Data-driven Intent Models for Visual Analysis Tools and Chatbot Platforms

Vidya Setlur
Tableau Research
Palo Alto, CA 94304
vsetlur@tableau.com

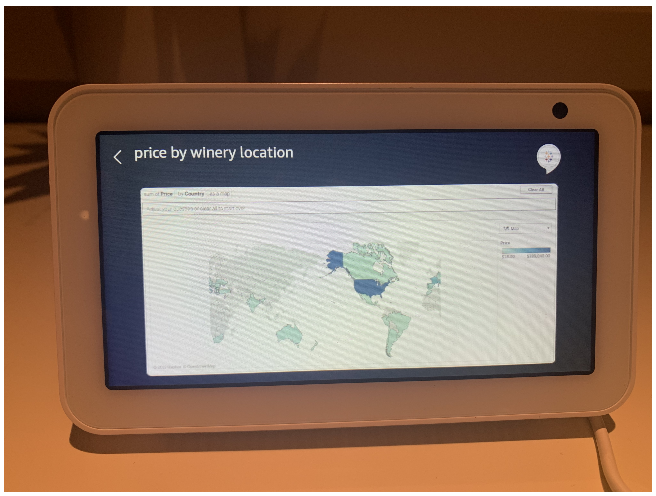

Figure 1: Natural language interaction with a wines dataset on an Echo Show device. This chatbot experience leverages deep learning techniques to transfer user intent learnings from Tableau's natural language feature, Ask Data (https://www.tableau.com/products/new-features/ask-data). Here, the user asks "show me the wine prices by location" and the chatbot responds with a map visualization.

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract

Natural language interaction in visual analysis tools supports expressive ways for users to interact with their data. Chatbot systems have recently garnered interest as conversational interfaces for a variety of tasks. Crafting the business logic for handling user intent in natural language input using a pre-defined grammar can be precise, but often covers a small set of intent models for a specific platform. More recently, machine learning approaches have shown to be promising for supporting complex responses based on the current conversational state of the interface. Such techniques could be employed for bootstrapping a range of chatbot interfaces for visual analysis. One approach is to use a labeled dataset of natural language interactions that capture user intent distribution, co-occurrence, and flow patterns. Another approach is employing deep learning techniques that approximate the heuristics and conversational cues for continuous learning in a chatbot interface. This paper explores the implications for these data-driven approaches in broadening the scope for visual analysis workflows across various chatbot experiences.

Author Keywords

Natural language interfaces; visual analysis; user intent; chatbots; data-driven approaches; machine learning.

CCS Concepts

•Human-centered computing → Information visualization; Natural language interfaces;

Motivaton

Conversational interfaces (CIs) such as smart assistant devices have become prevalent for a variety of tasks ranging from asking the weather to single turn question-and-answering for making a restaurant reservation [1, 2, 7].

In the context of visual analysis, effective tools help engage a user in the flow of analysis with a critical component being *interactivity*; rarely can a user's complex questions be answered by a single static chart. Users often interact with the current visualization and change the data display by filtering, navigating, and seeking details-on-demand. During these interactions it is critical to keep users in the flow of conversation; in other words, an *analytical conversation*. These conversations are typically referred to as information-seeking conversations, where the natural language interface provides visualization responses to a user. To build functionally intuitive CIs that can respond to more complicated analytical tasks, we need to understand how users interact in these information-seeking environments. It is necessary to develop techniques for effectively analyzing and characterize user interactions and utterance intent.

Natural language interfaces for visual analysis focus on user intent by attempting to match concepts in the input utterances with concepts known by the system. A successful user experience may involve refining the utterance and reformulating the query if the utterance is too broad, narrow, or ill-formulated. While inferencing can help provide sensible defaults, repair through follow-up utterances or employing a mixed initiative approach [4, 5], is prevalent in these systems. Most intelligent interface designs assume to one degree or another that the user will be “kept in the loop" to negotiate with the system by resolving ambiguities, making relevancy judgements, and revising utterances provided by the system [10]. These natural language interfaces are designed for a specific platform or modality, with user intent understanding constrained by the domain of the knowledge base or context in which the interaction occurs.

The promise that natural language brings to users for broadening the accessibility of visual analysis tools, has led to new entry points and proliferation into other platforms specifically around CIs. Merely repurposing how user intent is interpreted for one type of natural language interface in another, does not always lead to precise interpretation; the modalities for user repair and refinement are different (e.g. voice vs. text input). Hence, designing new natural language interfaces becomes a scalable challenge for developers. So, how can we leverage artificial intelligence (AI) techniques to broaden the reach for natural language interfaces to support visual analysis? Further, how can best practices from human-computer interaction (HCI) help AI algorithms to augment a better understanding of handling user intent?

