

Graef, Roland; Klier, Mathias; Kluge, Kilian; Zolitschka, Jan Felix

**Article — Published Version**

## Human-machine collaboration in online customer service – a long-term feedback-based approach

Electronic Markets

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Graef, Roland; Klier, Mathias; Kluge, Kilian; Zolitschka, Jan Felix (2020) : Human-machine collaboration in online customer service – a long-term feedback-based approach, Electronic Markets, ISSN 1422-8890, Springer, Berlin, Heidelberg, Vol. 31, Iss. 2, pp. 319-341, <https://doi.org/10.1007/s12525-020-00420-9>

This Version is available at:

<https://hdl.handle.net/10419/289148>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Human-machine collaboration in online customer service – a long-term feedback-based approach

Roland Graef<sup>1</sup> · Mathias Klier<sup>1</sup> · Kilian Kluge<sup>1</sup> · Jan Felix Zolitschka<sup>1</sup>

Received: 1 May 2019 / Accepted: 2 April 2020 / Published online: 6 May 2020  
© The Author(s) 2020

## Abstract

The rising expectations of customers have considerably contributed to the need for automated approaches supporting employees in online customer service. Since automated approaches still struggle to meet the challenge to fully grasp the semantics of texts, hybrid approaches combining the complementary strengths of human and artificial intelligence show great potential for assisting employees. While research in Case-Based Reasoning (CBR) already provides well-established approaches, they do not fully exploit the potential of CBR as hybrid intelligence. Against this background, we follow a design-oriented approach and develop an adapted textual CBR cycle that integrates employees' feedback on semantic similarity, which is collected during the Reuse phase, into the Retrieve phase by means of long-term feedback methods from information retrieval. Using a real-world data set, we demonstrate the practical applicability and evaluate our approach regarding performance in online customer service. Our novel approach surpasses human-based, machine-based, and hybrid approaches in terms of effectiveness due to a refined retrieval of semantically similar customer problems. It is further favorable regarding efficiency, reducing the average time required to solve a customer problem.

**Keywords** Human-machine collaboration · Online customer service · Textual case-based reasoning · Long-term feedback

## Introduction

Today's customers expect to be able to contact a company via e-mail, chat, and social media platforms, all while demanding ever shorter response times (Microsoft 2018; Salesforce Research 2016). For example, a study by Zendesk (2017) found that while in 2013 62% of the surveyed customers

expected a response to an e-mail within half a day, in 2016 this number had risen to 79%. Further, 64% of customers expected companies to respond to and interact with them in real-time (Salesforce Research 2016). In addition, customers are about five times more likely to view real-time messaging as important versus unimportant (Salesforce Research 2018). Thus, consumers' preferred channel matched directly with the method they thought was fastest (Gladly 2018). In 2018, half of the customers were disappointed with machine-based customer service such as chatbots, based primarily on the need for fast and yet personal service. As a consequence, companies face the challenge of meeting customers' demand for both a short response time and a high level of service quality (Forrester 2018; Mero 2018). Hybrid approaches combining the complementary strengths of human and artificial intelligence show great potential for deployment in online customer service where an advanced text comprehension is required to fulfill customers' needs. While humans have superior capabilities for empathic, complex, or intuitive tasks such as the semantic understanding of texts, artificial intelligence is particularly good at consistently solving repetitive or routine tasks in a specific area where fast processing of huge amounts of data is required (Dellermann et al. 2018; Forrester 2016; Guzmán and Pathania 2016). Besides, human involvement

---

Responsible Editor: Philipp Alexander Ebel

This article is part of the Topical Collection on *Hybrid Intelligence in Business Networks*

---

✉ Mathias Klier  
mathias.klier@uni-ulm.de

Roland Graef  
roland.graef@uni-ulm.de

Kilian Kluge  
kilian.kluge@uni-ulm.de

Jan Felix Zolitschka  
jan.zolitschka@uni-ulm.de

<sup>1</sup> Institute of Technology and Process Management, University of Ulm, Helmholtzstraße 22, 89081 Ulm, Germany

in responding to customer requests is favorable. Many customers, while increasingly preferring digital channels and demanding fast response times, nevertheless prefer to receive their information from a person rather than from a computer (Mero 2018; Parature 2014; Salesforce Research 2016). For instance, a study by Parature (2014) found that 60% of customers chose to interact with a human representative over a self-service system.

Case-Based Reasoning (CBR) is a promising means to augment human judgments by assisting employees in online customer service (Acorn and Walden 1992; Bedué et al. 2018; Heras et al. 2009; Lenz and Burkhard 1997; Lenz et al. 1999; Lenz et al. 1998b). CBR is often used to retrieve past and already solved customer problems – so-called cases – similar to the one currently encountered, allowing employees to draw information from past textual communication and reuse respective solutions. In this regard, CBR constitutes a methodology based on human-machine collaboration. On the one hand, CBR provides employees with past cases suitable to solve new customer problems. On the other hand, by solving new customer problems employees constantly increase the number of cases to draw from. The potential of CBR as a hybrid intelligence approach where humans and machines act as teammates however has not yet been fully exploited. Particularly in approaches focusing on textual information (Burke et al. 1997; El-Sappagh and Elmogy 2015; Lenz et al. 1998b) the human capability to understand and judge the semantic relationship between potential solutions and the currently encountered problem has not been tapped to enhance case retrieval.

To further the goal of exploiting human-machine collaboration regarding case retrieval, we follow a design-oriented approach (Hevner et al. 2004; Peffers et al. 2007) and propose a novel long-term feedback-based approach to retrieve semantically similar cases. To this end, our study focuses on incorporating feedback from employees in the long term to infer the semantic relation of texts. To this end, we assume a conventional textual CBR approach as the starting point. In a first step, we collect unary feedback regarding the semantic similarity of customer problems from employees. In the context of semantic similarity, unary feedback is a sensible choice for a rating scale, as it provides a single, unambiguous rating and clearly distinguishes the fundamentally different implications of feedback “not semantically similar” and “semantically similar”. Second, we reuse and link the collected feedback in order to instantiate a machine-learning model based on methods from the area of information retrieval that is able to generate the semantic context of a customer problem. Third, we use this semantic context to enhance the case retrieval for new customer problems by creating an adapted customer problem as the weighted combination of the new customer problem and its semantic context. This way, we take advantage of the human capability to judge the semantics of solutions and the machine’s ability to comprehensively learn from

employees’ input. We demonstrate the applicability and the capabilities of our hybrid intelligence approach by using publicly available open-domain customer problems from the popular service website Quora. Our contribution is twofold: First, our approach builds upon research in CBR dealing with textual information and extends this line of thought by integrating long-term feedback following established approaches from the field of information retrieval. Second, it fosters human-machine collaboration, using human-generated feedback to improve a computer system.

Guided by the Design Science Research (DSR) process by Peffers et al. (2007), the remainder of this paper is organized as follows (cf. Figure 1): In the next section, we describe and illustrate the problem context that motivates our research. Subsequently, we present the prior state of relevant research in the areas of CBR and information retrieval and conclude this section with the research gap. Following a description of our research method, we detail the design of our long-term feedback-based approach. In the subsequent section, we demonstrate our novel approach’s practical applicability using a publicly available real-world data set of a popular service website. Then, we conduct a summative evaluation based on a standard metric and compare the approach’s performance to competing artifacts from the literature. Our paper concludes with a discussion of implications and limitations, identification of possible future research opportunities, and a summary of our findings.

## Problem context

The starting point of our problem-centered research (Peffers et al. 2007) is the observation that an increasing fraction of customer service interactions is handled through online channels (Forrester 2018; Microsoft 2018). Indeed, online customer service has become ubiquitous across industries (e.g. in the areas of information technology or telecommunication (Dell) or telecommunication (AT&T)<sup>1</sup>) with most customers expecting to be able to reach a company by e-mail, chat, social media, or via specialized online platforms (Altitude and Spider Marketing 2016; Microsoft 2018). Accordingly, customer service employees face the challenge of providing correct, reliable, as well as consistent solutions to a huge amount of customer problems in written form within a short time frame.

The following example illustrates our problem context: Consider the online customer service of a telecom company that is contacted by a customer having an issue concerning their phone and expecting immediate support. As in virtually all of the most prominent online customer service channels, the customer sends a request (*customer problem*) as free-text in form of a question: “My phone does not turn on anymore.

<sup>1</sup> Dell: <https://www.dell.com/community/>; AT&T: <https://forums.att.com/>

What should I do?”. The incoming customer problem is answered by an employee who as a domain expert relies on their individual knowledge, experience, and available domain-specific information to respond with troubleshooting instructions in time. The free-text *solution* sent as the response together with the incoming customer problem constitutes a *case*. Over time a customer service department accumulates a large number of cases that comprise the *case base*.

Since many of the incoming customer problems are not unique but have been solved previously, there is often a past case in the case base that contains a semantically similar customer problem – generally with syntactical differences – and a solution that can serve as a basis for the solution to the newly incoming problem. In the example of a faulty phone, the solution might comprise standard troubleshooting instructions applicable in many related scenarios. Hence, by reusing solutions, the knowledge contained in the case base can be used to both reduce response times and increase the quality and consistency of solutions. The key challenge is to find a suitable solution among the cases in the case base in a fast and reliable way. Due to the unparalleled speed at which they can search and process large amounts of data, computer systems are prime candidates for this task. However, the large and inconsistent vocabulary as well as missing or only implicitly contained information in free-text customer problems poses a major problem for traditional approaches. Considering the examples in Table 1, a second customer problem “My phone’s screen stays dark and it does not seem to start up” (Problem B) is semantically similar to the first (Problem A) in the sense that the given and the desired information are similar despite a very different description of the issue. Hence the solution from the previous case (Solution A) can be reused (Solution B). In contrast, “I dropped my phone and now it does not turn on anymore” (Problem C) and “I dropped my phone in the sink and now it does not turn on anymore” (Problem D) are similar in terms of vocabulary and phrasing, but refer to different types of damage (mechanical damage vs. water damage), likely demanding different solutions (Solution C, Solution D). Understanding the semantics of text is a task humans excel at. It therefore appears likely that a solution which combines the respective strengths in form of human-computer collaboration yields a performance superior to that of a computer system or an entirely manual approach alone. In this context, we distinguish the performance dimensions efficiency and effectiveness. We define efficiency as the time and effort required by employees to solve a customer problem. In contrast, we define effectiveness as the ability to provide customers with a textual solution that contains the requested knowledge.

## Related work and research gap

Prior to designing a solution to the problem of combining the strengths of humans and computers in the context of online

customer service, its objectives need to be defined (Peffer et al. 2007). In the following, informed by the literature in the areas of CBR and information retrieval, we identify the targeted research gap and state the desired properties and functionality of our approach.

## Related work in textual CBR and information retrieval

Irrespective of a specific application domain, all approaches for online customer service automatically providing employees with similar cases and learning from human knowledge in terms of already solved customer problems are based on CBR (Acorn and Walden 1992; Bedué et al. 2018; Heras et al. 2009; Lenz et al. 1999; Lenz and Burkhard 1997; Lenz et al. 1998b). Thus, the well-established CBR methodology paves the way for a hybrid intelligence where humans and machines act as teammates (Gu et al. 2017; Martin et al. 2017; Reuss et al. 2015). As customer interactions in online customer service are generally based on textual messages, particularly the research stream of textual CBR seems to provide promising approaches to cope with the task of enabling hybrid intelligence.

Textual CBR approaches are based on the CBR cycle introduced by Aamodt and Plaza (1994): A case base contains all existing and solved cases  $c_j$ , each consisting of a textual description of the customer problem  $p_j$  and a solution  $s_j$  (Burke et al. 1997; Cunningham et al. 2004; Lenz et al. 1998b; Wang et al. 2006b). Case retrieval starts with an incoming customer problem  $p_i$  and aims to quantify the degree of resemblance between  $p_i$  and all customer problems  $p_j$  of existing cases  $c_j$  in the case base by means of a similarity function  $\text{sim}(p_i, p_j)$  (Liao et al. 1998). Subsequently, the  $k$  most similar cases are presented to the employee, who can choose to reuse one or

**Table 1** Case structure and illustrative examples

	Case A	Case B	Case C	Case D
<b>Customer Problem</b>	<i>My phone does not turn on anymore. What should I do?</i>	<i>My phone's screen stays dark and it does not seem to start up.</i>	<i>I dropped my phone and now it does not turn on anymore.</i>	<i>I dropped my phone in the sink and it does not turn on anymore.</i>
<b>Solution</b>	<i>Dear customer, to find out why your phone does not work anymore, please ...</i>	<i>Dear customer, to find out why your phone does not work anymore, please ...</i>	<i>Dear Customer, ... to assess the severity of the mechanical damage, please ...</i>	<i>Dear Customer, ... to immediately dry your phone, please ...</i>



more solutions  $s_j$ , revise them if necessary, and finally add the new case  $c_i$  (incoming customer problem  $p_i$  and corresponding new solution  $s_i$ ) to the case base (Lenz et al. 1999; Wang et al. 2006b, 2011). For a more detailed review of textual CBR we refer to Appendix I (section “Supporting online customer service with Case-Based Reasoning”).

While most existing textual CBR approaches are promising examples of human-machine collaboration, their retrieval results are primarily based on the information contained in the customer problem. In contrast to machines, a human reader has multiple skills that are challenging to attain for automated approaches. First of all, humans are able to understand and interpret rhetoric or linguistic specificities (e.g. irony or sarcasm). Moreover, content written by humans often contains indirectly stated information, such as social background, level of education, or age of the author, all of which could hardly be identified without human life experiences. Further, despite very different (similar) vocabulary, two customer problems could refer to quite similar (different) types of problems (cf. Table 1). Dealing with understanding the semantics of texts is a task humans excel at. Human judgment has shown beneficial in enhancing and guiding a computer system (Salton and Buckley 1990; Sarwar, Foley, & Allan, 2018; Trstenjak and Donko 2016). Hence, to fulfill customers’ needs in online customer service an understanding of the semantic meaning in textual messages exchanged between customers and employees is important to provide a correct solution concerning customers’ problems. As a result, many approaches take employees or users into account to validate automated solutions for customer problems (Balakrishnan, Ahmadi, & Ravana, 2016; Kunze and Hübner 1998; Lenz et al. 1999; Weis 2013). However, the potential in CBR as hybrid intelligence where humans and machines collaborate on equal terms harbors potential for improvement.

To exploit this potential, some authors in the research area of CBR have started to introduce approaches which incorporate feedback from the system user into the retrieval of new cases (Branting 2001; Cheng and Hüllermeier 2008; Coyle and Cunningham 2003; Gabel and Stahl 2004; Leake and Dial 2008; Soh and Blank 2008; Stahl 2005; Stahl and Gabel 2006; Zhang and Yang 1999). Feedback approaches offer the great opportunity to enhance a CBR approach during its operation (Stahl 2003; Weis 2013).

