The CAMEO Project: A Holistic View for Conversational Agents

Discussion Paper

Tommaso Di Noia[1], Guglielmo Faggioli[2], Marco Ferrante[2], Nicola Ferro[2], Fedelucio Narducci[1], Raffaele Perego[3] and Giuseppe Santucci[4]

[1]*Polytechnic University of Bari, Italy*
[2]*University of Padua, Italy*
[3]*National Research Council, Italy*
[4]*Sapienza University of Rome, Italy*

Abstract

The interaction with information access systems through conversational interfaces is becoming increasingly popular. Among the systems that benefit the most from this kind of interaction between the user and the system, we mention *Information Retrieval (IR)* systems and *Recommender Systems (RS)*. A conversational interface in such scenarios has multiple advantages: first, it allows users to refine their needs, either in terms of information in an IR system or of an item in a RS, through multiple interactions also based on the system's reaction to the user's prompt. Secondly, it allows elderly, children and visually impaired people to access these kinds of systems easily. Nevertheless, conversational systems come with a plethora of additional challenges that need to be addressed, including a more complex querying language and a challenging evaluation – for which we often lack also evaluation data. Finally, Conversational IR and RS are often intended as separate tasks, with separate models and systems. We argue that such tasks could benefit from an integrated model capable of seamlessly dealing with them and exploiting the joint knowledge to improve its effectiveness. CAMEO is a project that aims to deal with such challenges. In this extended abstract, we outline the current state of the works within the CAMEO project and detail some future directions that we wish to explore.

1. Introduction

Conversational interactions are an effective solution to reach a larger group of users when it comes to information access systems. Indeed, a conversational interface is an accessible way to interact with an RS or an IR system. Furthermore, it alleviates the mental encumbering of formulating a textual query for all users, who can use natural language sentences to interact with the system. This allows refining the information need over multiple utterances when the user interacts with a *Conversational Search (CS)* but also correcting and directing a *Conversational Recommendation (CR)* towards the optimal items through a conversation with the system itself. At the same time, conversational systems, used for both CR and CS, are affected by some major challenges that require additional effort to obtain good-performing systems. First of all, the interaction between the user and the system, typically easier for the user as they can use natural language, is harder for the system. The system has to deal with common dialog constructs,

IIR 2024: 14th Italian Workshop on Information Retrieval, September 5–6, 2024, Udine, Italy

such as anaphoras, ellipses, and coreferences, which are complex to be handled automatically by a machine. Secondly, the system needs to have an internal knowledge of the state of the conversation to provide the most effective answer. Finally, the evaluation of these systems is particularly challenging for several reasons. First, we still lack rich and extensive evaluation collections. Secondly, we lack proper evaluation measures and paradigms. Indeed, most of the current evaluation approaches [1, 2, 3] rely on procedures similar to those used in classic IR and RS and do not keep into account the structure of the conversation, nor take into consideration the fact that different users might interact in different ways with the system [4, 5, 6]. At the same time, in the current state of the art linked to CS and CR, these tasks are dealt with orthogonality: two completely different systems are often used to address them. We argue that an integrated approach able to provide additional information to the user through a CS component while also recommending the most suited products through a CR module would improve the satisfaction of the users. Furthermore, past efforts in the joint recommendation and search [7, 8, 9] have highlighted the positive effects that joint modelling can have on both CS and CR tasks.

The "*Conversational Agents: Mastering, Evaluating, Optimizing (CAMEO)*" project[1] [10] addresses the abovementioned limitations while fostering the development of conversational agents designed for joint CS and CR. CAMEO is being developed under the *Progetti di Rilevante Interesse Nazionale* (PRIN) framework and involves partners from four institutions: The National Research Council (CNR), the Polytechnic University of Bari (POLIBA), the Sapienza University of Rome and the University of Padua (UNIPD, coordinator).

