Flipping the Script: Inverse Information Seeking Dialogues for Market Research

Josh Seltzer
Nexxt Intelligence, Toronto, Canada

Kathy Cheng
Nexxt Intelligence, Toronto, Canada

Shi Zong
Nexxt Intelligence, Toronto, Canada
University of Waterloo
Waterloo, Canada

Jimmy Lin
University of Waterloo
Waterloo, Canada

ABSTRACT

Information retrieval has traditionally been framed in terms of searching and extracting information from mostly static resources. Interactive information retrieval (IIR) has widened the scope, with interactive dialogues largely playing the role of clarifying (i.e., making explicit, and/or refining) the information search space. Informed by market research practices, we seek to reframe IIR as a process of eliciting novel information from human interlocutors, with a chatbot-inspired virtual agent playing the role of an interviewer. This reframing flips conventional IIR into what we call an *inverse information seeking dialogue*, wherein the virtual agent recurrently extracts information from human utterances and poses questions intended to elicit related information. In this work, we introduce and provide a formal definition of an inverse information seeking agent, outline some of its unique challenges, and propose our novel framework to tackle this problem based on techniques from natural language processing (NLP) and IIR.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**.

KEYWORDS

Market Research, Inverse Information Seeking, Virtual Agent

ACM Reference Format:
Josh Seltzer, Kathy Cheng, Shi Zong, and Jimmy Lin. 2022. Flipping the Script: Inverse Information Seeking Dialogues for Market Research. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3477495.3536326

1 INTRODUCTION

Within the market research industry, researchers seek to understand consumers' perceptions and experiences through a variety of methodologies, most commonly dichotomized into *qualitative research* and *quantitative research*. Focus groups, in-depth interviews and ethnography are common forms of qualitative research that are exploratory in nature, with the goal of establishing baseline assumptions. Quantitative research is predominantly undertaken in the form of surveys, where close-ended questions prevail with the goal of quantifying and validating baseline assumptions.

While quantitative research provides statistical confidence, it lacks the flexibility and opportunity for unanticipated discoveries that characterizes qualitative research, and vice versa qualitative research affords more in-depth understanding which may not generalize to other consumers. With a growing emphasis on human-centered innovation, businesses are under increasing pressure to uncover quantifiable consumer-led insights quickly so as to drive the next new product, service, or experience [2, 9, 20].

Uncovering such insights necessitates a hybrid approach that can benefit from both qualitative and quantitative research, i.e., qualitative insights at scale. To achieve this, a technological breakthrough is required to overcome the current onerous process of qualitative research — one that involves significant human resources (including professionally trained moderators, recruiters, and coordinators), time (typically a minimum of 4 weeks for a project with 4 focus groups/32 participants), and cost (on average $25K for a project of aforementioned scale, compared to $15K for a survey with 500 participants[1]).

In this work, we propose to use a virtual agent to address the above business challenges. The virtual agent is designed to imitate the qualitative research process of in-depth interviews conducted by human moderators: qualitative researchers trained to build rapport with research participants, to ask good questions which delve deep into consumers' hearts and minds, and to distill a large amount of unstructured data into insights that can benefit marketers. Through the virtual agent, we are striving to considerably minimize the resources required to collect and analyze qualitative insights, and consequently the capacity to scale up the number of participants.

2 INVERSE INFORMATION SEEKING

2.1 Background

Interacting with humans is considered an essential component of IIR systems, which are designed to facilitate information seeking through methods such as search, clustering, question answering,

[1]Costs cited are in USD, and are intended for illustrative purposes only. Actual project costs vary by market, participant profile, etc.

SIGIR '22, July 11–15, 2022, Madrid, Spain

ACM ISBN 978-1-4503-8732-3/22/07…$15.00
https://doi.org/10.1145/3477495.3536326

and information extraction [5]. Methods from NLP are systematically related to, and widely employed, in information retrieval [8], with some prior work even focusing on agents which conversationally interact with humans in order to better understand their information seeking intention [28]. Despite a large literature dedicated to this area of research, inverse information seeking — the task of a virtual agent eliciting information from a human interlocutor — is largely unexplored. Moreover, most work related to question asking and answering has been limited to questions which seek factoid-type information [6].

