UvA-DARE (Digital Academic Repository)

Towards Building Economic Models of Conversational Search

Azzopardi, L.; Aliannejadi, M.; Kanoulas, E.

Citation for published version (APA):

Towards Building Economic Models of Conversational Search

Leif Azzopardi[1(✉)], Mohammad Aliannejadi[2], and Evangelos Kanoulas[2]

[1] University of Strathclyde, Glasgow, UK
leif.azzopardi@strath.ac.uk
[2] University of Amsterdam, Amsterdam, NL, The Netherlands
{m.aliannejadi,e.kanoulas}@uva.nl

Abstract. Various conceptual and descriptive models of conversational search have been proposed in the literature – while useful, they do not provide insights into how interaction between the agent and user would change in response to the costs and benefits of the different interactions. In this paper, we develop two economic models of conversational search based on patterns previously observed during conversational search sessions, which we refer to as: *Feedback First* where the agent asks clarifying questions then presents results, and *Feedback After* where the agent presents results, and then asks follow up questions. Our models show that the amount of feedback given/requested depends on its efficiency at improving the initial or subsequent query and the relative cost of providing said feedback. This theoretical framework for conversational search provides a number of insights that can be used to guide and inform the development of conversational search agents. However, empirical work is needed to estimate the parameters in order to make predictions specific to a given conversational search setting.

1 Introduction

Conversational Search is an emerging area of research that aims to couch the information seeking process within a conversational format [5,11] – whereby the system and the user interact through a dialogue, rather than the traditional query-response paradigm [16,17]. Much like interactive search, conversational search presents the opportunity for the system to ask for feedback, either through clarifying questions or follow up questions in order to refine or progress the search [4,9,10]. While there have been numerous studies trying to develop methods to improve the query clarifications or follow up questions, and to better rank/select results to present/use during a conversational search (e.g. [2,3,12–14,18]) – less attention has been paid to modelling conversational search and understanding the trade-offs between querying, assessing, and requesting feedback [8]. In this paper, we take as a reference point the work of Vakulenko et al. [15] and the work of Azzopardi [6,7]. The former is an empirically derived model of conversational search, called **QRFA**, which involves querying (Q), receiving/requesting feedback (RF) and assessing (A), while the latter works

are economic models of querying (Q) and assessing (A). In this paper, we consider two ways in which we can extend the economic model of search to include feedback. And then we use these models to better understand the relationships and trade-offs between querying, giving/requesting feedback and assessing given their relative efficiencies and their relative costs. We do so by analysing what the optimal course of interaction would be given the different models in order to minimise the cost of the conversation while maximising the gain. As a result, this work provides a number of theoretical insights that can be used to guide and inform the development of conversational search agents.

2   Background

In their analysis of conversational search sessions, Vakulenko et al. [15] found that two common conversational patterns emerged:

1. the user issues a (Q), the system would respond by requesting feedback (RF), the user would provide the said feedback, and the system would either continue to ask further rounds of feedback (RF), or present the results where user assess items (A), and then repeats the process by issuing a new query (or stops); or,
2. the user issues a (Q), the system would present results where the user assesses items (A), and then, the system requests feedback given the results (RF), the user would provide the said feedback, and the system would present more results (A), the user would assess more items (A), and the system would request further feedback (RF) until the user issues a new query (or stops).

Inspired by our previous work [1], we can think of these two conversational patterns as **Model 1**: Feedback First, and **Model 2**: Feedback After. There, of course, are many other possible patterns i.e. feedback before and after, or not at all, and combinations of. In this work, we shall focus on modelling these two "*pure*" approaches for conversational search. But, before doing so, we first present the original economic model of search.

An Economic Model of Search: In [6], Azzopardi proposed a model of search focused on modelling the traditional query response paradigm, where a user issues a query (Q), the system presents a list of results where the user assesses A items. The user continues by issuing a new query (or stops). We can call this **Model 0**. During the process, the user issues Q queries and is assumed to assess A items per query (on average). The gain that the user received was modelled using a Cobbs-Douglas production function (see Eq. 1). The exponents α and β denote the relative efficiency of querying and assessing. If $\alpha = \beta = 1$ then it suggests that for every item assessed per query, the user would receive one unit of gain (i.e. an ideal system). However, in practice, α and β are less than one, and so the more querying or more assessing naturally leads to diminishing returns. That is, assessing one more item is less likely to yield as much gain as the previous item, and similarly issuing another query will retrieve less new information (as the pool of relevant items is being reduced for a given topic).

