Rayner et al., 2012; Jurgens and Navigli, 2014).
  - However, it takes much time and effort to gamify a whole system.

Purpose
- To explore the possibilities of using of a small game module instead of gamifying a whole system.
- Can a dialogue game make users actively use the whole system?

Contribution
- Analyzed vast amounts of dialogue data, i.e., more than tens of millions of user utterances cumulated via a running app of a spoken dialogue system.
- Discovered that game plays increased the average number of utterances by about 150%, despite excluding utterances about the game.

Onsei-Assist
- Japanese spoken dialogue Android/iOS app.
  - Siri-like personal assistant developed by Yahoo Japan Corporation.
  - Produced more than 20 million of utterances within a year via pre-installs to smartphones and downloads (more than one million) in Google Play.

Word-chain game

Rules of Japanese word-chain game, Shiritori
- The principle of the word-chain game, called Shiritori in Japanese, is to say a word based on rotation so that its head character is the same character as the tail character of the previous word, e.g., (apple, eel, lip, pine, ...).
- Each player must say a word satisfying the following four conditions:
  1. The head of the word must be the same as the tail of the previous word.
  2. The word must not be a word already said in the game.
  3. The word must be a noun.
  4. The tail of the word must not end with “な(n).”

Natural dictionary by crowdsourcing
- Prepared 1,150 seed words from dozens of employees in our company by using a simple word-chain game program developed only for this purpose.
- Created a crowdsourcing task asking workers to answer an appropriate word for each seed word based on the above rule.
- Obtained a sufficient amount of words (6,148) with their frequencies after repeating the task three times.
- Extracted the top 20 words based on frequency for each of the 66 Japanese head characters in the extracted words.

Results of crowd sourcing task

<table>
<thead>
<tr>
<th>Stage</th>
<th>#Words</th>
<th>#Answers</th>
<th>#Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,403</td>
<td>3,379</td>
<td>71</td>
</tr>
<tr>
<td>2</td>
<td>2,951</td>
<td>9,314</td>
<td>826</td>
</tr>
<tr>
<td>3</td>
<td>6,148</td>
<td>25,645</td>
<td>2,285</td>
</tr>
</tbody>
</table>

- #Words: # of obtained words
- #Answers: # of user answers
- #Errors: # of answers breaking the above rules

Conclusion
- Discovered a fact that a game can help increase user engagement with a spoken dialogue system.
  - This suggests it is important to consider adding an entertaining module, such as a game, when developing a spoken dialogue system, as well as a useful module such as a route search.
  - Future research includes to examine other games such as a word association quiz games.

Log analysis

Comparison of reply rates after game’s win/lose
- Reply rate (by users) of a system response $R$ is defined as follows.
  \[
  (\text{Reply rate of } R) = \frac{\# \text{ of user replies to } R}{\# \text{ of } R \text{ in the log}}.
  \]
- Reply rates after a user won or failed (especially for saying a word already said) were 0.90 and 0.96, respectively.
  - This clearly indicates that users tend to retry to win after they failed.
- Those reply rates are quite high, considering the fact that even a question-type system response “どうしました？(What's happening?)” is 0.80.
  - This implies that the game leverages users’ natural desires for competition.

Comparison of new users (played/non-played)
- Average number of utterances over new users versus elapsed weeks.
  - Played and Non-played represent users who had played and had not played the game on the first day, respectively.
- The figure clearly indicates that Played tended to use the system more frequently than Non-played.

Comparison of existing users before/after each play
- Average number of utterances over game plays of existing users a week before and after each game play.
  - Game plays increased the average number of utterances by about 150% (24.60 → 43.61) despite the fact that we excluded utterances about the game.
  - A possible reason is that users have become more familiar with this assistant agent through playing the game. Thus they began to use non-game modules more frequently.

Average number of utterances w/o game utterances