While a CBR approach could be trained ex-ante by domain experts linking semantically similar cases, from an economic perspective this would be a waste of resources for two reasons. First, a CBR approach can already greatly support employees to some extent right from the start, even with a small case base and in the absence of feedback. Second, employees have to be released from work to label the past cases required to instantiate the CBR approach. In contrast, feedback approaches leverage synergies, since employees already profit from a CBR approach while providing feedback.

Nevertheless, only few researchers take feedback into account when developing textual CBR approaches (Balakrishnan et al. 2016; Daniels and Rissland 1997; Weis 2013). Some authors (Balakrishnan et al. 2016; Weis 2013) use feedback from system users resulting in a re-ranking of previously retrieved cases. However, re-ranking approaches discard potentially relevant cases that have not been returned by the initial retrieval. Thus, a re-ranking approach does not seem suitable to find new cases within similar contexts as the incoming customer problem. In contrast, Daniels and Rissland (1997) use so-called Pseudo-Relevance Feedback (Salton and Buckley 1990) by assuming the top two retrieved cases as relevant. By doing so, the terms in the cases treated as relevant are added to the initial customer problem, which is expected to lead to the retrieval of semantically more similar cases.

To the best of our knowledge, besides a few preliminary studies that we review in Appendix I (section “Feedback in case retrieval”), studies in the textual CBR literature do not focus on or consider feedback in depth. Thus, in the following we investigate feedback approaches from the related area of information retrieval that attempt to capture and utilize human knowledge on semantic similarity.

Research in information retrieval offers a wide range of feedback approaches for the retrieval of text documents that aim at taking humans’ superior capability to semantically understand texts into account. Although the retrieval process in this context is similar to textual CBR, the objectives differ slightly. While textual CBR approaches aim at retrieving helpful solutions with respect to a full-text description of a problem (Burke et al. 1997; Weber, Ashley, & Brüninghaus, 2005), approaches in information retrieval try to retrieve relevant text documents regarding a query which expresses the user’s request in a few keywords (Baeza-Yates, Ribeiro-Neto, & others, 1999). Nevertheless, approaches from information retrieval show high potential for adaption to textual CBR (Burke et al. 1997; Lenz et al. 1998b; Shekhar et al. 2014). Taking a feedback-oriented perspective, literature in information retrieval can particularly be classified into short-term (Rocchio 1971; Chen, et al. 2006a; Lagun, Sud, White, Bailey, & Buscher, 2013; Salton and Buckley 1990; Sarwar et al. 2018; Zhai and Lafferty 2001) and long-term feedback approaches (Crestani 1994, 2000; Mandl 2000; Lin et al. 2011; Mitra and Craswell 2017). Short-term feedback approaches can be characterized as using feedback only once for a single query, without storing it for use for further similar queries. Thus, these approaches require feedback for each query to enhance the retrieval of text documents, even if the query is nearly identical to previous queries. In contrast, long-term feedback approaches are identified by the storage of feedback to conserve the expressed interconnections between queries and relevant text documents for later use. One popular long-term feedback approach is to instantiate an artificial neural network on the collected feedback, as these algorithms are

able to learn complex mappings between patterns (Cöster and Asker 2000; Crestani 1994, 2000; Mitra and Craswell 2017). These models can be used to adapt new queries without the need for new user feedback. For a more detailed review of these approaches, we again refer the interested reader to Appendix I (section “Feedback in information retrieval”).

Especially long-term relevance feedback approaches from information retrieval, which incorporate the human capability to understand and judge the semantic relationship between a query and retrieval results to enhance future retrievals seem a promising means to foster human-machine collaboration in online customer service through a feedback-based textual CBR approach.

## Research gap and objective

Having surveyed related research in textual CBR and information retrieval, in the following we identify the research gap our novel approach seeks to close, concluding with the definition of the solution’s objectives as the next step in the DSR process (Pefferers et al. 2007).

Prior studies in textual CBR offer well-suited approaches as starting point for our research (Ashley 1991; Burke et al. 1997; Daniels and Rissland 1997; Jayanthi et al. 2010; Lenz et al. 1998b; Weber et al. 2005). Since automated approaches still struggle to meet the challenge of truly understanding the full semantic meaning within texts (Berners-Lee et al. 2001; Embley 2004; Khanapure and Chirchi 2013; Wang et al. 2006b, 2011), human-generated guidance through feedback is still necessary to improve computer systems. However, there is still a lack of feedback-based approaches in textual CBR that take humans’ superior capability to semantically understand texts into account in order to support employees’ search for solutions to new customer problems. As system users in organizations can be expected to be experts in their domain and possess similar knowledge, their feedback can be stored and reused to improve the retrieval for all users. In turn, employees in online customer service could work more efficiently and effectively when supported by advanced case retrieval which has been enhanced based on their own feedback. However, to the best of our knowledge, approaches integrating recent long-term feedback approaches from information retrieval into textual CBR approaches for online customer service bringing together concepts and findings from both research streams are still missing.

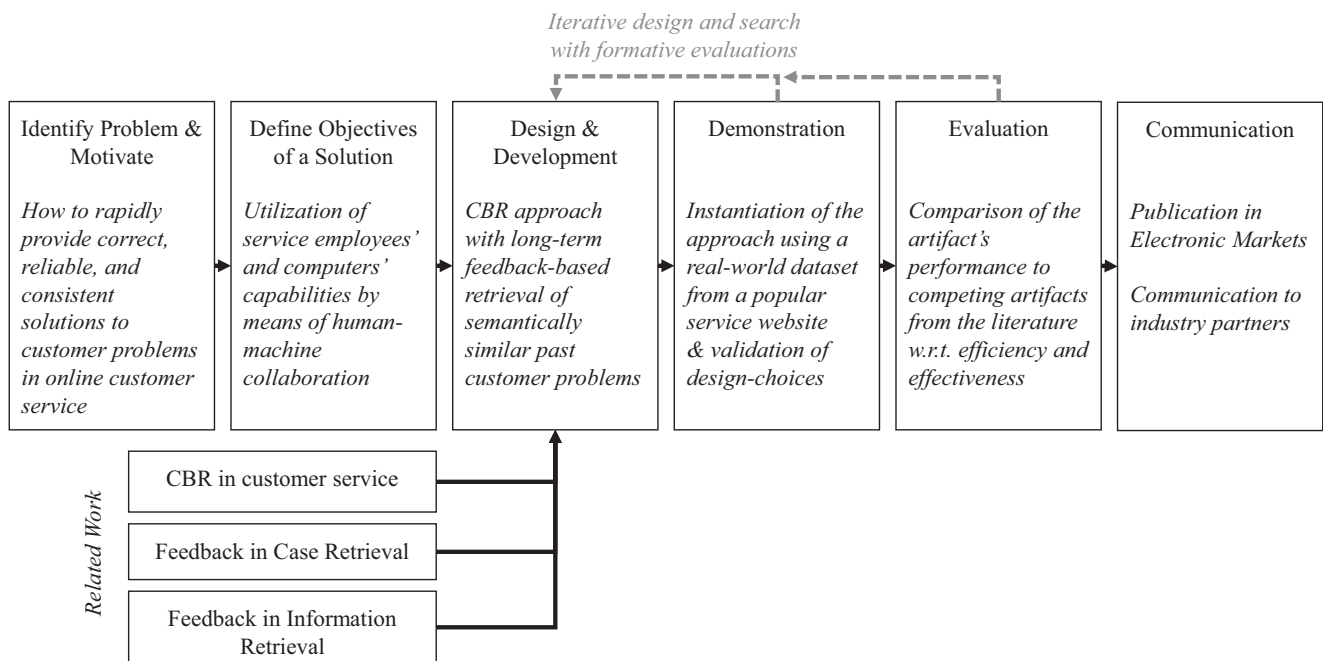
This conclusion drawn from an extensive review of the literature enables us to define the objectives for our solution (Pefferers et al. 2007): Based on well-established methods from information retrieval, we aim to develop a novel textual CBR approach which improves the case retrieval through long-term user feedback. The approach should enhance human-machine collaboration in textual CBR by making use of feedback from

employees during operation of the CBR approach, leveraging the inherent human-machine synergies of CBR: With each added solution and accompanying feedback, employees increase the effectiveness of the CBR approach. At the same time, they profit from the unparalleled speed at which machines can search and process large amounts of data, hence improving their efficiency when solving customer problems. However, the involvement of employees generally can require additional effort on the side of employees. In order to maintain a high level of efficiency, when designing our approach, we intend to keep this additional effort as low as possible. In summary, the objective of our approach is to provide employees with consistent and high-quality knowledge in a short time frame. Thereby, it contributes to an improved online customer service regarding effectiveness as well as efficiency.

## Research method

We conducted and report our research according to the Design Science research (DSR) process by Pefferers et al. (2007), carrying out its six activities as visualized in Fig. 1. First, within the realm of online customer service we identify rapidly responding to customer problems with correct, reliable, and consistent solutions as a key challenge and open research question. Second, we uncover that using the complementary strengths of humans and computers – understanding the semantics of texts and searching vast amounts of data, respectively – appears a promising avenue. Together with the review of previous research on CBR in customer service as well as the incorporation of feedback in both case and information retrieval, this sets the stage for the third activity: The design of our research artifact, a novel long-term feedback-based approach for the retrieval of semantically similar customer problems. Fourth, we instantiate the artifact using a real-world data set and fifth rigorously evaluate its efficacy by comparing its performance to competing artifacts from the literature. The research process concludes with the sixth activity, the communication of the entire research process and findings in the present paper.

DSR efforts can be characterized by their knowledge creation strategy and theorizing mode (Baskerville et al. 2018) as well as their kind of outcome and contribution to knowledge (Gregor and Hevner 2013). Since our research is concerned with the development of a novel approach, it constitutes a contribution of nascent design theory (Gregor and Hevner 2013). As the focus of the conducted research is the design, implementation, and evaluation of an artifact, it is work in interior mode (Baskerville et al. 2018; Gregor 2009; Sonnenberg and Brocke 2012). In our research, we mainly employ an inductive, iterative knowledge creation strategy, producing prescriptive knowledge (Gregor 2009; Sonnenberg and Brocke 2012). More precisely, we start from



**Fig. 1** Overview of the DSR process used for the conducted research (Peppers et al. 2007)

the well-established conventional textual CBR process. Throughout the development, we draw from prior research on feedback, text retrieval, and incorporation of feedback into text retrieval as justificatory knowledge in which the design is grounded (Gregor and Hevner 2013). In terms of the knowledge contribution framework by Gregor and Hevner (2013) our artifact constitutes an “improvement”, striving to improve efficiency as well as effectiveness of online customer service.

Our approach for validation and evaluation follows the “Technical Risk & Efficacy” strategy of the Framework for Evaluation in Design Science (FEDS) put forward by Venable et al. (2016) which structures the evaluation as a four-step process: goal explication, choice of evaluation strategy, determination of the properties to evaluate, and design of the evaluation episodes. The overarching goal of the evaluation is to demonstrate that our novel hybrid approach improves effectiveness and efficiency of online customer service. Hence, a customer sending the description of their problem shall be provided more often with the requested knowledge to raise effectiveness. Further, the time required by an employee to solve the customer problem should be reduced for improving efficiency. We chose the “Technical Risk & Efficacy” strategy as the design of our artefact is subject mainly to technical design risks. The formative, artificial evaluations throughout the earlier stages of the design process prescribed by the strategy ensure the design choices made indeed contribute towards the overall objective, while a final summative evaluation in comparison to competing artifacts demonstrates that the artifact as a whole indeed constitutes an improvement (Gregor and Hevner 2013). In the following, we present the design, demonstration, and evaluation of the artifact as a single

sequence of the process depicted in Fig. 1. Indeed, our artifact consists of a series of steps and elements, which in line with the evaluation strategy have been developed, tested, and validated throughout the search process (Gregor and Hevner 2013; Hevner et al. 2004; Sonnenberg and Brocke 2012). To give an example, we have validated two core elements – the semantic cluster vectors and the semantic context generator – empirically while developing our approach. For sake of brevity and communicative clarity, we defer the description of this validation to the sub-section “Instantiation and application” and refrain from using any test data throughout the artifact’s description (cf. Gregor and Hevner 2013).

## Design of the long-term feedback-based approach

To attain the goal of leveraging humans’ capability to judge the semantic similarity of texts to enhance the retrieval of semantically related customer problems in textual CBR, in our approach each incoming customer problem is adapted prior to the Retrieve phase of the CBR cycle. Specifically, the customer problem is transformed such that *semantically* similar past problems are retrieved rather than past problems which are solely *syntactically* similar with respect to the CBR approach’s similarity function. The knowledge necessary for this adaption is gained from human feedback on semantic similarity of customer problems collected from employees during use of the approach. In this way, by incorporating human feedback our approach enhances efficiency of online customer service by providing employees with semantically similar cases in a short

time frame, so that the time required to solve a customer problem is reduced. Further, it increases effectiveness as employees are consistently provided with past cases semantically similar to the newly incoming customer problem, whose solutions contain knowledge to solve the customer problem. In turn, they can incorporate the best of this and their own knowledge to provide the customer with the requested solution.

## Basic idea and overview

In conventional textual CBR approaches the similarity between an incoming customer problem  $p_i$  and each customer problem  $p_j$  associated with a case in the case base is determined with respect to a similarity function  $\text{sim}(p_i, p_j)$  (Burke et al. 1997; Lenz et al. 1999). As a result, conventional textual CBR approaches suffer from the drawback that a customer problem  $p_j$  similar to an incoming customer problem  $p_i$  with respect to  $\text{sim}(p_i, p_j)$  is not necessarily *semantically* similar to  $p_i$  as well. With our adapted CBR approach, we aim to assist employees with a fast and at the same time high quality retrieval of semantically similar cases by exploiting the complementary strengths of human and artificial intelligence through long-term feedback. To this end, in our approach, each incoming customer problem is adapted based on human knowledge on semantic similarity prior to the Retrieve phase of the textual CBR cycle. Specifically, we draw on employees' feedback on semantic similarity of customer problems collected during the Reuse phase for previously solved customer problems. This way, humans' superior capability to interpret texts is incorporated into the textual CBR cycle.

Our approach consists of three steps (cf. Figure 2).

First, to preserve and later benefit from the information contained in the dismissal and selection of retrieved cases, our approach enables employees to provide feedback on whether retrieved cases are indeed semantically similar to the considered customer problem (cf. step "Gathering Human Knowledge"). This feedback is collected in a feedback base comprising knowledge on semantic relationships of customer problems and in turn used to improve retrieval for further incoming customer problems.