In this work, we aim at using a virtual agent to proactively mine opinions from users through agent-driven conversations. There are two primary goals of this virtual agent: (1) to extract structured information representing opinions expressed over multiple turns of dialogue; and (2) to generate follow-up and clarification questions with the objective of eliciting details which extend the breadth and depth of the system's knowledge about the human's opinion. These goals are constrained by the specific research objectives of a market researcher for any particular project; although the nature of qualitative research is exploratory, there are often specific topics and aspects which a researcher wants to focus the conversation on.

2.2 Problem Definition

Our problem of an inverse information seeking dialogue consists of information exchanges between a virtual agent and a human. Formally, during each information exchange time step t, the agent poses a question q_t to which the human responds with an utterance u_t. Then the task of inverse information seeking is to recurrently generate a followup question q_{t+1} and to construct an *opinion graph* G_t, which incorporates the latest question/utterance pair (q_t, u_t) on top of the previous opinion graph G_{t-1}:

$$\mathcal{A} = (G_{t-1}, q_t, u_t) \rightarrow (G_t, q_{t+1}). \tag{1}$$

The opinion graph represents the aggregated information deduced from a dialogue, and is defined by $G = (N, E)$. N is a set of opinion nodes and an edge $(n_i \rightarrow n_j) \in E$ represents an explanation relationship such that the opinion node n_j is explained (at least in part) by the opinion node n_i. *Opinion nodes* are constructed as pairs $n = (a, I)$, where information slots $i \in I$ are extracted from question/utterance pairs referring to an aspect a. For example, if the agent asks a question "Do you like the shoes?" to which the human replies "yes, very much", then the resulting opinion node will consist of a = "the shoes" and I = {affect : "likes very much"}.

Unlike related work in opinion modeling [4], the information slots do not necessarily need to correspond exactly to an utterance substring, since the information may be deduced based on multiple turns of dialogue (as in the above example).

2.3 Evaluation Metrics

The performance of the virtual agent on the inverse information seeking task can be evaluated through a variety of metrics. With the aim of evaluating both its component parts as well as the holistic ability of the agent to mimic a human moderator, we focus on three primary evaluation metrics:

(1) Opinion graphs updated at each time step are evaluated using the F-measure against a test set of ground truth annotations, made by a human market researcher on the same information exchange pair (q_t, u_t) and prior opinion graph G_{t-1} [21].
(2) Questions generated at each time step are evaluated via human rating of the generated question in terms of its perceived specificity and sensibleness to the conversation's context, yielding a sensibleness and specificity average (SSA) [1].
(3) The extent of information elicited over the course of a conversation is evaluated by a human market researcher in terms of its SSA with respect to the research objective.

3 CHALLENGES

3.1 Constructing Opinion Graphs on the Fly

Representing opinions and constructing opinion graphs from text is a common task in opinion mining [4], however comparatively little work focuses on incorporating discourse structures as part of opinion analysis [25]. Unlike prior work in opinion mining, information elicited by a virtual agent may be expressed over the course of multiple utterances, and may sometimes only be deduced rather than extracted from utterances. For example, it is common for a human participant to qualify a previous utterance or to correct a perceived misunderstanding by the agent. In such cases, an utterance like "no I didn't mean that" has the intentional act of contradicting the agent's prior representation of the human's opinion.

Similarly, expressions of opinions may be discontinuous with respect to the mention of the aspect being opinionated on; or the aspect may even be left implicit within the dialogue. For example, after being shown a video advertisement, the utterance "I couldn't read it!" implicitly refers to a text within the video. It is thus imperative to keep track of the current subject being discussed across multiple turns of dialogue, and where necessary to trigger clarification questions which are intended to disambiguate any uncertainty in the generated opinion graph.

3.2 Generating Elicitation Questions

As discussed, the virtual agent's goal is to elicit information pertaining to peoples' opinions and experiences through the constraints of a research project's objective. In practice this means that the moderator needs to be sufficiently open-domain so as to explore unanticipated aspects and opinions; but at the same time sufficiently closed-domain so as to ensure that the information elicited is relevant to the research objectives. This task can be accomplished by alternating between *depth questions*, which are intended to elicit more information about an aspect which has already been mentioned; and *breadth questions*, which are intended to elicit new opinions about other aspects directly related to the research objectives.

The dilemma of choosing between breadth versus depth search is known as the breadth/depth trade-off [19]. Although heuristics exist for choosing breadth versus depth decisions optimally based on search capacity, in this task the optimization is further constrained by the uncertainty of the payoffs: more detailed information about an unexplored opinion may have diminishing returns, and conversely the number of aspects which someone is willing to share their opinion on may be exhausted.