2. The CAMEO Project

The core of CAMEO will include the following components: the *dialog manager*, responsible for the course of the dialog, the CS and CR engines. The *dialog manager* takes the user's input, processes it, and forwards it to the RS and IR engines. It is also in charge of post-processing the system's answer, considering the current state of the conversation, to output a natural language sentence as an answer. The *conversational RS* and *conversational IR* engines receive the input from the dialog manager and identify respectively items and documents that might satisfy the user's need. The joint nature of the CAMEO envisioned system is obtained through shared *internal knowledge*. This final component of the system interacts with and is updated by all the other elements of CAMEO. Ideally, it would contain information on the current state of the conversation as well as some form of description of the user, to provide personalized recommendations and more effective search results. Furthermore, within CAMEO, we plan to develop a visual analytics environment that can support researchers in making decisions on how to tweak the system and adapt it to improve its performance.

We will deploy CAMEO system for the conversational product search task. As a possible example of interactions with the system, assume the user is interested in a certain product of which they have limited knowledge. In this case, the user might start with a generic question, such as "What types of this product exist?": the CS component will then obtain documents satisfying this information need. The conversation continues with the user learning more and more. The system collects details on the user's interests and preferences based on their

[1] https://cameo.dei.unipd.it

questions and reactions to the system's answers. When the system has enough information, the CR component activates and starts suggesting products that might be relevant.

3. Future Work and Research paths

We describe here some research paths that we will investigate in the next phases of the project.
Dimension Importance Estimation Framework. Faggioli et al. [11] recently defined the so-called Dimension Importance Estimation (DIME) task. They observed that, given a neural encoder $\phi : \{V\}^* \rightarrow \mathbb{R}^d$ that projects strings $s \in \{V\}^*$ (i.e., queries and documents), in a d-dimensional latent space, it is possible to determine a subset of query-dependent important dimensions $\mathcal{D}$ with $|\mathcal{D}| < d$ such that if $\phi_{\mathcal{D}}$ is used (i.e., an encoder identical to ϕ except for that dimensions not in $\mathcal{D}$ are removed), the effectiveness of a dense IR system increases. Faggioli et al. [11] experimented exclusively with ad-hoc retrieval. We argue that both the conversational framework as well as RS can benefit from DIME models: for example, the state of the conversation and the previous utterances can be used in the CS task to identify which dimensions are the most important to retrieve relevant documents given the current utterance. At the same time, the relevance information provided by the user's feedback in the RS domain can provide a useful signal to guide dimension selection and the identification of the optimal $\mathcal{D}$.
Pareto-optimal solutions in Search and Recommendation. Many IR and RS tasks, including those in conversational scenarios, are moving from computing a ranking of final results based on a single metric to multi-objective problems where the metrics to optimize are multiple and often model contrasting goals. Solving these problems leads to a set of Pareto-optimal solutions, known as the Pareto frontier, in which no objective can be further improved without hurting others. Paparella et al. [12] propose a novel, post-hoc, theoretically-justified technique, named "Population Distance from Utopia" (PDU) to select one—best—Pareto-optimal solution among the ones lying in the Pareto frontier in search and recommendation scenarios. PDU is the only selection technique in the literature that can be "calibrated", i.e., it can choose the best Pareto-optimal solution based on ideal targets expressed on single queries or users. The extensive experimental evaluation presented in [12] focuses on IR and RS scenarios and shows that PDU's formulation and calibration feature notably impacts the solution's selection. This work is still ongoing with the formulation of a new loss function based on the PDU derivation to train effective multi-objective ranking models directly. The conversational framework devised in CAMEO will be an exemplary scenario for integrating and evaluating our PDU solution.
Answer Generation for Conversational Agents. Large Language Models (LLMs) have gained noteworthy importance and attention across different domains and fields in recent years. IR and RS are surely among the domains they impacted the most, as witnessed by the recent increase in the number of IR and RS solutions incorporating generative models. Retrieval Augmented Generation (RAG) is an emerging paradigm particularly interesting for CAMEO as it integrates existing knowledge from large-scale document corpora into the generation process, enabling the model to generate more coherent, contextually relevant, and accurate text across various tasks, including IR and RS dialogue systems. Recent studies have highlighted the significant positional dependence exhibited by RAG systems. Such studies observed how the placement of information within the LLM input prompt drastically affects the generated output. In a recent

paper, [13], Alessio et. al. focused on this property by investigating alternative strategies for ordering sentences within the LLM prompt to improve the average quality of the generated responses in the user and conversational system dialogues. The proposed end-to-end RAG architecture focuses on a conversational assistant use case and is empirically evaluated using the TREC CAsT 2022 collection. Our experiments highlight significant differences between distinct arrangement strategies. By employing an evaluation methodology based on RANKVICUNA, Alessio et. al. show that their best approach achieves improvements up to 54% in terms of overall conversational response quality over baseline methods.