Second, based on employees' feedback on semantic similarities stored in the feedback base we learn a so-called semantic context generator (cf. step "Learning the Semantic Context Generator") that derives the semantic context of incoming customer problems. Based on long-term feedback approaches from literature (Crestani 1994, 2000; Mandl 2000; Lin et al. 2011; Mitra and Craswell 2017) the semantic context generator draws on the combined knowledge on semantic similarity contained in the feedback base. This way, a semantic context  $p_i^s$  for an incoming customer problem  $p_i$  is derived taking into account humans' superior capability to interpret texts.

Finally, in order to integrate humans' knowledge into our approach, in the third step an adapted customer problem  $p_i^a$  is created (cf. step "Adapting the Customer Problem") from the incoming customer problem  $p_i$  and its semantic context  $p_i^s$  generated by the semantic context generator. On the one hand, the resulting adapted customer problem contains human knowledge on semantic similarity of customer problems leading to retrieval of semantically similar cases. On the other hand, the machines' superior capability is exploited as semantically similar problems are retrieved automatically and at a rapid pace.

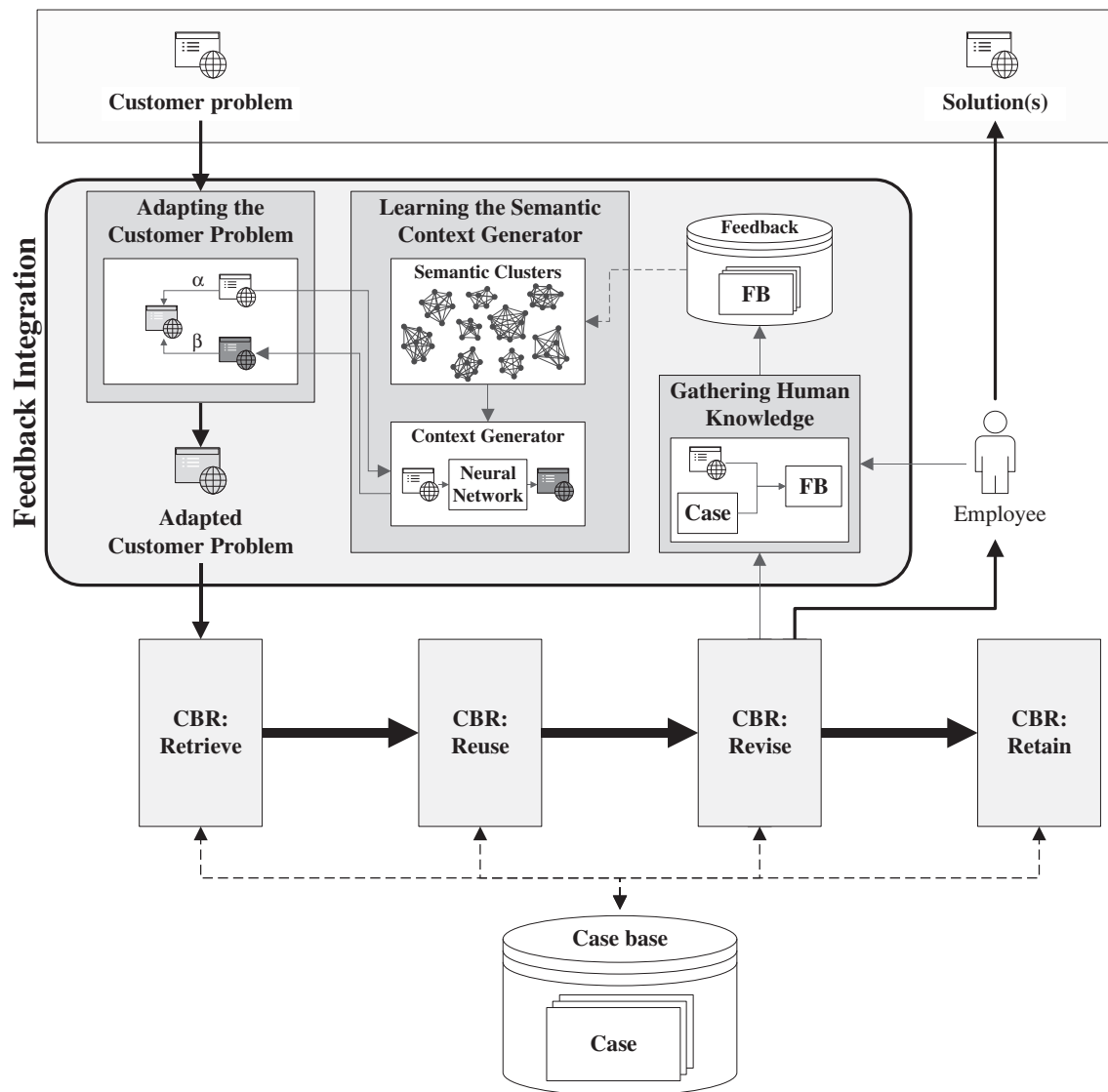
Combining the three steps "Gathering Human Knowledge", "Learning the Semantic Context Generator", and "Adapting the Customer Problem" results in a novel long-term feedback-based approach incorporating humans' superior capability to semantically understand texts in online customer service while at the same time merging concepts from research in textual CBR and long-term feedback approaches from information retrieval. In the following, we detail the three steps of our approach and thereby illustrate how employees' knowledge can be leveraged to incorporate the semantic relationship between texts into textual CBR approaches.

## Gathering human knowledge

We intend to exploit the complementary strengths of human and artificial intelligence in solving incoming customer problems. To do so, we aim to incorporate employees' capability to infer the semantics of texts in terms of feedback into the conventional CBR cycle (Aamodt and Plaza 1994) which serves as a starting point and well-founded basis for our approach (cf. Figure 2). To gather human knowledge, in a first step employees' feedback on the semantic similarity of customer problems is collected and stored in the feedback base  $FB$  as an integral part of the Reuse phase of our adapted CBR cycle. More precisely, following the conventional CBR cycle, in the Reuse phase the  $k$  most similar cases are presented to the employee (Burke et al. 1997; Lenz et al. 1998b). The employee examines these cases and selects those with a customer problem semantically similar to the incoming customer problem  $p_i$  and therefore most suitable to serve as a basis for its solution. In order to later benefit from this intellectual human effort and the manifested knowledge, in our adapted CBR we treat the employee's selection as feedback on semantic similarity. Since employees operating a CBR system always have to identify the semantically most similar cases as a basis for their solution, we are able to create our feedback on-the-fly. This allows our approach to function without a dedicated, potentially time-consuming and costly feedback-collection effort, avoiding a detrimental effect on efficiency.

The choice of feedback scale constitutes an important step in the design of our novel approach. In the literature, various different feedback scales are discussed, which can, in





**Fig. 2** Adapted CBR cycle in online customer service with feedback integration

particular, be classified by the number of scale points (e.g. unary, binary, or multi-point scale) (Boynton and Greenhalgh 2004; Cena et al. 2010; Cena et al. 2011; Krosnick and Fabrigar 1997). To find a suitable rating scale for collecting the employees' feedback, the properties of the quantity to be measured – the semantic similarity of customer problems – need to be taken into account (Krosnick and Fabrigar 1997).

On the one hand, the statement that two customer problems  $p_i$  and  $p_j$  are semantically similar differs fundamentally from the statement that  $p_i$  and  $p_j$  are *not* semantically similar. Only the rating of two customer problems  $p_i$  and  $p_j$  as semantically similar results in a transitive relationship: If a third problem  $p_k$  is semantically similar to  $p_j$ , then  $p_k$  and  $p_i$  are semantically similar as well. If however  $p_i$  and  $p_j$  are not semantically similar, and  $p_j$  and  $p_k$  are not semantically similar either, no information regarding the semantic similarity of  $p_i$

and  $p_k$  can be inferred. This property is best captured by a unipolar scale (cf. Krosnick and Fabrigar 1997).

On the other hand, it appears infeasible to find clear and unambiguous criteria for rating semantic similarity on a multi-point scale. Assume, for example, a five-point rating scale for semantic similarity, with a rating of 5 implying semantically identical and a rating of 1 representing semantically completely distinct customer problems. Further, assume the two customer problems “My phone does not turn on anymore. What should I do?” and “I dropped my phone and now it does not turn on anymore” (cf. Problems A and C in Table 1). While both customer problems are related to a malfunctioning phone, the type of damage differs. Thereby, it appears difficult and context-dependent to decide if this difference leads to a rating of 4, 3, 2, or 1. Hence, a rating on a multi-point scale is necessarily ambiguous. Further, two

customer problems whose similarity to a third customer problem is each rated as 3 might be semantically identical or entirely unrelated to each other.

Taking these considerations into account, a unary scale is a sensible choice to collect feedback on semantic similarity since it provides a single, unambiguous rating and clearly distinguishes the fundamentally different implications of feedback  $FB_{ij} = \text{"not semantically similar"}$  and  $FB_{ij} = \text{"semantically similar"}$ . Further, it can be assumed that employees in online customer service are domain experts who rate semantic similarity of customer problems on a unary scale consistently, avoiding the commonly encountered problem that users tend to use the very same feedback scales differently (Adomavicius and Tuzhilin 2005; Cena et al. 2010; Goldberg et al. 2001; Herlocker et al. 2004).

To sum up, employees' knowledge on the semantic similarity of a customer problem  $p_i$  to a past customer problem  $p_j$  retrieved from the case base is collected using a unary feedback scale during the Reuse phase of our adapted CBR cycle (cf. "Gathering Human Knowledge" in Fig. 2). The employees' feedbacks  $FB_{ij}$  are stored in the feedback base  $FB$ , the set of all feedbacks.

## Learning the semantic context generator

As outlined above, we intend to adapt each incoming customer problem based on human knowledge on semantic similarity. To do so, we learn a model that captures the semantic context of customer problems based on the feedback collected in the step "Gathering Human Knowledge". The semantic context of a customer problem encompasses different descriptions of the same underlying issue and in the following can be used to contribute to an improved retrieval. Literature developing long-term feedback approaches in the related area of information retrieval has often applied different forms of semantic contexts to enhance retrieval accuracy (Crestani 1994, 2000; Huang et al. 2013). To derive the semantic context, we propose to instantiate a so-called semantic context generator  $G(p)$  that learns the semantic relationship between customer problems from the employees' knowledge contained in the feedback base  $FB$  (Huang et al. 2013; Jung et al. 2007; Morrison, Marchand-Maillet, & Bruno, 2008). This way, we encode human knowledge in terms of employees' feedback into a machine-interpretable semantic context while laying a foundation for an advanced human-machine collaboration.

Prior to learning the semantic context generator, the transitive nature of semantic similarity discussed above is used to augment the information regarding the semantic relationships of customer problems. As semantic similarity constitutes a transitive relationship, if two feedbacks  $FB_{ij}$  and  $FB_{jk}$  link  $(p_i, p_j)$  and  $(p_j, p_k)$  as

semantically similar this unambiguously implies that  $p_j$  and  $p_k$  are semantically similar as well, even though there might not be a feedback  $FB_{jk} \in FB$  explicitly indicating this relationship. Consider, for instance, the two customer problems "My phone does not turn on anymore. What should I do?" (Problem A) and "My phone's screen stays dark and it does not seem to start up." (Problem B) (cf. Table 1) as well as feedback  $FB_{AB}$  indicating the semantic relationship between them. Further, assume a third customer problem "My phone is no longer running. How can I start it up?" (Problem E) which is semantically linked to Problem B by feedback  $FB_{BE}$ . Obviously, Problem E is semantically similar to Problem A as well, illustrating the transitivity implied by semantic similarity. Hence, in the absence of feedback  $FB_{AE}$ , this relationship can be deduced from  $FB_{AB}$  and  $FB_{BE}$  (Morrison et al. 2008). Consequently, the set  $c = \{p_i, p_j, p_k, \dots\}$  of customer problems linked through transitive semantic relationships implied by multiple feedbacks contains all customer problems concerning a common issue. Thus, we determine each so-called semantic cluster  $c$  to obtain all sets of customer problems that are semantically similar to each other. To learn a model for the adaption of customer problems, a representation of the semantic clusters is required such that a customer problem can be associated with the joint semantics of the respective cluster (Crestani 1994). Therefore, we encode a semantic cluster  $c$  in terms of a semantic cluster vector  $p_c^s$  by proposing the encoding function  $E(c)$ .  $E(c)$  takes into account all customer problems  $p_i, p_j, p_k, \dots$  part of a semantic cluster  $c$  to find  $p_c^s = E(c)$ . As a result, we can draw on the relationship between each customer problem  $p_i$  and its corresponding semantic cluster vector  $p_c^s$ . In order to do so, we define the semantic context generator  $G(p)$  as a machine-learning model that learns to associate a customer problem with its semantic context. To this end, it is trained pairwise with each customer problem  $p_i$  from the case base and its corresponding semantic cluster vector  $p_c^s$  to maximize the similarity function  $\text{sim}(G(p_i), p_c^s)$ . By this means, we enable the semantic context generator  $G(p)$  to derive a semantic context  $p_i^s = G(p_i)$  based on the incoming customer problem  $p_i$  (Crestani 1994).

To sum up, based on the step "Gathering Human Knowledge" the semantic context generator  $G(p_i)$  learns from employees' joint feedback in terms of semantic cluster vectors  $p_c^s$  to generate the semantic context  $p_i^s$  for incoming customer problems  $p_i$ .

## Adapting the customer problem

Finally and in order to ensure a synergistic human-machine collaboration in online customer service by merging humans' superior capability to semantically understand texts with machine's ability to process data at a rapid pace, we adapt the incoming customer problem  $p_i$  using its semantic context obtained in the previous step. Based on the well-known short-

term feedback approach Relevance Feedback, we adapt the incoming customer problem using the popular query reweighing technique derived from the Rocchio Algorithm (Carpineto and Romano 2012; Manning, Raghavan, & Schütze, 2008; Rocchio 1971; Salton and Buckley 1990). However, in contrast to standard Relevance Feedback, we do not collect feedback to adapt the incoming customer problem prior to retrieval. Instead, we incorporate employees' feedback through the semantic context  $p_i^s$  derived by our semantic context generator  $G(p_i)$ . Thereby we combine the incoming customer problem  $p_i$  with its semantic context  $p_i^s$ . Thus, we enable a well-founded retrieval of semantically similar cases by means of human-machine collaboration.