$$g_0(Q_0, A_0) = Q_0^{\alpha}.A_0^{\beta} \tag{1}$$

Given A_0 and Q_0, a simple cost model was proposed (Eq. 2), where the total cost to the user is proportional to the number of queries issued and the total number of items inspected (i.e. $Q_0 \times A_0$), where the cost per query is C_q and the cost per assessment is C_a.

$$c_0(Q_0, A_0) = Q_0.C_q + Q_0.A_0.C_a \tag{2}$$

By framing the problem as an optimisation problem [7], where the user wants a certain amount of gain, then the number of assessments per query that would minimise the overall cost to the user is given by:

$$A_0^{\star} = \frac{\beta.C_q}{(\alpha - \beta).C_a} \tag{3}$$

Here, we can see that as the cost of querying increases, then, on average, users should assess more items. While if the cost of assessing increases, on average, users should assess fewer items per query. If the relative efficiency of querying (α) increases, then users should assess fewer items per query, while if the relative efficiency of assessing (β) increases then the users should assess more. Moreover, there is a natural trade-off between querying and assessing, such that to obtain a given level of gain $g(Q_0, A_0) = G$ then as the A_0 increases, Q_0 decreases, and vice versa. In the following sections, we look at how we can extend this model to also include the two different forms of feedback (first and after).

3 Models

Below we present two possible models for conversational search given the two pure strategies of feedback first and feedback after. In discussing the models, we will employ a technique called *comparatives statics* – that is we make statements regarding the changes to the outcomes (i.e. how users would change their behaviour) given a change in a particular variable assuming all other variables remain the same.

3.1 An Economic Model of Conversational Search - Feedback First

Under Model 1 (Feedback First), each round of feedback, aims to improve the initial query – and thus drive up the efficiency of querying. To model this, we need to introduce a function, called $\Gamma(F_1)$ where the efficiency of queries is directly related to how many rounds of feedback are given. We assume for simplicity that the relationship is linear, each round of feedback increases the efficiency of querying by a fixed amount, which we denote as γ_1. We can then set: $\Gamma(F_1) = \gamma_1.F_1 + \alpha$. If no rounds of feedback are given/requested, then $\Gamma(F_1) = \alpha$ – which results in the original gain function. The gain function under feedback first is:

$$g_1(Q_1, F_1, A_1) = Q_1^{\Gamma(F_1)}.A_1^{\beta} \tag{4}$$

Since the rounds of feedback are only given per query, then the cost model becomes:

$$c_1(Q_1, F_1, A_1) = Q_1.C_q + Q_1.F_1.C_f + Q_1.A_1.C_a \qquad (5)$$

Again, we can frame the problem as an optimisation problem, where we would like to minimise the cost given a desired level of gain G. Then by using a Lagrangian multiplier, differentiating and solving the equations we are at the following expressions:

$$A_1^\star = \frac{\beta.C_q + F_1.C_f}{(\gamma_1.F_1 + \alpha - \beta).C_a} \qquad (6)$$

$$F_1^\star = \frac{\beta.C_q + (\alpha - \beta)A_1.C_a}{\gamma_1.A_1.C_a + \beta.C_f} \qquad (7)$$

From these expressions, we can see that there is a dependency between the two main actions of assessing and providing feedback (unfortunately, we could not reduce down the expression any further). With respect to the optimal number of assessments, we can see that when $F = 0$ the model reverts back to the original model. And similarly, if C_q increases, then $A_1^\star$ increases. If C_a increases, $A_1^\star$ decreases. And if C_f increases and $F_1 > 0$, then $A_1^\star$ increases. If the number of rounds of feedback increases, then the model stipulates that a user should inspect fewer items (as the query after feedback is more efficient – and so more effort should be invested into assessing). For feedback, we see that if C_q or C_a increases, then the user should provide more feedback, or alternatively the system should request more feedback (i.e. $F_1^\star$ increases). While if the cost of feedback increases then users should provide less feedback (or the system should request less feedback before returning results). Interestingly, if γ_1 increases, then fewer rounds of feedback are required. This makes sense because if the clarifying question improves the quality of the original query sufficiently, then there is no need for further clarifications of the information need. More clarifications would only increase the cost of the conversation but not necessarily increase the amount of gain. On the other hand, if γ_1 is very small, then it suggests that the clarifying questions are not increasing the quality of the query such that it retrieves more relevant material. Thus, asking clarifying questions, in this case, is unlikely to be worthwhile as it will drive up the cost of conversation without increasing the reward by very much.