In addition to navigating the breadth/depth trade-off, the virtual agent is further tasked with asking questions which are most likely to elicit relevant information. Trivially, a moderator can always

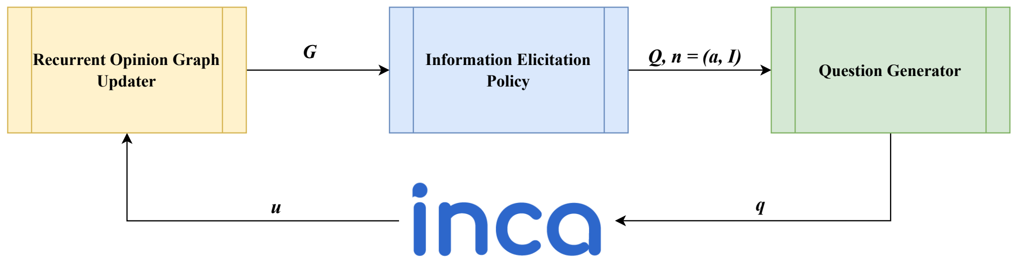

Figure 1: Our Inquisitive Natural Conversation Agent (INCA). A user utterance u triggers the updating of the opinion graph G, which is used to determine the next action as indicated by the question recipe Q and the relevant opinion node n containing an aspect a and informational slots I. Finally, the question generator produces a question q which INCA displays to the user.

attempt to elicit new information by asking a generic depth question such as "Can you tell me more about that?", or else a generic breadth question such as "Are there other things you'd like to tell me?". However, asking questions which are specific to the prior dialogue are likely to be the most successful in eliciting useful information, with the caveat that non-sensible and irrelevant questions should be avoided. Thus, there is the added challenge of generating semantically and syntactically sensible questions which refer specifically to prior utterances.

4 FRAMEWORK FOR INVERSE INFORMATION SEEKING

In this section, we present our **I**nquisitive **N**atural **C**onversation **A**gent (INCA) framework. The framework addresses above challenges through the coordination of three modules: (1) a recurrent opinion graph updater; (2) an information elicitation policy; and (3) a question generator. The overview of this framework is in Figure 1.

4.1 Recurrent Opinion Graph Updater

To address the challenge of constructing opinion graphs recurrently based on the question and utterance pair at each time step, we propose a system which represents opinions as nodes defined by the aspect being opinionated, coupled with corresponding information slots. Our system classifies each question according to the modifications to the opinion graph (*update rules*) which would result from the corresponding utterance response. In particular, we consider the aspects mentioned and the intended dialogue act within each utterance, and update the opinion graph based on the question's update rules. This approach allows for interpretable consequences of any question posed by the virtual agent, such that the dialogue act and aspects extracted from each utterance response will deterministically update the opinion graph.

Three variables are required to determine the update: (1) the update rules associated with the posed question, which we take as prespecified rules designed by a human (see Section 4.3); (2) a dialogue act predicted from the utterance; and (3) any aspects or corresponding information extracted from the utterance content. Dialogue acts are defined as the meaning of any utterance with respect to its illocutionary force, and have been used widely in conversational analysis to classify the actions intended by interlocutors [3, 17]. We adapt existing methods for detecting dialogue acts in utterances using deep learning models [11, 13]. Since each opinion node is uniquely defined by an aspect, this module employs methods from opinion aspect extraction and named entity recognition [14, 18], and adapts named entity matching [16] so as to group novel articulations of an entity with one that has already been added to the opinion graph. Aspect-based sentiment analysis models which classify aspects according to affective content [15, 27] are used to fill information slots related to each aspect, whenever affective valuations of aspects are made explicit in an utterance. Whenever novel aspects are mentioned in an utterance, we adapt existing work on inferring explanatory relations, which can be represented as edges in the opinion graph [4].

4.2 Information Elicitation Policy

The optimal type of information to seek depends crucially on both the research objectives as well as the opinions expressed so far within a conversation. To strike a balance between these needs, we use a rule-based dialog flow management system which determines what type of question to ask next based on the state of the conversation thus far, as well as controls set by a human researcher. For example, a human researcher may wish to constrain the number of aspects to collect information about; or to specify the target aspects which are of specific interest and therefore will be prioritized above other aspects; or the types of information slots of most importance.

