Analyzing and Characterizing User Intent in Information-seeking Conversations

Chen Qu¹, Liu Yang¹, W. Bruce Croft¹, Johanne R. Trippas², Yongfeng Zhang³, Minghui Qiu⁴
¹ University of Massachusetts, Amherst, ² RMIT University, ³ Rutgers University, ⁴ Alibaba Group

Abstract

Research Problem: understanding and characterizing how people interact in information-seeking conversations

MSDialog:
- QA interactions between information seekers and providers
- Multi-turn technical support: dialogs on Microsoft products
- More than 2,000 QA dialogs with 10,000 utterances
- Annotated with user intent on the utterance level using crowdsourcing
- Freely available to encourage exploration of information-seeking conversation models

https://ciir.cs.umass.edu/downloads/msdialog/

Data Characterization: user intent distribution, co-occurrence, and flow patterns.

Key Properties and Comparisons

Dataset | Multi-turn | Human | Information-seeking | User intent label
---|---|---|---|---
DSTC 1-3 | ✓ | ✓ | ✓ | ✓
DSTC 4-5 | ✓ | ✓ | ✓ | ✓
Switchboard | ✓ | ✓ | ✓ | ✓
Twitter Corpus | ✓ | ✓ | ✓ | ✓
DSTC 6 (2nd Track) | ✓ | ✓ | ✓ | ✓
Ubuntu Dialog Corpus | ✓ | ✓ | ✓ | ✓
MSDialog | ✓ | ✓ | ✓ | ✓

Data Collection and Filtering

Data Collection:
- Crawled over 35,000 dialogs from Microsoft Community
- The forum is well-moderated and contains user-generated questions with high-quality answers
- The answers are provided by Microsoft staff, community moderators, article authors, and other experienced users including Microsoft Most Valuable Professionals.

Data Selection Criteria: To ensure quality and consistency, we use the following criteria:
- With 3 to 10 turns
- With 2 to 4 participants
- With at least one correct answer selected by the community
- Falls into one of the categories of Windows, Office, Bing, and Skype, which are the major categories of Microsoft products.

Data Taxonomy

<table>
<thead>
<tr>
<th>Code Label</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>OQ</td>
<td>Original Question</td>
<td>The first question by a user that initiates the QA dialog.</td>
</tr>
<tr>
<td>RQ</td>
<td>Repeat Question</td>
<td>Posters other than the user repeat a previous question.</td>
</tr>
<tr>
<td>CQ</td>
<td>Clarifying Question</td>
<td>Users or agents ask for clarification to get more details.</td>
</tr>
<tr>
<td>FD</td>
<td>Further Details</td>
<td>Users or agents provide more details.</td>
</tr>
<tr>
<td>FQ</td>
<td>Follow Up Question</td>
<td>Users ask follow up questions about relevant issues.</td>
</tr>
<tr>
<td>IR</td>
<td>Information Request</td>
<td>Agents ask for information of users.</td>
</tr>
<tr>
<td>PA</td>
<td>Potential Answer</td>
<td>A potential answer or solution provided by agents.</td>
</tr>
<tr>
<td>PF</td>
<td>Positive Feedback</td>
<td>Users provide positive feedback for working solutions.</td>
</tr>
<tr>
<td>NF</td>
<td>Negative Feedback</td>
<td>Users provide negative feedback for useless solutions.</td>
</tr>
<tr>
<td>GG</td>
<td>Greetings/Gratitude</td>
<td>Users or agents greet each others or express gratitude.</td>
</tr>
<tr>
<td>JK</td>
<td>Junk</td>
<td>There is no useful information in the post.</td>
</tr>
<tr>
<td>O</td>
<td>Others</td>
<td>Posts that cannot be categorized using other classes.</td>
</tr>
</tbody>
</table>

 Acknowledgement

This work was supported in part by the CIIR and in part by NSF grant #IIS-1419693 and NSF grant #IIS-1715095. Travel was sponsored in part by CIIR and in part by SIGIR Student Travel Grant.