Data-Driven Approaches for Analytical Intent

The handling of intent in analytical conversation involves the understanding of complicated visual analysis tasks with multiple turns of utterance exchange. AI techniques provide mechanisms to capture intent from conversations, operationalizing on training data, and applying them in CIs. The general premise behind these techniques is the extraction of features based on the content, structural, and sentiment characteristics of a given utterance, and apply the learnings to new contexts and visual analysis workflows. Developing intelligent conversational interfaces is still an ongoing research problem that raises many challenges in the AI and HCI communities. The discussion of prevalent state-of-the-art techniques and offering insights, can help provide effective methods of user intent prediction in analytical conversation.

Training datasets are employed in the analysis of user intent for response ranking and user intent prediction. To support natural dialogs, conversational systems should be modeled closely to human behavior; the data should come from conversation interactions with *actual* humans. The challenge with such data is ensuring that it is appropriate for user intent analysis, specifically for visual analysis turn-taking tasks. More recently, several bodies of research have focused on collecting conversation data to create training sets for natural language systems, annotated with fine-grained user intent types [8]. With the increased presence of such large-scale, annotated datasets, there are unique research opportunities for analyzing user intent distribution, co-occurrence patterns and flow patterns of large-scale analytical conversations. We need to develop techniques for gaining insights on human intent dynamics, where humans provide feedback about the system's response with critical information about how to improve the previous response [6]. The nature of ambiguity that is specific to visual analysis [9], along with the semantics of the underlying data, provide additional input to how these datasets can be leveraged for training chatbot platforms.

Deep learning approaches are programmed to learn by example and determine whether their output is accurate or not on their own. Sequence-to-sequence learning and reinforcement learning employ Recurrent Neural Networks to tackle complex sequence-to-sequence prediction problems, extract relevant information from large corpora, and perform shallow reasoning [11]. Bidirectional Encoder Representations from Transformers (BERT) has gained popularity in the application to various natural language processing tasks, including text classification for question answering [3]. Figure 1 shows an initial exploration of BERT applied to transfer user intent from telemetry data from Tableau's natural language feature *Ask Data*, to an Echo Show [1] chatbot interface.

Discussion and Research Opportunities

While these AI techniques hold promise for building conversational interfaces, they do have limitations. Based on our initial exploration in appropriating user intent to a chatbot platform, we found that many of the models are trained to address only a limited class of problems. Current approaches to predict user intent often require human input to correct system choices and are not specific to the domain of visual analysis. Using a dataset such as telemetry from an existing natural language visual analysis tool, can help with domain coverage. We also found that natural language processing techniques such as statistical parsers [12] can augment AI models to better understand the context and meaning of users' requests. The challenge is to develop methods that are generalizable and can be easily applied to various intent domains in visual analysis.

In this workshop paper, we explore data-driven approaches that could be employed to bridge the gap between complete automation and user interaction and refinement for developing natural language experiences for CIs. We hope that this topic can elicit a conversation (pun intended) on drawing strengths from *both* the AI and HCI communities with the goal of democratizing visual analysis tools through natural language.

REFERENCES

[1] Amazon. 2020. Amazon Echo. (2020). https://www.amazon.com/slp/smart-speakers/nv8sf8ertbsj8gz

[2] Apple. 2020. Apple Siri. (2020). https://support.apple.com/en-us/HT204389

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2018).

[4] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G. Karahalios. 2015. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software Technology (UIST 2015)*. ACM, New York, NY, USA, 489–500.

[5] Enamul Hoque, Vidya Setlur, Melanie Tory, and Isaac Dykeman. 2018. Applying Pragmatics Principles for Interaction with Visual Analytics. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 309–318.

[6] Jian Hu, Gang Wang, Fred Lochovsky, Jian-tao Sun, and Zheng Chen. 2009. Understanding User's Query Intent with Wikipedia. In *Proceedings of the 18th International Conference on World Wide Web (WWW '09)*. Association for Computing Machinery, New York, NY, USA, 10. DOI: http://dx.doi.org/10.1145/1526709.1526773

[7] Microsoft. 2020. Microsoft Cortana. (2020). https://www.microsoft.com/en-us/cortana

[8] Chen Qu, Liu Yang, W. Bruce Croft, Johanne R. Trippas, Yongfeng Zhang, and Minghui Qiu. 2018. Analyzing and Characterizing User Intent in Information-Seeking Conversations. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 4. DOI: http://dx.doi.org/10.1145/3209978.3210124

[9] Vidya Setlur, Melanie Tory, and Alex Djalali. 2019. Inferencing Underspecified Natural Language Utterances in Visual Analysis. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. Association for Computing Machinery, New York, NY, USA, 12. DOI: http://dx.doi.org/10.1145/3301275.3302270

[10] Ben Shneiderman. 1997. Direct Manipulation for Comprehensible, Predictable and Controllable User Interfaces. In *Proc.2nd Intl Conf. Intelligent User Interfaces (IUI '97)*. ACM, New York, NY, USA, 33–39. DOI:http://dx.doi.org/10.1145/238218.238281

[11] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 3104–3112. http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf

[12] Min Tang, Xiaoqiang Luo, and Salim Roukos. 2002. Active Learning for Statistical Natural Language Parsing. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02)*. Association for Computational Linguistics, USA, 8. DOI:http://dx.doi.org/10.3115/1073083.1073105