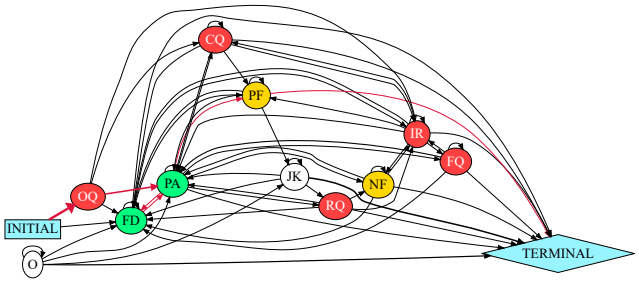

Figure 3: Flow pattern with a Markov model. Node colors: red (questions), green (answer related), yellow (feedback). Edges are directed and weighted by transition probability.

Negative Feedback tends to lead to *Potential Answer* or *Further Details*. (5) Dialogs tend to end after *Others* or *Junk*.

Besides the Markov transition graph, we use a different perspective to inspect the flow pattern by focusing on the user intent transition between turns in each dialog. We find that a quite significant flow path across turns is "*INITIAL*→*OQ*→(*PA*→*FD*)×3→*PA*→*PF*→*TERMINAL*". The "*PA*↔*FD*" circle pattern is typically caused by the "*PA+IR*", "*PA+CQ*", "*NF+FD*" co-occurrences described in Section 4.3 and the "*IR*→*FD*", "*CQ*→*FD*", "*NF*→*PA*" sequential relationship suggested in Figure 3.

4.5 Comparison with Ubuntu Dialog Corpus

Although UDC is less suitable for user intent analysis due to the informal language style, we investigate the characterizations of UDC and compare them to MSDialog since they are both in the technical support domain. We sampled 200 UDC dialogs and annotated user intent with MTurk using the same method with MSDialog. The informal language style of UDC may impact the annotation quality.

4.5.1 Statistics. For this section, we present the statistics for UDC (complete set) and MSDialog (complete set) instead of the dialogs we sampled. As shown in Table 4, UDC dialogs have shorter utterances because of the informal language style.

Table 4: Statistics of UDC & MSDialog (both complete sets)

Items	Ubuntu Dialog Corpus	MSDialog
# Dialogs	930,000	35,000
# Utterances	7,100,000	300,000
# Words (in total)	100,000,000	24,000,000
Avg. # Participants	2	3.18
Avg. # Turns Per Dialog	7.71	8.94
Avg. # Words Per Utterance	10.34	75.91

4.5.2 Data Characterization. *Potential Answer* and *Further Details* are the most significant user intent in UDC, which is consistent with MSDialog. Interestingly, the most common user intent in MSDialog, *Greetings/Gratitude*, is quite rare in UDC. In addition, we observe the exact same top 5 label co-occurrences in UDC as described in Section 4.3. Note that they are not necessarily in the same order. Finally, we found that the flow patterns observed in MSDialog also hold in UDC, except for the tendency from *Positive Feedback* to *TERMINAL*. This can be explained by the scarcity of *Positive Feedback* in UDC. Although the UDC dialogs with informal language style are drastically different from the formal written style of MSDialog, the resemblance in user intent characterizations indicates that human QA conversations, regardless of the communication medium, follow similar patterns.

5 DISCUSSION

In this section we discuss the limitation of our findings. The patterns we discovered are closely related to several design choices, including using dialogs from a well moderated forum in a specific domain. These choices were made to keep the setting as clean as possible as the research community is at an initial stage of this study. Although MSDialog does not cover every aspect of the highly diverse information-seeking conversations, it should be a first step to analyze and predict user intent in an information-seeking setting.

6 CONCLUSIONS

In this paper, we create and annotate a large multi-turn question answering data for research in conversational search. We perform in-depth characterization and analysis of this data to gain insights on the distribution, co-occurrence and flow pattern of user intent in information-seeking conversations. We will make our dataset freely available to inspire future research. Future work will consider using neural architectures for user intent prediction tasks.

ACKNOWLEDGMENTS

This work was supported in part by the Center for Intelligent Information Retrieval and in part by NSF grant #IIS-1419693 and NSF grant #IIS-1715095. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

REFERENCES

[1] N. J. Belkin, C. Cool, A. Stein, and U. Thiel. Cases, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems. 1970.
[2] S. Bhatia, P. Biyani, and P. Mitra. Classifying User Messages For Managing Web Forum Data. In *WebDB '12*, 2012.
[3] J. J. Godfrey and E. Holliman. Switchboard-1 Release 2. *Jour. of Env. Health*, 1997.
[4] M. Henderson, B. Thomson, and J. D. Williams. The Third Dialog State Tracking Challenge. In *SLT '15*, 2015.
[5] C. Hori and T. Hori. End-to-end Conversation Modeling Track in DSTC6. *arXiv:1706.07440*, 2017.
[6] S. Kim, L. F. D'Haro, R. E. Banchs, J. D Williams, M. Henderson, and K. Yoshino. The Fifth Dialog State Tracking Challenge. In *SLT '16*, 2016.
[7] A. P. Kirilenko and S. Stepchenkova. Inter-Coder Agreement in One-to-Many Classification: Fuzzy Kappa. *Plos One*, 2016.
[8] R. Lowe, N. Pow, I. Serban, and J. Pineau. The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems. In *SIGDIAL '15*, 2015.
[9] G. Marchionini. Exploratory Search: From Finding to Understanding. *Commun. ACM*, 2006.
[10] R. N. Oddy. *Information Retrieval through Man-Machine Dialogue.* 1977.
[11] F. Radlinski and N. Craswell. A Theoretical Framework for Conversational Search. In *CHIIR'17*, 2017.
[12] A. Ritter, C. Cherry, and W. B. Dolan. Unsupervised Modeling of Twitter Conversations. In *ACL '10*, 2010.
[13] A. Ritter, C. Cherry, and W. B. Dolan. Data-driven Response Generation in Social Media. In *EMNLP '11*, 2011.
[14] C. Shah and J. Pomerantz. Evaluating and Predicting Answer Quality in Community QA. In *SIGIR '10*, 2010.
[15] P. Thomas, D. McDuff, M. Czerwinski, and N. Craswell. MISC: A Data Set of Information-seeking Conversations. In *CAIR '17*, 2017.
[16] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. How Do People Interact in Conversational Speech-Only Search Tasks: A Preliminary Analysis. In *CHIIR '17*, 2017.
[17] R. W. White and R. A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm.* Synthesis Lectures on Information Concepts, Retrieval, and Services. 2009.