3.2 An Economic Model of Conversational Search - Feedback After

Given Model 2 (Feedback After), we can extended the original economic model of search by adding in feedback (F_2), such that a user issues a query, assess A_2 items, then provides feedback to refine the query, followed by assessing another A items. The process of giving feedback and assessing, then repeats this F_2 times. For this kind of conversational interaction, we can define the gain to be proportional to the number of queries and the number of rounds of feedback per query and the number of assessments per query or round of feedback:

$$g_2(Q_2, F_2, A_2) = Q_2^\alpha.(1 + F_2)^{\gamma_2}.A_2^\beta \qquad (8)$$

As before, α and β denote the efficiency of querying and assessing, while γ_2 expresses the efficiency of providing feedback. We can see that if $F = 0$ (e.g. no feedback), then the model reverts back to the original model in Eq. 1.

$$c_2(Q_2, F_2, A_2) = Q_2.C_q + Q_2.F_2.C_f + Q_2.(1 + F_2).A_2.C_a \qquad (9)$$

The cost function is also extended to include the rounds of feedback, and the additional assessments per feedback round, where if $F_2 = 0$ then the cost model reverts back to the original cost model. The amount of feedback per query will depend on the cost of the feedback, and its relative efficiency (γ_2). So if $\gamma_2 = 0$ and $F_2 > 0$ then there is no additional benefit for providing feedback – only added cost.

To determine the optimal number of assessments and feedback, for a given level of gain G where we want to minimise the total cost to the user given G we formulated the problem as an optimisation problem. Then by using a Lagrangian multiplier, differentiating and solving the equations we are at the following expressions for the optimal number of assessments ($A^\star$), and the optimal number of rounds of feedback ($F_2^\star$) :

$$A_2^\star = \frac{\gamma.(C_q + C_f) - \alpha.(1 + F_2).C_f}{(\alpha - \gamma).(1 + F_2).C_a} \qquad (10)$$

$$F_2^\star = \frac{(\gamma - \beta).C_q + (\beta - \alpha).C_f}{(\alpha - \gamma).C_f} \qquad (11)$$

First, we can see that under this model, it is possible to solve the equations fully. While this may seem quite different – during the intermediate steps the optimal $A_2^\star$ was:

$$A_2^\star = \frac{\beta.(C_q + F_2.C_f)}{(\alpha - \beta).(F_2 + 1)C_a} \qquad (12)$$

where we can see the relationship between assessing and giving feedback. And, if we set $F_2 = 0$, then the model, again, falls back to the original model in Eq. 1. Specifically, assuming feedback is to be given/requested (i.e. $F_2 > 0$), then if the cost of performing the feedback (C_f) increases, on average, a user should examine more items per query and per round of feedback. If the cost of assessing (C_a) increases, then, on average, a user should examine fewer items per query/feedback. Intuitively, this makes sense, as it suggests a user should invest more in refining their need to bring back a richer set of results, than inspecting additional results for the current query or round of feedback. Now, if the relative efficiency of assessing (β_2) increases, then a user should examine more items per query/feedback.

In terms of feedback, we can see that as the cost of querying (C_q) increases, then it motivates giving more feedback. While if the cost of feedback (C_f) increases, it warrants providing less feedback. This is because querying is a natural alternative to providing another round of feedback. We can also see that the relative efficiencies of querying (α_2) to feedback (γ_2) also play a role in determining the optimal amount of feedback – such that as γ_2 increases they users should give more feedback, while if α_2 increase they should query more.

4 Discussion and Future Work

In this paper, we have proposed two economic models of conversational search that encode the observed conversational patterns of *feedback first* and *feedback after* given the QFRA work. While these models represent two possible conversational strategies, they do, however, provide a number of interesting observations and hypotheses regarding conversational search paradigms and the role of feedback during conversational sessions. We do, of course, acknowledge that in practice conversational sessions are likely to be more varied. Nonetheless, the insights are still applicable. Firstly, and intuitively, if the cost of giving feedback either before/first or after increases, then the number of rounds of feedback will be fewer. While if the cost of querying increases, then it motivates/warrants requesting/giving more feedback. The amount of which depends on the relative costs and efficiencies of each action. However, a key difference arises between whether feedback is given before (first) or after. Under feedback first, if the relative efficiency γ_1 (e.g. answering clarifying questions, etc.) increases, then perhaps ironically less feedback is required. This is because the initial query will be enhanced quicker than when γ_1 is low. Moreover, if the relative efficiency of querying is initially high, then it also suggests that little (perhaps even no) feedback would be required because the query is sufficiently good to begin with and clarifications or elaborations will only result in increased costs. These are important points to consider when designing and developing a conversational search system. As the decision to give/request feedback is decided by both the gains and the costs involved (i.e. is it economically viable?). And thus both need to be considered when evaluating conversational agents and strategies.