More precisely, we create an adapted customer problem  $p_i^a$  such that the given similarity function  $\text{sim}(p_i^a, p_j)$  is maximized for all  $p_j$  which are *semantically* similar to  $p_i$  but minimized for all  $p_j$  for which this is not the case. In order to achieve this, we combine employees' knowledge on semantic similarity – the semantic context introduced in the previous step – and the incoming customer problem. We define  $p_i^a$  as the weighted combination of the incoming customer problem  $p_i$  and its semantic context  $p_i^s$  (cf. Equation (1)). On this account, the weights  $\alpha \in [0, \infty)$  and  $\beta \in [0, \infty)$  determine the extent to which the incoming customer problem  $p_i$  and its semantic context  $p_i^s$  are represented within the adapted customer problem  $p_i^a$ :

$$p_i^a = \alpha \cdot p_i + \beta \cdot p_i^s \quad (1)$$

To sum up, the adapted customer problem  $p_i^a$  is defined as the weighted combination of the incoming customer problem  $p_i$  and its semantic context  $p_i^s$ . In our approach, the adapted customer problem  $p_i^a$  is used in place of the incoming customer problem  $p_i$  in the Retrieve phase of the CBR cycle. As outlined above, the resulting refined retrieval of semantically similar problems is expected to lead to improved performance in online customer service. In this way, our approach leverages the complementary strengths of human and artificial intelligence through the incorporation of human feedback into the CBR cycle.

## Demonstration of the approach

As an essential part of the DSR process (Gregor and Hevner 2013; Hevner et al. 2004; Peffers et al. 2007) we demonstrate the practical applicability of our approach. To this end, we use a publicly available real-world data set of the popular service website Quora. In the following, we first introduce the data set of suitable customer problems and elucidate the general validation and evaluation setup. Following this, we instantiate the

components of our approach and in the process validate the design choices made during its development (Venable et al. 2016).

## Data set

To instantiate our approach and evaluate its efficiency as well as its effectiveness we use a publicly available real-world data set containing 404,288 pairs of customer questions and their corresponding semantic relationships published by the popular service website Quora (Iyer, Dandekar, & Csemai, 2017). On Quora, users ask questions on a wide variety of topics that are answered by an international community comprised of both laypeople and topic experts (Wang et al. 2013). These answers are discussed and judged through voting by fellow community members (Wang et al. 2013). To keep their knowledge base redundancy-free, Quora aims to have each semantically distinct question answered only once (Bodnick 2015; Iyer et al. 2017). To facilitate this, all registered users can merge questions (Scharff 2015), with some complex merges requiring review by Quora staff (Wacker 2016). If questions are merged, future visitors will be redirected to the incarnation of this question deemed to be phrased best by the merging user (Scharff 2015). In total, the data set comprises 149,496 distinct customer questions that are linked by feedbacks indicating whether two customer questions are semantic duplicates. One example is (“How do I get my iPhone out of recovery mode?”, “What do I do if my iPhone is stuck in recovery mode?”, Duplicate). The questions asked on Quora and contained in the chosen data set are similar in style to online customer service. For this reason, the data set provides an appropriate setting to demonstrate our novel long-term feedback-based approach.

## Instantiation and application

For the demonstration and later evaluation of our approach, we use the whole set of 149,496 distinct questions contained in the Quora data set as customer problems  $p_j$ . Following conventional textual CBR approaches, we handle and store customer problems as representations in the well-established vector space model which has proven successful in many retrieval applications (Burke et al. 1997; Manning et al. 2008; Salton, Wong, & Yang, 1975). To transform the customer problems into vector representations, we preprocess the questions by removing stop words (Manning et al. 2008) and using a Porter stemmer (Manning et al. 2008; Porter 1980). Then, we convert each question into a tf-idf vector (Hua et al. 2009; Manning et al. 2008; Salton and Buckley 1988; Sebastiani 2002), limiting the total vocabulary size to  $N_{\text{voc}} = 10,082$  terms by omitting words which appear fewer than four

times in the whole corpus (cf. Turney and Pantel 2010). As the semantic context generator  $G(p)$  learns to maximize  $\text{sim}(G(p_i), p_c^s)$ , it is capable of adapting to any textual CBR similarity measure. Here, we use the cosine similarity  $\cos(p_i, p_j)$  which is commonly used for similarity determination in textual CBR with tf-idf vectorization (Bedué et al. 2018; Burke et al. 1997; Lenz et al. 1998b; Manning et al. 2008; Salton and McGill 1984).

In line with CBR literature (Bedué et al. 2018; Burke et al. 1997; Lenz et al. 1998a), we refer to the retrieval for a specific customer problem as successful if at least one semantically similar case is contained within the  $k$  retrieved cases. In the following, we refer to the proportion of customer problems for which the retrieval was successful with respect to the total number of customer problems as *successful retrievals* (Bedué et al. 2018). Whenever necessary, we set  $k=5$  as this is a reasonable number of cases to display at once for an employee to scan, a commonly made assumption in similar application contexts (cf. Balakrishnan et al. 2016; Burke et al. 1997). On this basis, we instantiate and validate our approach focusing on its major steps “Gathering Human Knowledge”, “Learning the Semantic Context Generator”, and “Adapting the Customer Problem”.

## Gathering human knowledge

In our approach, employees’ knowledge on the semantic similarity of customer problems is gathered in terms of feedback during the Reuse phase. In the chosen data set feedback has already been provided by humans and consolidated (Scharff 2015), which allows us to simulate the collection of feedback from employees. To this end, we assume that the pairs of questions contained in the Quora data set that are marked as semantic duplicates are semantically similar customer problems. In this regard, the Quora feedback process corresponds to an employee providing unary feedback  $FB_{ij}$  on customer problems, enabling us to create our employee feedback base  $FB$  directly from the data set. As a result, we obtain a feedback base  $FB$  comprising 149,263 distinct unary feedbacks  $FB_{ij}$  linking two semantically similar customer problems  $p_i$  and  $p_j$ .

## Learning the semantic context generator

To derive the semantic context of customer problems based on the feedback stored in the feedback base  $FB$ , the semantic context generator has to be instantiated. To do so, we first determine the semantic clusters  $c$ ; second, specify the encoding function  $E(c)$  to determine the semantic cluster vectors  $p_c^s$ ; and third, learn the semantic context generator  $G(p)$ .

First, we determine the semantic clusters  $c$  by identifying all customer problems in the feedback base  $FB$  linked as semantically similar by feedback or a transitive semantic relationship. This results in 6,279 semantic clusters comprising between 2 and 109 customer problems. To learn the semantic relationship between customer problems, a sufficient number of them is needed within each cluster  $c$ . Therefore, and to support adequate splits in training, validation, and test data, we focus on the semantic clusters comprising more than 50 customer problems. This leads to 13 semantic clusters covering a total of 933 distinct customer problems and comprising between 51 and 109 customer problems each, which in the following we use for demonstration and evaluation purposes. To enable a thorough validation and evaluation of our approach, we conduct a five-fold cross-validation, splitting the 933 customer problems into equally sized balanced test sets and performing the subsequent steps for each of these (cf. Goodfellow, Bengio, & Courville, 2016). For each fold, the test set is removed prior to any processing.

Second, we specify the encoding function  $E(c)$  to create a semantic cluster vector  $p_c^s$  for each semantic cluster  $c$  maximizing  $\text{sim}(p_i, p_c^s)$  for all customer problems  $p_i \in c$ . In our vector space representation with  $\cos(p_i, p_j)$  the cluster’s centroid is the vector with the highest average similarity to all customer problems  $p_i \in c$  and is thus well-suited to represent the semantic cluster vector  $p_c^s$  of a cluster (Chen et al. 2006a; Crestani 1994). This is precisely the relationship exploited in Relevance Feedback, where the users’ short-term feedback is used to approximate the centroid of the cluster of documents relevant to a query (Manning et al. 2008). Hence, we define the encoding function  $E(c)$  as the cluster’s centroid:

$$E(c) = \frac{1}{|c|} \sum_{p_i \in c} p_i = p_c^s \quad (2)$$

To empirically validate our choice of the cluster’s centroid as the semantic cluster vector  $p_c^s$ , for each incoming customer problem  $p_i$  we use the centroid  $p_c^p$  of the semantic cluster  $c$  it belongs to for retrieval ( $p_i^a = p_c^p$ ). We find that this maximizes our evaluation metric, i.e. for each cross-validation fold each retrieval yields at least one customer problem that is semantically similar to the incoming customer problem, fulfilling the expectation that the cluster centroid yields optimal retrieval performance (cf. Manning et al. 2008).

Third, we instantiate and learn the semantic context generator  $G(p_i)$  based on the customer problems (samples) and the corresponding semantic cluster vectors  $p_c^s$  (targets). In order to instantiate the semantic context generator  $G(p_i)$ , we use an artificial neural network due to its great potential to perform well on tasks involving large and sparse vectors, such as those present in our application context (Goodfellow et al. 2016). To find a suitable parameterization for the given task, we follow the procedure outlined by



Ng (2018) and start with an initial configuration, validate how well it performs regarding a loss function, and then iteratively adjust the configuration to improve upon this baseline. As the loss function measuring the distance between the semantic context generator's output  $G(p_i)$  and the desired target  $p_c^s$  we choose the negative of the similarity function  $-sim(G(p_i), p_c^s) = -\cos(G(p_i), p_c^s)$  that reaches its minimum when the similarity function  $\cos(G(p_i), p_c^s)$  of our textual CBR approach is maximized. Further, we opt to use the ReLU activation function (Glorot et al. 2011; Goodfellow et al. 2016) for all layers since its range of values  $[0, \infty)$  matches that of the vector for  $p_i$ . To learn the semantic context generator, for each fold of the cross-validation we split the available data (note that the test set has been removed) into a training and validation set of which only the former is used for training (Goodfellow et al. 2016; Ng 2018). To verify that the semantic context generator was learned successfully, the model's output for a customer problem  $p_i$  is compared with the associated target from the training and validation set by means of the loss function. While the length of  $p_i$  – the number of words in the vocabulary  $N_{voc}$  – determines the number of neurons in the input and output layers, both the number of hidden layers and their size have to be determined by iterative experimentation. We find that two fully connected hidden layers with 5,000 neurons each are well suited to the task at hand. We further use dropout (Srivastava et al. 2014) for better generalization and stop the training of our neural network as soon as the loss on the validation set does not decrease for five epochs in order to prevent overfitting (cf. Prechelt 2012).

We find that the neural network chosen for  $G(p_i)$  is well suited to infer the semantic cluster vector  $p_c^s$  for an incoming customer problem  $p_i$  as the average losses are nearly  $-1.0$  which represents an optimal result for the chosen loss function. Specifically, the average loss on our training sets yields  $-0.98$  while the average loss on the validation set is  $-0.97$ , indicating good generalization. The average test set loss is comparable as well ( $-0.97$ ).

## Adapting the customer problem

Finally, the output of the semantic context generator  $p_i^s = G(p_i)$  is used to adapt the incoming customer problem  $p_i$ , creating the adapted customer problem  $p_i^a$ . Subsequently, we have to determine the weights  $\alpha \in [0, \infty)$  and  $\beta \in [0, \infty)$  for merging the incoming customer problem  $p_i$  and its semantic context  $p_i^s$ . To approach this issue, we use the customer problems from the validation set as input in the Retrieve phase of our adapted CBR cycle and optimize  $\alpha$  and  $\beta$  regarding the proportion of *successful retrievals*. As the semantic context

generator  $G(p_i)$  outputs the semantic clusters' centroids almost perfectly,  $\alpha = 0.0$  and  $\beta = 1.0$  are expected to yield the optimum proportion of successful retrievals. Indeed, we find that varying  $\alpha$  decreases the metric. If however, customer problems in other application contexts deal with multiple topics simultaneously,  $\alpha > 0.0$  is reasonable in order to consider both the customer problem and its semantic context.

Finally, we applied the instantiated approach and used the customer problems from the test set as input for the Retrieve phase of our adapted CBR cycle. On this account, we incorporate employees' feedback on semantic similarity of customer problems contained in the feedback base  $FB$  by generating the semantic context of a customer problem contained in the test set. Subsequently, the semantic context is used to adapt the customer problem to improve the retrieval of semantically similar customer problems. Preprocessing the customer problem, generating the semantic context, and adapting the customer problem requires just a few milliseconds even on a standard laptop computer and thus significantly less time than the CBR retrieval from a large case base. By this means, we are able to adequately assist employees in online customer service by a fast and at the same time high-quality retrieval of semantically similar cases.

## Evaluation of the approach

As demanded by the DSR process (Peppers et al. 2007), we conduct a summative evaluation (Venable et al. 2016) of our approach. The goal of this evaluation is to show that our long-term feedback-based approach is indeed an improvement (Gregor and Hevner 2013) in terms of both effectiveness and efficiency over existing approaches regarding our problem context. To this end, we compare the performance of our novel hybrid approach on the Quora data set chosen for its instantiation both with solely machine-based (fully automated) and hybrid as well as entirely human-based (manual) approaches.

To evaluate our approach against machine-based as well as hybrid approaches, we compare its performance against that of competing artifacts in textual CBR. To ensure comparability, all considered approaches are based on the same conventional textual CBR approach. As the baseline (BL) we take the solely machine-based CBR retrieval of related customer problems  $p_j$  in absence of human feedback and hence without any further adaption of the incoming customer problem ( $p_i^{a-BL} = p_i$ ) (Bedué et al. 2018). Additionally, we use Relevance Feedback (RF) and Pseudo-Relevance Feedback (PRF) as the two probably most established short-term feedback approaches from information retrieval (cf. Manning et al. 2008) that have already been integrated into textual CBR approaches (Daniels and Rissland 1997). As in our approach, we utilize the well-known Rocchio query reweighting technique

to integrate these approaches into the conventional textual CBR cycle.