Finding the correct answer to a given question is the primary goal of a conversational agent. Although most conversational agents perform this task well, some intrinsic natural language issues cannot be solved by only analyzing the submitted question. The definition of a well-disambiguated request through a natural language question is not a trivial task. Indeed, it requires the usage of specific terms plus a cognitive effort that is not affordable for everyone. As stated by Biancofiore et al. [14], in a high-level perspective, a conversational agent mainly deals with two classes of interaction: Disambiguation and Exploration.

Exploration. After the system returns the answer $\hat{a}$ to an initial query q, a new set of queries Q' can be suggested by the system or posed by the user to explore relevant topics related to $\hat{a}$.

Disambiguation. In the case that there are *too many* or *too few* eligible answers for a given question, or the request is ambiguous, the system can ask a new question to the user.

In a conversational spectrum ranging from exploration to disambiguation for IR and RS, CS is placed on the exploration side, while CR is on the disambiguation side. In a holistic view, we need to balance the two interaction strategies to define an integrated CS-CR system.

Visual Analytics-based Evaluation of Conversational Agents. The evaluation of CS and CR is a challenging task, and CAMEO will investigate using Visual Analytics solutions to help the system designers evaluate and improve the system. The research path will investigate and integrate three different strategies. The first approach will rely on the definition and collection of concersation-related metrics. The second one will define a model capturing the state of the conversation and its temporal evolution, and the ordering of the sentences. The third one will investigate the possibility of adapting models and visual solutions from explainable artificial intelligence (XAI), see, e.g., [15]. The models and the data produced by these three research strategies will be used to design a Visual Analytics system supporting the design and evaluation of both single CS and CR and their integration.

4. Conclusion

In this work, we described our envisioned holistic conversational search and recommendation agent that will be developed within the CAMEO research project framework. Furthermore, we have detailed how a set of recent approaches can be employed in developing such a holistic agent. More in detail, such approaches rely on estimating the importance of each dimension in an embedding space, working on the optimization aspect to select the Pareto-optimal solution, or permuting documents to improve the quality of the response in a RAG scenario. Finally, we described some visualization strategies that we intend to use to help the designers of the system evaluate its performance and tune it according to their needs.

Acknowledgments

This work was partially supported by CAMEO, PRIN 2022 n. 2022ZLL7MW.

References

[1] V. Stamatis, L. Azzopardi, A. Wilson, VES team at TREC conversational assistance track (cast) 2019, in: E. M. Voorhees, A. Ellis (Eds.), Proceedings of the Twenty-Eighth Text REtrieval Conference, TREC 2019, Gaithersburg, Maryland, USA, November 13-15, 2019, volume 1250 of *NIST Special Publication*, National Institute of Standards and Technology (NIST), 2019. URL: https://trec.nist.gov/pubs/trec28/papers/VES.C.pdf.

[2] J. Dalton, C. Xiong, J. Callan, Cast 2020: The conversational assistance track overview, in: E. M. Voorhees, A. Ellis (Eds.), Proceedings of the Twenty-Ninth Text REtrieval Conference, TREC 2020, Virtual Event [Gaithersburg, Maryland, USA], November 16-20, 2020, volume 1266 of *NIST Special Publication*, National Institute of Standards and Technology (NIST), 2020. URL: https://trec.nist.gov/pubs/trec29/papers/OVERVIEW.C.pdf.

[3] J. Dalton, C. Xiong, J. Callan, TREC cast 2021: The conversational assistance track overview, in: I. Soboroff, A. Ellis (Eds.), Proceedings of the Thirtieth Text REtrieval Conference, TREC 2021, online, November 15-19, 2021, volume 500-335 of *NIST Special Publication*, National Institute of Standards and Technology (NIST), 2021. URL: https://trec.nist.gov/pubs/trec30/papers/Overview-CAsT.pdf.