Of course, such discussions are purely theoretical. More analysis is required, both computationally through simulations and empirically through experiments to explore and test these models in practice. With grounded user data, it will be possible to estimate the different parameters of the proposed models – to see whether they provide a reasonable fit and valid predictions in real settings. It is also worth noting that another limitation of these models is that they model the average conversational search process over a population of user sessions – rather than an individual's conversational search process. For example, the quality of clarifying questions asked during the feedback first model is likely to vary depending on the question, this, in turn, suggests that each question will result in different improvements to the original query (i.e. γ_1 is not fixed, but is drawn from a distribution). However, the model is still informative, because we can consider what would happen for different values of γ_1 and determine when feedback would be viable, and at what point it would not be. Once estimates of the costs and relative efficiencies are obtained for a given setting, it will also be possible to further reason about how or what in the conversational process needs to be improved. Finally, more sophisticated models of conversational search could also be further developed to analyse different possible mixed strategies. However, we leave such the empirical investigations and further modelling for future work.

References

1. Aliannejadi, M., Azzopardi, L., Zamani, H., Kanoulas, E., Thomas, P., Craswell, N.: Analysing mixed initiatives and search strategies during conversational search. In: CIKM, pp. 16–26. ACM (2021)
2. Aliannejadi, M., Kiseleva, J., Chuklin, A., Dalton, J., Burtsev, M.S.: Convai3: generating clarifying questions for open-domain dialogue systems (ClariQ). CoRR abs/2009.11352 (2020)
3. Aliannejadi, M., Zamani, H., Crestani, F., Croft, W.B.: Asking clarifying questions in open-domain information-seeking conversations. In: SIGIR, pp. 475–484. ACM (2019)
4. Allen, J., Guinn, C.I., Horvtz, E.: Mixed-initiative interaction. IEEE Intell. Syst. Their Appl. **14**(5), 14–23 (1999)
5. Anand, A., Cavedon, L., Joho, H., Sanderson, M., Stein, B.: Conversational search (dagstuhl seminar 19461). Dagstuhl Rep. **9**(11), 34–83 (2019)
6. Azzopardi, L.: The economics in interactive information retrieval. In: SIGIR, pp. 15–24. ACM (2011)
7. Azzopardi, L.: Modelling interaction with economic models of search. In: SIGIR, pp. 3–12. ACM (2014)
8. Azzopardi, L., Dubiel, M., Halvey, M., Dalton, J.: Conceptualizing agent-human interactions during the conversational search process. The Second International Workshop on Conversational Approaches to Information Retrieval, CAIR (2018)
9. Belkin, N.J., Cool, C., Stein, A., Thiel, U.: Cases, scripts, and information-seeking strategies: on the design of interactive information retrieval systems. Expert Syst. Appl. **9**(3), 379–395 (1995)
10. Croft, W.B., Thompson, R.H.: I^{3}r: a new approach to the design of document retrieval systems. JASIS **38**(6), 389–404 (1987)
11. Culpepper, J.S., Diaz, F., Smucker, M.D.: Research frontiers in information retrieval: report from the third strategic workshop on information retrieval in Lorne (SWIRL 2018). SIGIR Forum **52**(1), 34–90 (2018)
12. Hashemi, H., Zamani, H., Croft, W.B.: Guided transformer: leveraging multiple external sources for representation learning in conversational search. In: SIGIR, pp. 1131–1140. ACM (2020)
13. Kiesel, J., Bahrami, A., Stein, B., Anand, A., Hagen, M.: Toward voice query clarification. In: SIGIR, pp. 1257–1260 (2018)
14. Rao, S., Daumé, H.: Learning to ask good questions: ranking clarification questions using neural expected value of perfect information. In: ACL (1), pp. 2736–2745 (2018)
15. Vakulenko, S., Revoredo, K., Di Ciccio, C., de Rijke, M.: QRFA: a data-driven model of information-seeking dialogues. In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (eds.) ECIR 2019. LNCS, vol. 11437, pp. 541–557. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15712-8_35
16. Vtyurina, A., Savenkov, D., Agichtein, E., Clarke, C.L.A.: Exploring conversational search with humans, assistants, and wizards. In: CHI Extended Abstracts, pp. 2187–2193 (2017)

17. Yan, R., Song, Y., Wu, H.: Learning to respond with deep neural networks for retrieval-based human-computer conversation system. In: SIGIR, pp. 55–64 (2016)
18. Zamani, H., Dumais, S.T., Craswell, N., Bennett, P.N., Lueck, G.: Generating clarifying questions for information retrieval. In: WWW, pp. 418–428. ACM / IW3C2 (2020)