Relevance Feedback (Manning et al. 2008; Salton and Buckley 1990), the first competing artifact, requires the customer service employee to provide feedback on the top five retrieved customer problems on the semantic similarity to the incoming customer problem. As the Quora data set already contains the semantic relationship between customer problems (cf. “Gathering Human Knowledge”), we can simulate the respective human-machine collaboration. On this basis, the adapted customer problem  $p_i^{a-RF}$  is obtained as the weighted combination of the incoming customer problem  $p_i$  as well as the simulated human feedback regarding semantically similar and not semantically similar cases:

$$p_i^{a-RF} = \alpha p_i + \beta \sum_{\text{similar}} p_k - \gamma \sum_{\text{not similar}} p_k \quad (3)$$

We integrate Pseudo-Relevance Feedback as the second competing artifact. Here, each of the top five customer problems returned by the conventional textual CBR retrieval is taken to be relevant and used to create the adapted customer problem  $p_i^{a-PRF}$  (Manning et al. 2008). Thus, the adapted customer problem  $p_i^{a-PRF}$  is modeled as the weighted combination of the incoming customer problem  $p_i$  and the top five retrieved customer problems marked as semantically similar:

$$p_i^{a-PRF} = \alpha p_i + \frac{1}{5} \beta \sum_{l=1}^5 p_l \quad (4)$$

For both competing approaches, in absence of a deterministic algorithm (Moschitti 2003) we empirically determine the optimal pre-factors  $\alpha$ ,  $\beta$ , and  $\gamma$  on the validation set by setting  $\alpha = 1.0$  and using a simple hill-climbing algorithm (Russell and Norvig 2010). In the case of Relevance Feedback, we first obtain an optimal value for  $\frac{\alpha}{\beta}$  keeping  $\gamma = 0$  and subsequently repeat the procedure varying  $\gamma$  while keeping  $\frac{\alpha}{\beta}$  fixed.

As the metric to evaluate the effectiveness of the machine-based as well as hybrid approaches, we calculate the proportion of *successful retrievals* by comparing the cases retrieved with the set of semantically similar cases in the corresponding semantic cluster as ground truth. This metric is chosen as it clearly reflects the effectiveness of an approach with regard to providing customers with the requested knowledge: Retrieving semantically similar cases more often increases the effectiveness of employees in online customer service as the knowledge to craft a solution is readily available.

For each of the approaches, we conduct a five-fold cross-validation as introduced above, measuring the proportion of successful retrievals for each of the five test sets. Averaged over the cross-validation folds (cf. Fig. 3) our approach retrieves a semantically similar customer problem among the top five retrieved problems (i.e.,  $k = 5$ ) 97.9% of the time. In

contrast, the baseline approach performed quite poorly, as only 86.3% of retrievals yield a semantically similar customer problem for  $k = 5$ , indicating an inferior effectiveness. The retrieval results of the competing artifact based on Pseudo-Relevance Feedback are affected negatively by incorrect cases among the top five retrieval results of the baseline approach. Thus, the Pseudo-Relevance Feedback approach slightly outperforms or fails to surpass the baseline depending on  $k$ , retrieving a semantically similar customer problem for  $k = 5$  in 85.1% of all cases. In contrast, Relevance Feedback returns a semantically similar customer problem in 88.9% of all cases. Here we can see that since in Relevance Feedback only the top five retrieved problems are considered, its effectiveness regarding the very first retrieved customer problem is approximately equal to the baseline performance for  $k = 5$  and thus better than our approach, but does not increase much beyond. In contrast, our approach surpasses Relevance Feedback already for  $k = 2$ .

To compare the efficiency of our hybrid approach against the competing solely machine-based and hybrid approaches, we first focus on the time required for retrieval. As the approaches are based on the same conventional CBR approach, the steps required by an employee to work with the approaches are the same. Our hybrid approach merely requires about 0.04 s more computing time (on a standard laptop) than the baseline approach, whereas due to the twofold retrieval Relevance and Pseudo-Relevance Feedback require at least double the baseline computing time. In the case of Relevance Feedback, the manual feedback phase in practice likely dominates the overall retrieval duration. Hence, in terms of retrieval duration our approach yields virtually the same level of efficiency as the baseline approach while outperforming the competing short-term feedback approaches. Further, in the evaluated application scenario, this advantage of our approach regarding efficiency is increased by the high rate of successful retrievals: Since it returns at least one semantically

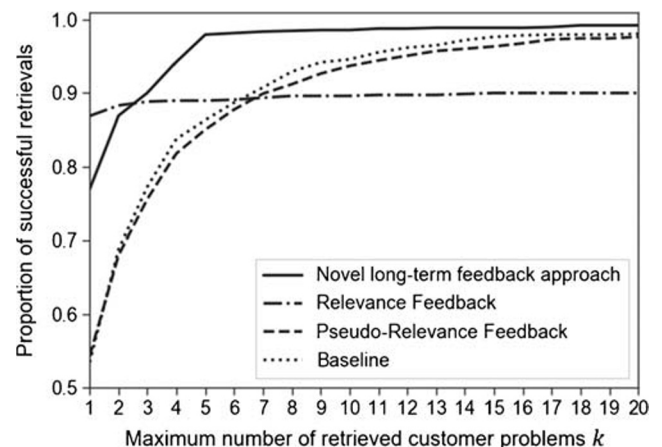


Fig. 3 Evaluation of our approach in comparison to competing artifacts

similar customer problem in close to 98% of cases, employees need to undergo the lengthy process of crafting a solution from scratch significantly less often than when using the baseline or short-term feedback approaches.

Comparing our hybrid approach to an entirely human-based approach in terms of effectiveness and efficiency is hardly possible without an additional field study where service experts are timed while they phrase solutions to customer problems that are then rated by customers regarding their quality. While this is out of scope of the present study, we can nevertheless compare the approach by relying on fair assumptions. Regarding the effectiveness of our approach, we argue that providing a service expert with semantically similar cases in addition to their own knowledge does not lower effectiveness in terms of providing the customer with the requested knowledge. To the contrary, providing employees with relevant knowledge to solve customer problems should reduce the limitations and errors inherent to employees phrasing solutions from scratch only based on their own knowledge. With regard to efficiency, we rely on empirical data on typical reading and typing speeds to estimate that a service expert on average requires 4 min to phrase a solution to a customer problem from scratch (under the generous assumption that service experts are able to immediately type out solutions to customer problems without prior deliberation). In contrast, for our hybrid approach we estimate an average total time of up to 3:30 min to provide a solution, taking into account the cases where no semantically similar past customer problem can be retrieved and allowing for up to 1 min for selection and adaption of the reused solution. For details on the assumptions underlying these estimates, see Appendix II.

Taken together, our long-term feedback-based approach clearly outperforms solely machine-based and hybrid approaches in terms of effectiveness and outperforms both the baseline textual CBR approach as well as the competing short-term feedback approaches in terms of efficiency. Finally, our hybrid approach is arguably more effective and clearly more efficient than an entirely human-based approach.

## Discussion, limitations, and future research

Truly understanding the semantic meaning of texts is still a challenge for computer systems. In particular, prior textual CBR approaches do not fully leverage the complementary strengths of human and artificial intelligence. Against this background, we proposed a novel long-term feedback-based approach taking humans' capability to semantically understand texts into account. Our results illustrate that integrating employees' knowledge into textual CBR by means of long-

term feedback leads to superior results. In detail, our contributions to theory and practice are threefold.

First, our approach improves upon existing approaches (cf. Gregor and Hevner 2013) by fostering a synergistic human-machine collaboration. Compared to conventional CBR approaches which only evolve with the number of customer problems solved, our approach additionally incorporates humans' feedback as a learning component by exploiting the Reuse phase and augmenting the Retrieve phase. Thus, on the one hand the adapted customer problem of our approach contains human knowledge regarding the semantic similarity of customer problems leading to a refined retrieval of semantically similar cases. On the other hand, machines' superior capabilities are exploited as semantically similar problems are retrieved at a rapid pace. Indeed, this fosters solving customer problems efficiently. Further on, our approach provides all employees with relevant and detailed knowledge beyond their own to solve customer problems. Hence, we also contribute to more effective as well as consistent solutions in online customer service. As a result, through incorporation of feedback our approach enables a service quality level greater (successful retrieval) or equal (no successful retrieval) than without the approach. In the rare case that our approach does not return a semantically similar case, the employees can still phrase a solution from scratch, just as in the entirely manual approach. If, however, employees instead reuse solutions from the semantically different cases presented by the system, effectiveness might be affected adversely. Hence, to avoid this competing effect, employees must be provided with clear instructions on how to work with our approach.

Second, our novel long-term feedback-based approach contributes to the development of a refined textual CBR approach incorporating humans' superior capability to semantically understand the relation between customer problems. Since conventional textual CBR approaches suffer from the drawback that a close match regarding the similarity of two customer problems does not necessarily mean that they are *semantically* similar as well, human guidance is still needed to determine the semantic meaning of customer problems. Here, our approach improves upon previous approaches by leveraging employees' capability to capture the semantic relation between customer problems in terms of feedback. Furthermore, as we provide an approach to adequately assist employees with a fast and at the same time high-quality retrieval of semantically similar, employees and organizations benefit from our approach insofar as the quality of provided solutions improves over time. Thus, we take a step towards fully grasping the semantic meaning of texts in customer service, which is still a challenge for automated approaches (Berners-Lee et al. 2001; Embley 2004; Khanapure and Chirchi 2013; Wang et al. 2006b, 2011).

Finally, our study builds upon research in textual CBR (Burke et al. 1997; Lenz and Burkhard 1997) and extends its

line of thoughts by integrating a long-term feedback approach from information retrieval (e.g. Crestani 1994, 2000; Mandl 2000; Lin et al. 2011; Mitra and Craswell 2017). While doing so, we merge concepts from research in textual CBR and recent long-term feedback approaches in information retrieval. Since prior literature does not provide such an integrated perspective of textual CBR and long-term feedback approaches, we address this gap and substantially extend existing contributions.

Besides its benefits, our approach and study also implicate limitations that can serve as starting points for future research. First, regarding the demonstration and evaluation of our approach, the implementation and evaluation of our hybrid approach within an operating online customer service department remain a desideratum. Our work paves the way to empirically investigate the influence that hybrid approaches, as ours, may have on efficiency and effectiveness in online customer service as well as the trade-off that could appear between them. Further, for the demonstration and evaluation, we considered only one data set. Although the customer problems contained in the open-domain data set published by the popular service website Quora conform to the properties of customer problems in online customer service, one single customer problem could refer to multiple different relevant topics and might contain significantly more text. However, the feedback on duplicates in a service context closely resembles the feedback on semantic similarity of customer problems and the free availability of the data set enables direct and rigorous quantitative comparison of (future) feedback-based approaches. Nevertheless, as one possible next step in exploring long-term feedback-based approaches, we encourage researchers to apply and evaluate our approach in real-world customer service settings. Second, while we focused on integrating long-term feedback by adapting the customer problem prior to the core CBR process, it also seems promising to investigate the integration of long-term feedback into other parts of the CBR cycle. Promising starting points include adapting the similarity function (Huang et al. 2013; Weis 2013) or gathering implicit feedback on the basis of employees' behavior when reviewing and selecting cases. Finally, while we only consider employees' feedback on customer problems, it could also prove beneficial to integrate further information during query adaption. For example, various kinds of data available about the customer (e.g. past communication, purchases, personal information) might be included. This additional context information may help to further increase the quality of the proposed solutions. For instance, if an incoming customer problem contains references to previous requests or omits important order details that can be deduced from the customer's purchase history. Further, it might be beneficial to weight recent feedback higher than earlier feedback to reflect refinements in the employees' understanding of semantic similarities or changes in the

company's policies (e.g., one of two similar products is discontinued by an organization and therefore earlier feedbacks linking these two products are outdated).

## Conclusion

Nowadays, organizations face the challenge of meeting customers' demand for reduced response times while handling customer requests with a consistently high level of service quality. Since until now automated approaches still struggle to meet the challenge of truly understanding the full semantic meaning of texts (Berners-Lee et al. 2001; Khanapure and Chirchi 2013), human guidance through feedback is still necessary. Despite extensive scientific work in the field of textual CBR and information retrieval, so far no study has considered intensifying hybrid human-machine collaboration to enhance case retrieval for new customer problems by investigating the semantic relationships between free-text cases through long-term feedback. To this end, we propose a novel approach in textual CBR which incorporates human knowledge in terms of long-term feedback. We gather employees' feedback regarding the semantic similarity of customer problems in the Reuse phase of the CBR cycle. The collected feedback from all employees is used to create semantic clusters and to train a semantic context generator. Finally, the semantic context of incoming customer problems is determined to enhance the retrieval of semantically similar cases. The demonstration and evaluation based on a real-world data set illustrates that our long-term feedback-based approach clearly outperforms solely machine-based and hybrid approaches in terms of effectiveness, retrieving semantically similar customer problems in 98% of cases compared to 87% (baseline CBR) and at most 89% (Relevance Feedback), respectively. Further, our approach outperforms the baseline textual CBR approach in terms of efficiency, as employees need to provide a solution from scratch less frequently, reducing the average time to provide a solution by at least 12.5%. It is also more efficient than the competing short-term feedback approaches, as it requires only a single retrieval. Additionally, it is arguably more effective and clearly more efficient than an entirely human-based approach. Thus, our approach improves performance in online customer service. It fosters a synergistic human-machine collaboration and contributes to the development of a more refined textual CBR approach regarding the semantic relation of customer problems by merging concepts from research in textual CBR and long-term feedback approaches in information retrieval. Against this background, our approach constitutes a promising first step in order to overcome current challenges in understanding the semantic meaning of texts in textual CBR and beyond.

**Funding Information** Open Access funding provided by Projekt DEAL.



## Appendix I: Expanded review of related work

In this appendix we expand on the section “Related work in textual CBR and information retrieval” and give a more detailed review of the application of textual Case-Based Reasoning (CBR) in Online Customer Service as well as approaches incorporating user feedback in both case and information retrieval.

### Supporting online customer service with Case-Based Reasoning

Irrespective of a specific application domain, all approaches for online customer service automatically providing employees with similar cases and learning from human knowledge in terms of already solved customer problems are based on CBR (Acorn and Walden 1992; Bedué et al. 2018; Heras et al. 2009; Lenz et al. 1999; Lenz and Burkhard 1997; Lenz et al. 1998b). Thus, CBR paves the way for a hybrid intelligence where humans and machines act as teammates (Gu et al. 2017; Martin et al. 2017; Reuss et al. 2015).

CBR is a well-established methodology in artificial intelligence for solving problems through reusing solutions of previously solved similar cases (Aamodt and Plaza 1994; Yan et al. 2014). To do so, solutions from a huge amount of cases contained in a case base are automatically retrieved and suggested as most suitable answers for a new customer problem (Acorn and Walden 1992; Kriegsman and Barletta 1993; Simoudis 1992). In case of unsuitable suggestions (based on human judgements), an employee creates a new solution and the respective case is added to the case base. Thus, CBR constitutes a self-learning approach that ensures consistent, fast and high-quality solutions and evolves with the number of problems solved by a sound human-machine collaboration. Against this background, CBR is capable of supporting human-machine collaboration to retrieve the most relevant solutions in online customer service. By this means, companies can benefit from the superior capabilities of machines to search and process large amounts of data in an efficient way to provide solutions in time regardless of the responding employee. Therefore, CBR seems predestined to meet the rising expectations of customers (Forrester 2018) in all of the frequently used online channels (Statista 2017). As customer interactions in online customer service are generally based on textual messages, particularly the research stream of textual CBR seems to provide promising approaches to cope with the task of enabling a hybrid intelligence.