[4] G. Faggioli, M. Ferrante, N. Ferro, R. Perego, N. Tonellotto, Hierarchical dependence-aware evaluation measures for conversational search, in: F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai (Eds.), SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021, ACM, 2021, pp. 1935–1939. URL: https://doi.org/10.1145/3404835.3463090. doi:10.1145/3404835.3463090.

[5] G. Faggioli, M. Ferrante, N. Ferro, R. Perego, N. Tonellotto, A dependency-aware utterances permutation strategy to improve conversational evaluation, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), Advances in Information Retrieval - 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10-14, 2022, Proceedings, Part I, volume 13185 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 184–198. URL: https://doi.org/10.1007/978-3-030-99736-6_13. doi:10.1007/978-3-030-99736-6_13.

[6] T. Di Noia, F. M. Donini, D. Jannach, F. Narducci, C. Pomo, Conversational recommendation: Theoretical model and complexity analysis, Inf. Sci. 614 (2022) 325–347. URL: https://doi.org/10.1016/j.ins.2022.07.169. doi:10.1016/J.INS.2022.07.169.

[7] K. Zhao, Y. Zheng, T. Zhuang, X. Li, X. Zeng, Joint learning of e-commerce search and recommendation with a unified graph neural network, in: K. S. Candan, H. Liu, L. Akoglu, X. L. Dong, J. Tang (Eds.), WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining, Virtual Event / Tempe, AZ, USA, February 21 - 25, 2022, ACM, 2022, pp. 1461–1469. URL: https://doi.org/10.1145/3488560.3498414. doi:10.1145/3488560.3498414.

[8] T. Thonet, J. Renders, M. Choi, J. Kim, Joint personalized search and recommendation with hypergraph convolutional networks, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), Advances in Information Retrieval - 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10-14, 2022, Proceedings, Part I, volume 13185 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 443–456. URL: https://doi.org/10.1007/978-3-030-99736-6_30. doi:10.1007/978-3-030-99736-6_30.

[9] Z. Si, Z. Sun, X. Zhang, J. Xu, X. Zang, Y. Song, K. Gai, J. Wen, When search meets recommendation: Learning disentangled search representation for recommendation, in: H. Chen, W. E. Duh, H. Huang, M. P. Kato, J. Mothe, B. Poblete (Eds.), Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023, ACM, 2023, pp. 1313–1323. URL: https://doi.org/10.1145/3539618.3591786. doi:10.1145/3539618.3591786.

[10] T. Di Noia, G. Faggioli, M. Ferrante, N. Ferro, F. Narducci, R. Perego, G. Santucci, Cameo: Fostering joint conversational search and recommendation, in: SEBD 2024: Symposium on Advanced Database Systems, June 23–26, 2024, Villasimius, Sardinia, Italy, CEUR Workshop Proceedings, CEUR-WS.org, 2024.

[11] G. Faggioli, N. Ferro, R. Perego, N. Tonellotto, Dimension importance estimation for dense information retrieval, in: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July 14–18, 2024, Washington, DC, USA, ACM, 2024. URL: https://doi.org/10.1145/3626772.3657691. doi:10.1145/3626772.3657691.

[12] V. Paparella, V. W. Anelli, F. M. Nardini, R. Perego, T. Di Noia, Post-hoc selection of pareto-optimal solutions in search and recommendation, in: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 2013–2023. URL: https://doi.org/10.1145/3583780.3615010. doi:10.1145/3583780.3615010.

[13] M. Alessio, G. Faggioli, N. Ferro, F. M. Nardini, R. Perego, Improving rag systems via sentence clustering and reordering, in: RAG @ SIGIR 2024 workshop: The Information Retrieval's Role in RAG Systems, ACM, 2024, pp. 1–10. URL: toAppear.

[14] G. M. Biancofiore, Y. Deldjoo, T. Di Noia, E. Di Sciascio, F. Narducci, Interactive question answering systems: Literature review, ACM Comput. Surv. 56 (2024) 239:1–239:38. URL: https://doi.org/10.1145/3657631. doi:10.1145/3657631.

[15] B. La Rosa, G. Blasilli, R. Bourqui, D. Auber, G. Santucci, R. Capobianco, E. Bertini, R. Giot, M. Angelini, State of the art of visual analytics for explainable deep learning (2023).