This module takes the current opinion graph as an input, and outputs a target aspect a and the label of a *question recipe* Q, which specifies the type of information which the system will attempt to elicit as well as localized question template strings, which is subsequently used by the question generator. Future improvements to this module could adapt the Transformer Embedding Dialogue (TED) policy introduced by the Rasa team [26].

4.3 Question Generator

Given the aspect a and a question recipe Q determined by the information elicitation policy, the question generator finally generates a natural language question q_{t+1} that will be posed in the next information exchange.

Although large language models such as GPT-2 can be fine-tuned to generate non-templated questions based on training data [23], the risk of generating potentially toxic or otherwise undesirable questions is unacceptably high considering the business requirements. Instead, we opt for using templated-based methods, i.e., by generating questions from templates which specify the type of information to be filled by each slot, particular to each question recipe. Although slot filling on template strings has been used to generate conversational utterances [24], such method in practice can result in grammatical or coreferential errors, particularly when slots are filled not just with aspect names but also whole phrases from an utterance.

To address these limitations, we design a question generator which employs a pretrained language model to perform slot filling on a template question. Although language models pretrained using a masking objective have been shown to perform well on coreference resolution tasks [10] and grammatical error correction [12], it is insufficient when faced with complex tasks such as cross-domain reasoning [22]. We hypothesize that prudential design of question templates can lead to the generation of sensible and fluent natural language questions using pretrained language models. Furthermore, the use of multilingual pretrained language models [7] with multilingual question recipes presents the opportunity for our question generator to generalize across many languages.

5 CONCLUSION

In this work, we present a virtual agent which asks questions to human participants, intended to facilitate the active elicitation of information which is relevant to a market researcher's ad-hoc objectives. We first formulate our problem of inverse information seeking as a twofold task: the recurrent construction of an opinion graph and the generation of elicitation questions. We then discuss the unique challenges of making the virtual agent emulate professionally trained researchers, including accurately representing a participant's expressed opinions over the course of a dialogue, and determining both the aspects of interest as well as the most apt question to elicit further details. Our proposed framework, `INCA`, addresses above challenges through a novel combination of techniques from NLP and IIR. It is structured as a modular feedback system which first updates the opinion graph representation of the dialogue, then determines the next information elicitation action based on constraints provided by a human researcher, and finally generates a question using templates which are filled using a pretrained language model, before starting again with the participant's next utterance. By situating this problem as one of inverse information seeking, we hope to contribute toward a new vision of consumer-led market research, and to motivate future work which aims to uncover the nuanced and rich details of human opinions and experiences.

REFERENCES

[1] Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. 2021. Towards a Human-like Open-Domain Chatbot. *arXiv preprint arXiv:2001.09977* (2021).

[2] Abdullah Al Mamun, Muhammad Mohiuddin, Syed Ali Fazal, and Ghazali Bin Ahmad. 2018. Effect of Entrepreneurial and Market Orientation on Consumer Engagement and Performance of Manufacturing SMEs. *Management Research Review* (2018).

[3] John Langshaw Austin. 1975. *How To Do Things With Words.* Oxford University Press.

[4] Nofar Carmeli, Xiaolan Wang, Yoshihiko Suhara, Stefanos Angelidis, Yuliang Li, Jinfeng Li, and Wang-Chiew Tan. 2021. Constructing Explainable Opinion Graphs from Reviews. In *Proceedings of the Web Conference 2021.* 3419–3431.

[5] W. Bruce Croft. 2019. The Importance of Interaction for Information Retrieval. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 1–2.

[6] Pradeep Dasigi, Kyle Lo, Iz Beltagy, Arman Cohan, Noah A. Smith, and Matt Gardner. 2021. A Dataset of Information-Seeking Questions and Answers Anchored in Research Papers. *arXiv preprint arXiv:2105.03011* (2021).

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers).* Minneapolis, Minnesota, 4171–4186.

[8] Jens Dörpinghaus, Johannes Darms, and Marc Jacobs. 2018. What Was the Question? A Systematization of Information Retrieval and NLP Problems. In *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems (FedCSIS).* 471–478.

[9] Maria Vittoria Franceschelli, Gabriele Santoro, and Elena Candelo. 2018. Business Model Innovation for Sustainability: A Food Start-up Case Study. *British Food Journal* (2018).

[10] Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. SpanBERT: Improving Pre-training by Representing and Predicting Spans. *Transactions of the Association for Computational Linguistics* 8 (2020), 64–77.