Textual CBR approaches are based on the well-established CBR cycle introduced by Aamodt and Plaza (1994) with the goal of developing an approach for automated problem solving. In the context of online customer service, the textual CBR approach comprises a case base

of all existing and solved cases  $c_j$ , each consisting of a textual description of the customer problem  $p_j$  and a solution  $s_j$  (Burke et al. 1997; Cunningham et al. 2004; Lenz et al. 1998b; Wang et al. 2006b). Case retrieval starts with an incoming customer problem  $p_i$  and aims to quantify the degree of resemblance between  $p_i$  and all customer problems  $p_j$  of existing cases  $c_j$  in the case base by means of a similarity function  $sim(p_i, p_j)$  (Liao et al. 1998). Subsequently, the  $k$  most similar problems and their solutions are presented to the employee, who can choose to reuse one or more of the retrieved solutions  $s_j$ , revise these solutions regarding the current customer problem  $p_i$  if necessary, and finally adds the new case  $c_i$ , comprising the incoming customer problem  $p_i$  and the corresponding new solution  $s_i$ , to the case base (Lenz et al. 1999; Wang et al. 2011; Wang et al. 2006a).

Literature mainly focuses on the retrieval of similar cases based on the free-text representation of the customer problem or the query, respectively (Ashley 1991; Balakrishnan et al. 2016; Burke et al. 1997; Daniels and Rissland 1997; Jayanthi et al. 2010; Lenz et al. 1998b; Sizov et al. 2015; Wang et al. 2011; Weber et al. 2005). Most textual retrieval approaches make use of well-known methods from information retrieval to retrieve similar cases (Burke et al. 1997; Hammond et al. 1995; Kunze and Hübner 1998; Lenz and Burkhard 1997; Lenz et al. 1998b; Shekhar et al. 2014; Wilson and Bradshaw 1999). For instance, Burke et al. (1997) rely on the Vector Space Model (Salton et al. 1975). Their FAQ Finder retrieves the most similar questions and corresponding answers from a case base. Others (Kunze and Hübner 1998; Lenz and Burkhard 1997; Shekhar et al. 2014) rely on the Inference Network Model (Turtle and Croft 1990) by embedding cases into a network linked with Information Entities representing statistically identified or domain-specific phrases or terms.

Nevertheless, whereas most existing textual CBR approaches seem promising to collaborate with employees in online customer service by proposing similar already solved cases based on the customer problem, their retrieval results are primarily based on the information contained in the customer problem or query, respectively. Thus, a collaboration between employee and machine, taking the feedback of employees regarding the relevance of proposed cases into account, seems promising to extend and enhance existing textual CBR approaches.

### Feedback in case retrieval

In contrast to machines, a human reader has multiple skills which are challenging to attain for automated approaches. First of all, humans are able to understand and interpret rhetoric or linguistic specificities as for example irony or sarcasm. Moreover, content written by humans, e.g. customer

problems, often contains indirectly stated information, such as social background, level of education, or age of the author, all of which could hardly be identified without human life experiences. Further, despite very different (similar) vocabulary, two customer problems could refer to quite similar (different) types of problems (cf. Table 1). Understanding the semantics of text is a task humans excel at. Human judgement has shown beneficial in enhancing and guiding a computer system (Salton and Buckley 1990; Sarwar et al. 2018; Trstenjak and Donko 2016). Hence, to fulfill customers' needs in online customer service an understanding of the semantic meaning in textual messages exchanged between customers and employees is important to provide a correct solution concerning customers' problems. As a result, many approaches take employees or users into account to validate automated solutions for customer problems (Balakrishnan et al. 2016; Kunze and Hübner 1998; Lenz et al. 1999; Weis 2013). However, the potential in CBR as hybrid intelligence where humans and machines collaborate on equal terms harbors potential for improvement.

To exploit this potential, some authors in the research area of CBR have started to introduce approaches which incorporate feedback from the system user into the retrieval of new cases (Branting 2001; Cheng and Hüllermeier 2008; Coyle and Cunningham 2003; Gabel and Stahl 2004; Leake and Dial 2008; Soh and Blank 2008; Stahl 2005; Stahl and Gabel 2006; Zhang and Yang 1999). Feedback approaches offer the great opportunity to enhance a CBR approach during its operation. While a CBR approach could be trained *ex-ante* by domain experts linking semantically similar cases, from an economical perspective, this would be a waste of resources for two reasons. First, a CBR approach can already greatly support employees to some extent, even with a small case base and in the absence of feedback. Second, employees have to be released from work to label the required cases to train the CBR approach. In contrast, feedback approaches enable to leverage synergies, since employees profiting from a CBR approach, already identify the most suitable cases as basis for their solution. In addition, authors investigating feedback approaches in the area of CBR state that an on-the fly feedback generation is preferable to continuously improve the problem-solving competence of a CBR approach (Stahl 2003; Weis 2013). Depending of the specific domain, a continuous improvement in case retrieval can be crucial for the effectiveness of the approach. For instance, in case of a fast development of products or a high risk of frequently changing laws.

Nevertheless, only few researchers take feedback into account when developing their textual CBR approaches (Balakrishnan et al. 2016; Daniels and Rissland 1997; Weis 2013). Some authors (Balakrishnan et al. 2016; Weis 2013) use feedback from system users resulting in a re-ranking of previously retrieved cases. Balakrishnan et al. (2016), for instance, consider three different types of feedback, namely a

four-star rating, a referral with a dichotomous scale (i.e. Yes = 1, No = 0) and a textual comment. The textual comment is further analyzed to transform the text into a numerical rating by comparing specific key words within the comment against selected sentiment words. From this, a score is computed which is used to re-rank the retrieved cases. While the score is saved, it can only be utilized again if the very same query is performed again. Furthermore, Weis (2013) collects user annotations as feedback via CBR for re-ranking the answers of a question-answering approach. To do so, he represents cases as question-answer pairs in form of multilayered extended semantic networks which aim to represent the semantics of text by a graph structure. By this means, cases are retrieved and annotated to enrich the case base in order to collect relevant solutions. With this in mind, rank-optimizing decision trees are trained on features extracted from the case base and combined with existing answer validation features used by the initial question-answering approach. As a result, the feedback collected through the CBR approach enhances the performance of the question-answering system by re-ranking answers based on knowledge from the case base. Nevertheless, re-ranking approaches discard potentially relevant cases which have not been found by the initial retrieval. Therefore, a re-ranking approach does not seem suitable to find new cases within similar contexts as the query. In contrast, Daniels and Rissland (1997) use so-called Pseudo-Relevance Feedback (Salton and Buckley 1990) by assuming the top two retrieved cases as relevant. By doing so, the terms in the cases treated as relevant are added to the initial query resulting in a new query which is expected to lead to an improved retrieval performance of the textual CBR approach. To the best of our knowledge, besides these few examples, further studies in textual CBR literature do not focus on or consider feedback in depth.

To sum up, it seems very promising to foster human-machine collaboration in textual CBR by using feedback from users on case retrieval results. Thus, in the following we investigate further feedback approaches from the related area of information retrieval, showing appropriate characteristics to capture the semantic relationship between query and feedback in order to consider human knowledge on semantic similarity for new queries.

## Feedback in information retrieval

Research in information retrieval offers a wide range of feedback approaches for the retrieval of text documents which aim at taking humans' superior capability to semantically understand texts into account. Although the retrieval process in this context is similar to textual CBR, the objectives differ slightly. While textual CBR approaches aim at retrieving helpful solutions with respect to a full text description of a problem (Burke et al. 1997; Weber et al. 2005), approaches in information

retrieval try to retrieve relevant text documents regarding a query which expresses the user's request in a few keywords (Baeza-Yates and Ribeiro-Neto 1999). Nevertheless, approaches from information retrieval show high potential for adaption to textual CBR (Burke et al. 1997; Lenz et al. 1998b; Shekhar et al. 2014). Taking a feedback-oriented perspective, literature in information retrieval can particularly be classified into short-term (Rocchio 1971; Chen et al. 2006a; Lagun et al. 2013; Salton and Buckley 1990; Sarwar et al. 2018; Zhai and Lafferty 2001) and long-term feedback approaches (Crestani 1994, 2000; Mandl 2000; Lin et al. 2011; Mitra and Craswell 2017). Short-term feedback approaches can be characterized by using feedback only once for the query without storing feedback for further similar queries. Thus, these approaches require feedback for each query to enhance the retrieval of relevant text documents, even if the query is nearly identical to previous queries. In contrast, long-term feedback approaches are identified by the storage of feedback in order to conserve the expressed interconnections between queries and relevant text documents for later use. On this basis, a model can be trained on the collected feedback to improve the retrieval for new queries, without the need of new explicit feedback for each query.

One of the most well-known short-term feedback approaches is Relevance Feedback (Rocchio 1971; Salton and Buckley 1990): after an initial retrieval users decide for each retrieved document whether it is relevant to their query or not. Based on this feedback, an adapted query can be generated which results in more relevant documents being returned in a subsequent retrieval (Manning et al. 2008; Rocchio 1971). The main drawback of this and other short-term feedback approaches in general lies in their query-specific constitution. As feedback is not stored, further queries require new feedback; future retrievals do not benefit from already provided feedback. As collecting feedback from system users is a time-consuming task, other researchers concentrate on improving the query by simulating short-term feedback (Abderrahim 2013; Almasri et al. 2016; Buckley, Salton, Allan, & Singhal, 1995; Carpineto and Romano 2012; Xu et al. 2009). To do so, these authors rely on the established Pseudo-Relevance Feedback approach, treating the top-ranked documents in the initial retrieval results of the query as relevant (Buckley et al. 1995) and using terms from these documents to adapt the query (Abderrahim 2013; Buckley et al. 1995; Carpineto and Romano 2012; Xu et al. 2009). Although authors using Pseudo-Relevance Feedback report reasonable retrieval accuracy, the approach is based on the assumption that the top-ranked initially retrieved documents are indeed relevant. If this is not the case, it can even lead to worse retrieval accuracy (Cao et al. 2008; Lin et al. 2011).

Long-term feedback approaches store user feedback and use it for future retrievals (Crestani 1994, 2000; Mandl 2000; Lin et al. 2011; Mitra and Craswell 2017; Yin and Li

2006). A common way to collect user feedback on the semantic relationship between a query and a retrieved document is to explicitly ask the users to mark relevant results (Morrison et al. 2008), often through a rating scale (Yin and Li 2006). The collected feedback is stored in a feedback base to draw from the interrelationships for retrieval improvement (Heisterkamp 2002; Morrison et al. 2008; Yoshizawa and Schweitzer 2004). A common feedback augmentation strategy is clustering of semantically related documents based on the users' feedback (Chen et al. 2006b; Cord and Gosselin 2006; Crestani 1994; Morrison et al. 2008; Wang et al. 2006a; Wen et al. 2001; Yin et al. 2002). In the simplest case, all documents linked by a single user feedback are considered semantically related, hence comprising a cluster (Morrison et al. 2008; Wen et al. 2001). Once instantiated, semantic clusters are used to improve future retrievals. For example, in the approach by Jordan and Watters (2004) a query is matched to a semantic cluster by computing the similarity between query and the clusters so-called profile term vector. Yoshizawa and Schweitzer (2004) use the collected feedback to learn a distance metric, which places documents semantically related to a query closer to the query and vice versa. Feedback can also be used to learn a relationship between query and document terms, enabling query adaption aimed towards improving the retrieval of semantically similar documents (Cui et al. 2002). A popular approach based on long-term feedback is to learn an artificial neural network on the data contained in the feedback base, as these algorithms are able to learn complex mappings between patterns (Cöster and Asker 2000; Crestani 1994, 2000; Crestani and van Rijsbergen 1997; Fournier and Cord 2002; Huang et al. 2013; Lin et al. 2011; Mandl 2000; Mitra and Craswell 2017; Wang et al. 2006a). An overview of approaches using neural networks to capture the semantics of queries and retrieval results is given by Mitra and Craswell (2017). For instance, Lin et al. (2011), use neural networks to rank a set of query expansion terms according to their impact on retrieval performance. To train their model they use feedback from users which have associated a set of relevant terms with individual relevance scores to a given query. A similar approach is pursued by Crestani and van Rijsbergen (1997). Cöster and Asker (2000) use users' relevance feedback to train a neural network which predicts the difference of a given query to the optimal query. Other authors (Crestani 1994, 2000) trained neural networks based on tuples of queries and clusters of relevant documents, such that the trained network can be used to create improved queries. In contrast to these approaches concerned with adapting the query, Mandl (2000) uses a neural network to learn a cognitive similarity function based on human similarity judgments. As a result, the similarity between a query and documents is determined.

In summary, long-term relevance feedback approaches from information retrieval using the human capability to

understand and judge the semantic relationship between a query and retrieval results to enhance future retrievals seem a promising means to cope with the problem of exploiting human-machine collaboration in online customer service through a feedback-based textual CBR approach.

## Appendix II: Comparison to human-based approach

In this appendix, we detail the assumptions and estimates on which we based the comparison of our novel hybrid approach to an entirely human-based approach in terms of both effectiveness and efficiency.

Comparing our hybrid approach to an entirely human-based approach in terms of effectiveness is hardly possible without an additional field study where customers rate the answers phrased by a service expert from scratch. However, we argue that providing a service expert with semantically similar cases in addition to their own knowledge should not lower effectiveness in terms of providing the customer with the requested knowledge. To the contrary, providing employees with relevant knowledge to solve customer problems should reduce the limitations and errors inherent to employees phrasing solutions from scratch only based on their own knowledge. More precisely, with the help of our approach an employee having a false idea of the solution or missing some details can phrase a correct as well as more detailed solution, consequently providing a better solution to the customer. In the cases where a semantically similar customer problem is returned among the top five retrieved cases, the employee can subsequently reuse the solution. Hence, due to the high proportion of successful retrievals (97.94% of cases) our approach in most cases allows employees to phrase correct and high-quality solutions without knowing the (exact) solution from memory. Therefore, we are confident that our hybrid approach raises effectiveness of online customer service compared to entirely human-based customer service.

In order to compare our hybrid approach against an entirely human-based approach in terms of efficiency, we consider the time required for creating a solution for a customer problem. Starting with the human-based approach, the time required by a service employee to phrase a solution from scratch is comprised of the time required to read the customer problem, think about the solution and/or searching for additional knowledge, and typing up the solution. In contrast, the time required in our hybrid approach to create a solution for a customer problem is comprised of reading the customer problem, finding a solution based on the retrieved semantically similar customer problems, and adapting the solution. In the cases where no semantically similar past customer problem is retrieved, the employees need to follow the same steps as in the entirely manual approach. However, since our hybrid approach retrieves a semantically similar case within the top five cases in about

98% of cases (proportion of successful retrievals) the latter situation rarely arises. To demonstrate that on average the time required for creating a solution by the human-based approach is considerably more than with our hybrid approach, we step-wise estimate the time required by both approaches. Thereby, we favor the competing human-based approach by neglecting the time to think about phrasing a solution and to manually search for relevant knowledge.

With this in mind, we assume a service expert to be an average reader and a skilled typist, who is able to read 238 words per minute (Brysbaert 2019) and type 120 words per minute (Ayres 2005). Further, we calculate the average number of words contained in the customer problems of our data set (9.66 words per question) and consider the average number of words in answers on Quora, which is 473 with a rising trend (Rughiniş et al. 2014). On this basis, a service expert would require approximately 3 s to read the customer problem and close to 4 min to write an answer, resulting in a total time of about 4 minutes to provide a solution to a customer problem on average. In contrast, our hybrid approach requires a service expert to read the customer problem, wait for the retrieval of past customer problems, read the top five retrieved customer problems, and subsequently adapt and proofread the solution of an identified semantically similar customer problem. Based on the estimates above and very conservatively assuming an upper bound for the retrieval duration of 15 s this leads to a total time of 2:30 min. Accounting for the time required to manually write a solution in the 2% of cases where the retrieval was not successful, i.e. no semantically similar customer problem was retrieved, results in an average total time of about 2:35 min. If we further assume that on average it takes about 1 min to identify a semantically similar customer problem and adapt the solution (e.g. personalize the form of address), in the hybrid approach providing the solution to a customer problem takes on average about 3:30 min.

As a result, under these assumptions our hybrid approach requires at most about 85 % of the time a purely human based approach would require. This is despite the very generous assumption that the employees start to type their solution to an incoming customer problem immediately, without first pondering over the customer problem or researching further information, as well as the very conservative estimates for the time required for retrieval and adaption of the solution.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain



permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1), 39–59. <https://doi.org/10.3233/AIC-1994-7104>.
- Abderrahim, M. E. A. (2013). Concept based vs. Pseudo relevance feedback performance evaluation for information retrieval system. *International Journal of Computational Linguistics Research*, 4(4), 149–158.
- Acom, T. L., & Walden, S. H. (1992). *SMART: Support management automated reasoning technology for Compaq customer service*. In *Proceedings of the 4th Conference on Innovative Applications of Artificial Intelligence* (pp. 3–18). San Jose: CA.
- Almasri, M., Berrut, C., & Chevallet J.-P. (2016). A comparison of deep learning based query expansion with Pseudo-relevance feedback and mutual information. In *Proceedings of the 38th European Conference on Information Retrieval*. [https://doi.org/10.1007/978-3-319-30671-1\\_57](https://doi.org/10.1007/978-3-319-30671-1_57)
- Altitude & Spider Marketing (2016). The Omnichannel Evolution of Customer Experience. Retrieved from <https://http://bit.ly/The-Omnichannel-Evolution-of-Customer-Experience>
- Ashley, K. D. (1991). Reasoning with cases and hypotheticals in HYPO. *International Journal of Man-Machine Studies*, 34(6), 753–796. [https://doi.org/10.1016/0020-7373\(91\)90011-u](https://doi.org/10.1016/0020-7373(91)90011-u)
- Ayres, R. U. (2005). *On the reappraisal of microeconomics: economic growth and change in a material world*. Edward Elgar publishing. <https://doi.org/10.4337/9781845427948>
- Baeza-Yates, R., Ribeiro-Neto, B., & others (1999). *Modern information retrieval* (Vol. 463). New York: ACM Press. ISBN-13: 978-0321416919.
- Balakrishnan, V., Ahmadi, K., & Ravana, S. D. (2016). Improving retrieval relevance using users' explicit feedback. *Aslib Journal of Information Management*, 68(1), 76–98. <https://doi.org/10.1108/AJIM-07-2015-0106>
- Baskerville, R., Baiyere, A., Gregor, S., Hevner, A. R., & Rossi, M. (2018). Design science Research contributions: Finding a balance between artifact and theory. *Journal of the Association for Information Systems*, 19(5), 358–376. <https://doi.org/10.17705/1jais.00495>
- Bedué, P., Graef, R., Klier, M., & Zolitschka, J. F. (2018). A novel hybrid knowledge retrieval approach for online customer service platforms. In *Proceedings of the 26th European Conference on Information Systems*.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34–43.
- Bodnick, M. (2015). Quora & the importance of canonical questions. Retrieved from <https://blog.quora.com/Quora-the-importance-of-canonical-questions>
- Branting, L. K. (2001). Acquiring customer preferences from return-set selections. In D. W. Aha & I. Watson (chairs), *Case-Based Reasoning Research and Development: Proceedings of the 4th International Conference on Case-Based Reasoning*. [https://doi.org/10.1007/3-540-44593-5\\_5](https://doi.org/10.1007/3-540-44593-5_5)
- Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, 109, 104047. <https://doi.org/10.1016/j.jml.2019.104047>
- Buckley, C., Salton, G., Allan, J., & Singhal, A. (1995). Automatic query expansion using SMART: TREC 3. *NIST Special Publication (SP)*, 69–80.
- Burke, R., Hammond, K., Kulyukin, V., Lytinen, S., Tomuro, N., & Schoenberg, S. (1997). Question answering from frequently asked question files: Experiences with the FAQ FINDER system. *AI Magazine*, 18(2), 57–66. <https://doi.org/10.1609/aimag.v18i2.1294>
- Cao, G., Nie, J.-Y., Gao, J., & Robertson, S. (2008). Selecting good expansion terms for Pseudo-relevance feedback. In *Proceedings of the 31st Conference on Research and Development in Information Retrieval*. Symposium conducted at the meeting of: ACM. <https://doi.org/10.1145/1390334.1390377>
- Carpineto, C., & Romano, G. (2012). A survey of automatic query expansion in information retrieval. *ACM Computing Surveys*, 44(1), 1–50. <https://doi.org/10.1145/2071389.2071390>
- Chen, S.-M., Lin, H.-C. Hsi-Ching, Chang, Y.-C., & others (2006a). A new method for query reweighting for document retrieval based on neural networks. *International Journal of Information and Management Sciences*, 17(4), 95–110.
- Chen, Y., Rege, M., Dong, M., & Fotouhi, F. (2006b). Deriving Semantics for Image Clustering from Accumulated User Feedbacks. In *Proceedings of the 15th conference on Multimedia*. <https://doi.org/10.1145/1291233.1291300>
- Cheng, W., & Hüllermeier, E. (2008). Learning similarity functions from qualitative feedback. In *Proceedings of the 9th European Conference on Case-Based Reasoning*. [https://doi.org/10.1007/978-3-540-85502-6\\_8](https://doi.org/10.1007/978-3-540-85502-6_8)
- Chung, K.-P., Wong, K. W., & Fung C.-C. (2006). Reducing user log size in an inter-query learning content based image retrieval (CBIR) system with a cluster merging approach. In *The 2006 IEEE International Joint Conference on Neural Network*. <https://doi.org/10.1109/IJCNN.2006.246825>
- Cord, M., & Gosselin, P. (2006). Image retrieval using long-term semantic learning. In *2006 International Conference on Image Processing*. <https://doi.org/10.1109/icip.2006.313127>
- Cöster, R., & Asker, L. (2000). A similarity-based approach to relevance learning. In *Proceedings of the 14th European Conference on Artificial Intelligence*.
- Coyle, L., & Cunningham, P. (2003). Exploiting re-ranking information in a case-based personal travel assistant. In *Proceedings of the 5th International Conference on Case-Based Reasoning*.
- Crestani, F. (1994). Domain knowledge Acquisition for Information Retrieval using neural networks. *Journal of Applied Expert Systems*, 2(2), 101–116.
- Crestani, F. (2000). Neural relevance feedback for information retrieval. In B. Bouchon-Meunier, L. A. Zadeh, & R. Y. Yager (Eds.), *Uncertainty in intelligent and information systems* (pp. 197–208). Singapore: World Scientific. [https://doi.org/10.1142/9789812792563\\_0016](https://doi.org/10.1142/9789812792563_0016)
- Crestani, F., & van Rijsbergen, C. J. (1997). A model for adaptive information retrieval. *Journal of Intelligent Information Systems*, 8(1), 29–56. <https://doi.org/10.1023/A:1008601616486>
- Cui, H., Wen, J.-R., Nie, J.-Y., & Ma, W.-Y. (2002). Probabilistic query expansion using query logs. In D. Lassner, D. de Roure, & a. Iyengar (chairs), *Proceedings of the 11th International Conference on World Wide Web*. <https://doi.org/10.1145/511446.511489>
- Cunningham, C., Weber, R. O., Proctor, J. M., Fowler, C., & Murphy, M. (2004). Investigating graphs in textual case-based reasoning. In *Proceedings of the 7th European Conference on Case-Based Reasoning*. [https://doi.org/10.1007/978-3-540-28631-8\\_42](https://doi.org/10.1007/978-3-540-28631-8_42)
- Daniels, J. J., & Rissland, E. L. (1997). Integrating IR and CBR to locate relevant texts and passages. In *Proceedings of the 8th International Workshop on Database and Expert Systems Applications*. <https://doi.org/10.1109/dexa.1997.617270>
- Dellermann, D., Lipusch, N., Ebel, P., & Leimeister, J. M. (2018). Design principles for a hybrid intelligence decision support system for

- business model validation. *Electronic Markets*, 1–19. <https://doi.org/10.1007/s12525-018-0309-2>
- El-Sappagh, S. H., & Elmoggy, M. (2015). Case based reasoning: Case representation methodologies. *International Journal of Advanced Computer Science and Applications*, 6(11), 192–208. <https://doi.org/10.14569/ijacsa.2015.061126>
- Embley, D. W. (2004). Toward semantic understanding: An approach based on information extraction ontologies. In *Proceedings of the 15th Australasian Database Conference. Inc: Symposium conducted at the meeting of Australian Computer Society*.
- Forrester (2016). Your Customers Don't Want To Call You For Support. Retrieved from <https://http://bit.ly/Your-Customers-Dont-Want-To-Call-You-For-Support>
- Forrester. (2018). *2018 Customer Service Trends: How Operations Become Faster, Cheaper — And Yet, More Human* Retrieved from <https://http://bit.ly/2018-Customer-Service-Trends>.
- Fournier, J., & Cord, M. (2002). Long-term similarity learning in content-based image retrieval. In *Proceedings. International Conference on Image Processing*. <https://doi.org/10.1109/icip.2002.1038055>
- Gabel, T., & Stahl, A. (2004). Exploiting background knowledge when learning similarity measures. In *Proceedings of the 7th European Conference on Case-Based Reasoning*. Symposium conducted at the meeting of: Springer. [https://doi.org/10.1007/978-3-540-28631-8\\_14](https://doi.org/10.1007/978-3-540-28631-8_14)
- Gladly (2018). *Customer Service Expectations Survey: Trends and insights from consumers about customer service*. Retrieved from <https://http://bit.ly/Customer-Service-Expectations-Survey>
- Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In G. Gordon, D. Dunson, & M. Dudik (chairs), *International Conference on Artificial Intelligence and Statistics*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, Massachusetts, London, England: MIT Press. ISBN: 0262035618.
- Gregor, S. (2009). Building theory in the sciences of the artificial. In V. Vaishnavi & S. Purao (Eds.), *Proceedings of the 4th international conference on design science Research in information systems and technology*. New York, New York, USA: ACM Press. <https://doi.org/10.1145/1555619.1555625>
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science Research for maximum impact. *MIS Quarterly*, 37(2), 337–355. <https://doi.org/10.25300/MISQ/2013/37.2.01>
- Gu, D., Li, J., Bichindaritz, I., Deng, S., & Liang, C. (2017). The mechanism of influence of a case-based health knowledge system on hospital management systems. In *Proceedings of the 25th International Conference on Case-Based Reasoning*. [https://doi.org/10.1007/978-3-319-61030-6\\_10](https://doi.org/10.1007/978-3-319-61030-6_10)
- Guzmán, I., & Pathania, A. (2016). Chatbots in Customer Service. Retrieved from <https://http://bit.ly/Accenture-Chatbots-Customer-Service>
- Hammond, K., Burke, R., Martin, C., & Lytinen, S. (1995). FAQ finder: A case-based approach to knowledge navigation. In *Proceedings of the 11th Conference on Artificial Intelligence for Applications*. <https://doi.org/10.1109/caia.1995.378787>
- Heisterkamp, D. R. (2002). Building a latent semantic index of an image database from patterns of relevance feedback. In *16th International Conference on Pattern Recognition*. <https://doi.org/10.1109/icpr.2002.1047417>
- Heras, S., García-Pardo, J. Á., Ramos-Garijo, R., Palomares, A., Botti, V., Rebollo, M., & Julián, V. (2009). Multi-domain case-based module for customer support. *Expert Systems with Applications*, 36(3), 6866–6873. <https://doi.org/10.1016/j.eswa.2008.08.003>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems Research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Hua, J., Tembe, W. D., & Dougherty, E. R. (2009). Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognition*, 42(3), 409–424. <https://doi.org/10.1016/j.patcog.2008.08.001>
- Huang, P.-S., He, X., Gao, J., Deng, L., Acero, A., & Heck, L. (2013). Learning deep structured semantic models for web search using Clickthrough data. In *Proceedings of the 22nd International Conference on Information and Knowledge Management*. Symposium conducted at the meeting of: ACM. <https://doi.org/10.1145/2505515.2505665>
- Iyer, S., Dandekar, N., & Csernai, K. (2017). First Quora Dataset Release: Question Pairs. Retrieved from <https://data.quora.com/First-Quora-Dataset-Release-Question-Pairs>
- Jayanthi, K., Chakraborti, S., & Massie, S. (2010). Introspective knowledge revision in textual case-based reasoning. In *Proceedings of the 18th International Conference on Case-Based Reasoning*. [https://doi.org/10.1007/978-3-642-14274-1\\_14](https://doi.org/10.1007/978-3-642-14274-1_14)
- Jordan, C., & Watters, C. (2004). Extending the Rocchio relevance feedback algorithm to provide contextual retrieval. In *Proceedings of the 2nd International Atlantic Web Intelligence Conference*. [https://doi.org/10.1007/978-3-540-24681-7\\_16](https://doi.org/10.1007/978-3-540-24681-7_16)
- Jung, S., Herlocker, J. L., & Webster, J. (2007). Click data as implicit relevance feedback in web search. *Information Processing and Management*, 43(3), 791–807. <https://doi.org/10.1016/j.ipm.2006.07.021>
- Khanapure, V. M., & Chirchi, V. R. (2013). iAssist: An Intelligent Online Assistance System. *International Journal of Scientific and Research Publications*, 3(2). <https://doi.org/10.1109/64.248349>
- Kriegsman, M., & Barletta, R. (1993). Building a case-based help desk application. *IEEE Expert*, 8(6), 18–26.
- Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyber, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 141–164). Wiley. <https://doi.org/10.1002/9781118490013.ch6>
- Kunze, M., & Hübner, A. (1998). CBR on semi-structured documents: The experience book and the FAILQ project. In *Proceedings of 6th German Workshop on Case-Based Reasoning*.
- Lagun, D., Sud, A., White, R. W., Bailey, P., & Buscher, G. (2013). Explicit feedback in local search tasks. In *Proceedings of the 36th International Conference on Research and Development in Information Retrieval*. <https://doi.org/10.1145/2484028.2484123>
- Leake, D., & Dial, S. A. (2008). Using case provenance to propagate feedback to cases and adaptations. In *Proceedings of the 9th European Conference on Case-Based Reasoning*. [https://doi.org/10.1007/978-3-540-85502-6\\_17](https://doi.org/10.1007/978-3-540-85502-6_17)
- Lenz, M., & Burkhard, H.-D. (1997). CBR for document retrieval: The FAILQ project, In *Proceedings of the 2nd International Conference of Case-Based Reasoning Research and Development* (pp. 84–93). USA: Rhode Island. [https://doi.org/10.1007/3-540-63233-6\\_481](https://doi.org/10.1007/3-540-63233-6_481)
- Lenz, M., Bartsch-Spörl, B., Burkhard, H.-D., & Wess, S. (Eds.). (1998a). *Lecture notes in computer science: Vol. 1400. Case-based reasoning technology: From foundations to applications*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/3-540-69351-3>
- Lenz, M., Hübner, A., & Kunze, M. (1998b). Textual CBR. In M. Lenz, B. Bartsch-Spörl, H.-D. Burkhard, & S. Wess (Eds.), *Lecture notes in computer science, Case-based reasoning technology: From foundations to applications* (Vol. 1400, pp. 115–137). Berlin, Heidelberg: Springer. [https://doi.org/10.1007/3-540-69351-3\\_5](https://doi.org/10.1007/3-540-69351-3_5)
- Lenz, M., Busch, K.-H., Hübner, A., & Wess, S. (1999). The Simatic knowledge manager. In D. Aha, I. Becerra-Fernandez, F. Maurer, & H. Muñoz-Avila (Eds.), *Exploring synergies of knowledge management and case-based reasoning. Proceedings of the AAAI workshop* (pp. 40–45). Menlo Park, California: AAAI Press.
- Liao, T. W., Zhang, Z., & Mount, C. R. (1998). Similarity measures for retrieval in case-based reasoning systems. *Applied Artificial Intelligence*, 12(4), 267–288. <https://doi.org/10.1080/088395198117730>