[11] Daniel Jurafsky, Rebecca Bates, Noah Coccaro, Rachel Martin, Marie Meteer, Klaus Ries, Elizabeth Shriberg, Andreas Stolcke, Paul Taylor, and Carol Van Ess-Dykema. 1997. Automatic Detection of Discourse Structure for Speech Recognition and Understanding. In *Proceedings of the 1997 IEEE Workshop on Automatic Speech Recognition and Understanding.* 88–95.

[12] Masahiro Kaneko, Masato Mita, Shun Kiyono, Jun Suzuki, and Kentaro Inui. 2020. Encoder-Decoder Models Can Benefit From Pre-trained Masked Language Models in Grammatical Error Correction. *arXiv preprint arXiv:2005.00987* (2020).

[13] Hamed Khanpour, Nishitha Guntakandla, and Rodney Nielsen. 2016. Dialogue Act Classification in Domain-Independent Conversations Using a Deep Recurrent Neural Network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers.* 2012–2021.

[14] Joseph Lark, Emmanuel Morin, and Sebastián Peña Saldarriaga. 2017. A Comparative Study of Target-based and Entity-based Opinion Extraction. In *International Conference on Computational Linguistics and Intelligent Text Processing.* Springer, 211–223.

[15] Yuliang Li, Aaron Xixuan Feng, Jinfeng Li, Saran Mumick, Alon Halevy, Vivian Li, and Wang-Chiew Tan. 2019. Subjective Databases. *arXiv preprint arXiv:1902.09661* (2019).

[16] Yuliang Li, Jinfeng Li, Yoshihiko Suhara, AnHai Doan, and Wang-Chiew Tan. 2020. Deep Entity Matching with Pre-trained Language Models. *arXiv preprint arXiv:2004.00584* (2020).

[17] Kecheng Liu and Weizi Li. 2014. *Organisational Semiotics for Business Informatics.* Routledge.

[18] Deon Mai and Wei Emma Zhang. 2020. Aspect Extraction Using Coreference Resolution and Unsupervised Filtering. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop.* 124–129.

[19] Rubén Moreno-Bote, Jorge Ramírez-Ruiz, Jan Drugowitsch, and Benjamin Y. Hayden. 2020. Heuristics and Optimal Solutions to the Breadth–Depth Dilemma. *Proceedings of the National Academy of Sciences* 117, 33 (2020), 19799–19808.

[20] N Nuruzzaman, Deeksha Singh, and Chinmay Pattnaik. 2019. Competing to be Innovative: Foreign Competition and Imitative Innovation of Emerging Economy Firms. *International Business Review* 28, 5 (2019), 101490.

[21] Heiko Paulheim. 2017. Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. *Semantic Web* 8, 3 (2017), 489–508.

[22] Dheeraj Rajagopal, Vivek Khetan, Bogdan Sacaleanu, Anatole Gershman, Andrew Fano, and Eduard Hovy. 2021. Cross-Domain Reasoning via Template Filling. *arXiv preprint arXiv:2111.00539* (2021).

[23] Ivan Sekulić, Mohammad Aliannejadi, and Fabio Crestani. 2021. Towards Facet-Driven Generation of Clarifying Questions for Conversational Search. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval.* 167–175.

[24] Ming-Hsiang Su, Chung-Hsien Wu, and Liang-Yu Chen. 2019. Attention-Based Response Generation Using Parallel Double Q-learning for Dialog Policy Decision in a Conversational System. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2019), 131–143.

[25] Shiliang Sun, Chen Luo, and Junyu Chen. 2017. A Review of Natural Language Processing Techniques for Opinion Mining Systems. *Information Fusion* 36 (2017), 10–25.

[26] Vladimir Vlasov, Johannes E. M. Mosig, and Alan Nichol. 2019. Dialogue Transformers. *arXiv preprint arXiv:1910.00486* (2019).

[27] Albert Weichselbraun, Stefan Gindl, Fabian Fischer, Svitlana Vakulenko, and Arno Scharl. 2017. Aspect-based Extraction and Analysis of Affective Knowledge from Social Media Streams. *IEEE Intelligent Systems* 32, 3 (2017), 80–88.

[28] Hamed Zamani, Susan Dumais, Nick Craswell, Paul Bennett, and Gord Lueck. 2020. Generating Clarifying Questions for Information Retrieval. In *Proceedings of The Web Conference 2020.* 418–428.