- Lin, Y., Lin, H., Jin, S., & Ye, Z. (2011). Social annotation in query expansion: A machine learning approach. In *Proceedings of the 34th International Conference on Research and Development in Information Retrieval* (pp. 405–414). New York. <https://doi.org/10.1145/2009916.2009972>
- Mandl, T. (2000). Tolerant information retrieval with backpropagation networks. *Neural Computing and Applications*, 9(4), 280–289. <https://doi.org/10.1007/s005210070005>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York: Cambridge University Press. <https://doi.org/10.1017/cbo9780511809071>
- Martin, A., Emmenegger, S., Hinkelmann, K., & Thönssen, B. (2017). A viewpoint-based case-based reasoning approach Utilising an Enterprise architecture ontology for experience management. *Enterprise Information Systems*, 11(4), 551–575. <https://doi.org/10.1080/17517575.2016.1161239>
- Mero, J. (2018). The effects of two-way communication and chat service usage on consumer attitudes in the E-commerce retailing sector. *Electronic Markets*, 28(2), 205–217. <https://doi.org/10.1007/s12525-017-0281-2>
- Microsoft (2018). State of Global Customer Service Report. Retrieved from <https://bit.ly/State-of-Global-Customer-Service-Report>
- Mitra, B., & Craswell, N. (2017). Neural models for information retrieval. *ArXiv Preprint ArXiv*, 1705, 01509.
- Morrison, D., Marchand-Maillet, S., & Bruno, E. (2008). Semantic clustering of images using patterns of relevance feedback. In *Proceedings of the 6th International Workshop on Content-Based Multimedia Indexing*. <https://doi.org/10.1109/cbmi.2008.4564964>
- Moschitti, A. (2003). A study on optimal parameter tuning for Rocchio text classifier. In G. Goos, J. Hartmanis, J. van Leeuwen, & F. Sebastiani (Eds.), *Lecture Notes in Computer Science. Advances in Information Retrieval* (Vol. 2633, pp. 420–435). Berlin, Heidelberg: Springer. [https://doi.org/10.1007/3-540-36618-0\\_30](https://doi.org/10.1007/3-540-36618-0_30)
- Ng, A. (2018). Machine learning yearning: Technical strategy for AI engineers in the era of deep learning.
- Parature (2014). *2014 State of Multichannel Customer Service Study*. Retrieved from <https://bit.ly/2014-State-of-Multichannel-Customer-Service-Study>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science Research methodology for information systems Research. *Journal of Management Information Systems*, 24(3), 45–77.
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137. <https://doi.org/10.1108/eb046814>
- Prechelt, L. (2012). Early stopping — But when? In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Lecture notes in computer science* (pp. 53–67). Neural Networks: Tricks of the Trade. [https://doi.org/10.1007/978-3-642-35289-8\\_5](https://doi.org/10.1007/978-3-642-35289-8_5)
- Reuss, P., Althoff, K.-D., Henkel, W., Pfeiffer, M., Hankel, O., & Pick, R. (2015). Semi-automatic knowledge extraction from semi-structured and unstructured data within the OMAHA project. In *Proceedings of the 23rd International Conference on Case-Based Reasoning*. [https://doi.org/10.1007/978-3-319-24586-7\\_23](https://doi.org/10.1007/978-3-319-24586-7_23)
- Rocchio, J. J. (1971). *Relevance feedback in information retrieval* (pp. 313–323). The SMART Retrieval System: Experiments in Automatic Document Processing. Englewood Cliffs: Prentice-Hall.
- Rughiniş, R., Marinescu-Nenciu, A. P., Matei, Ş., & Rughiniş, C. (2014). Computer-supported collaborative questioning. Regimes of online sociality on Quora. In *2014 9th Iberian conference on information systems and technologies (CISTI)*. Symposium conducted at the meeting of IEEE. <https://doi.org/10.1109/cisti.2014.6876946>
- Russell, S. J., & Norvig, P. (2010). *Ai: A Modern Approach* (3rd edn): Pearson education. ISBN-13: 978-1292153964.
- Salesforce Research (2016). *State of the Connected Customer*. Retrieved from <https://bit.ly/State-of-the-Connected-Customer-first-edition>
- Salesforce Research (2018). *State of the Connected Customer*. Retrieved from <https://bit.ly/State-of-the-Connected-Customer-second-edition>
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5), 513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41(1), 288–297. [https://doi.org/10.1002/\(sici\)1097-4571\(199006\)41:4%3C288::aid-asi8%3E3.0.co;2-h](https://doi.org/10.1002/(sici)1097-4571(199006)41:4%3C288::aid-asi8%3E3.0.co;2-h)
- Salton, G., & McGill, M. J. (1984). *Introduction to modern information retrieval*. New York: McGraw-Hill Book Company. ISBN: 0-07-054484-0.
- Salton, G., Wong, A., & Yang, C.-S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613–620. <https://doi.org/10.1145/361219.361220>
- Sarwar, S. M., Foley, J., & Allan, J. (2018). Term relevance feedback for contextual named entity retrieval. In *Proceedings of the 3rd Conference on Human Information Interaction and Retrieval*. <https://doi.org/10.1145/3176349.3176886>
- Scharff, L. (2015). Introducing Question Merging. Retrieved from <https://blog.quora.com/Introducing-Question-Merging>
- Sebastiani, F. (2002). Machine learning in automated text categorization. *Computing Surveys*, 34(1), 1–47. <https://doi.org/10.1145/505282.505283>
- Shekhar, S., Chakraborti, S., & Khemani, D. (2014). Linking cases up: An extension to the case retrieval network. In *Proceedings of the 22nd International Conference on Case-Based Reasoning Research and Development*. [https://doi.org/10.1007/978-3-319-11209-1\\_32](https://doi.org/10.1007/978-3-319-11209-1_32)
- Simoudis, E. (1992). Using case-based retrieval for customer technical support. *IEEE Expert*, 7(5), 7–12. <https://doi.org/10.1109/64.163667>
- Sizov, G., Öztürk, P., & Aamodt, A. (2015). Evidence-driven retrieval in textual CBR: Bridging the gap between retrieval and reuse. In *Proceedings of the 23rd International Conference on Case-Based Reasoning Research and Development*. Symposium conducted at the meeting of: Springer. [https://doi.org/10.1007/978-3-319-24586-7\\_24](https://doi.org/10.1007/978-3-319-24586-7_24)
- Soh, L.-K., & Blank, T. (2008). Integrating case-based reasoning and meta-learning for a self-improving intelligent tutoring system. *International Journal of Artificial Intelligence in Education*, 18(1), 27–58.
- Sonnenberg, C., & vom Brocke J. (2012). Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, . . . B. Kuechler (Eds.), *Lecture Notes in Computer Science. Design Science Research in Information Systems. Advances in Theory and Practice* (Vol. 7286, pp. 381–397). Berlin, Heidelberg: Springer [https://doi.org/10.1007/978-3-642-29863-9\\_28](https://doi.org/10.1007/978-3-642-29863-9_28)
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from Overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958.
- Stahl, A. (2003). *Learning of knowledge-intensive similarity measures in case-based reasoning*. The University of Kaiserslautern, Kaiserslautern, Germany: Doctoral dissertation.
- Stahl, A. (2005). Learning similarity measures: A formal view based on a generalized CBR model. In *Proceedings of the 6th International Conference on Case-Based Reasoning Research and Development*. Symposium conducted at the meeting of: Springer. [https://doi.org/10.1007/11536406\\_39](https://doi.org/10.1007/11536406_39)
- Stahl, A., & Gabel, T. (2006). Optimizing similarity assessment in case-based reasoning. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2 AAAI Press*, Boston, MA, pp. 1667–1670.



- Statista (2017). Most Popular Channels. Retrieved from <https://bit.ly/Most-Popular-Channels>
- Trstenjak, B., & Donko, D. (2016). Case-based reasoning: A hybrid classification model improved with an Expert's knowledge for high-dimensional problems. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 10(6), 1184–1190. <https://doi.org/10.3233/his-160233>
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37, 141–188. <https://doi.org/10.1613/jair.2934>
- Turtle, H., & Croft, W. B. (1990). Inference networks for document retrieval. In *Proceedings of the 13th International Conference on Research and Development in Information Retrieval*. <https://doi.org/10.1145/96749.98006>
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science Research. *European Journal of Information Systems*, 25(1), 77–89. <https://doi.org/10.1057/ejis.2014.36>
- Wacker, J. (2016). Question Merging: Updates. Retrieved from <https://productupdates.quora.com/Question-Merging-Updates>
- Wang, B., Zhang, X., & Li, N. (2006a). Relevance Feedback Technique for Content-Based Image Retrieval using Neural Network Learning. In *Proceedings of the 5th International Conference on Machine Learning and Cybernetics*. <https://doi.org/10.1109/icmlc.2006.258628>
- Wang, K., Qi, L., & Zhong, Q. (2006b). A Research on improvement of customer Service Systems in Mobile Telecommunication Enterprises: A knowledge classification perspective. In *Proceedings of the 2nd International Conference on Service Operations and Logistics, and Informatics*. <https://doi.org/10.1109/soli.2006.328946>
- Wang, D., Li, T., Zhu, S., & Gong, Y. (2011). iHelp: An intelligent online helpdesk system. *IEEE Transactions on Systems, Man, and Cybernetics*, 41(1), 173–182. <https://doi.org/10.1109/tsmcb.2010.2049352>
- Wang, G., Gill, K., Mohanlal, M., Zheng, H., & Zhao B. Y., (2013). Wisdom in the social crowd: An analysis of Quora. In *International Conference on World Wide Web*. <https://doi.org/10.1145/2488388.2488506>
- Weber, R. O., Ashley, K. D., & Brüninghaus, S. (2005). Textual case-based reasoning. *The Knowledge Engineering Review*, 20(3), 255–260. <https://doi.org/10.1017/s0269888906000713>
- Weis, K.-H. (2013). A case based reasoning approach for answer Reranking in question answering. In *INFORMATIK 2013 – Informatik angepasst an Mensch, Organisation und Umwelt*. Bonn: Gesellschaft für Informatik e.V., pp. 93–104
- Wen, J.-R., Nie, J.-Y., & Zhang, H.-J. (2001). Clustering User Queries of a Search Engine. In *Proceedings of the 10th International Conference on World Wide Web*. <https://doi.org/10.1145/371920.371974>
- Wilson, D. C., & Bradshaw, S. (1999). CBR Textuality. In *Proceedings of the 4th UK Case-Based Reasoning Workshop*.
- Xu, Y., Jones, G. J. F., & Wang, B. (2009). Query dependent Pseudo-relevance feedback based on Wikipedia. In *Proceedings of the 32nd International Conference on Research and Development in Information Retrieval*. Symposium conducted at the meeting of ACM. <https://doi.org/10.1145/1571941.1571954>
- Yan, A., Qian, L., & Zhang, C. (2014). Memory and forgetting: An improved dynamic maintenance method for case-based reasoning. *Information Sciences*, 287, 50–60. <https://doi.org/10.1016/j.ins.2014.07.040>
- Yin, P.-Y., & Li, S.-H. (2006). Content-based image retrieval using association rule mining with soft relevance feedback. *Journal of Visual Communication and Image Representation*, 17(5), 1108–1125. <https://doi.org/10.1016/j.jvcir.2006.04.004>
- Yin, P.-Y., Bhanu, B., Chang, K.-C., & Dong, A. (2002). Improving retrieval performance by long-term relevance information. In *Proceedings of the 16th International Conference on Pattern Recognition*. <https://doi.org/10.1109/icpr.2002.1047994>
- Yoshizawa, T., & Schweitzer, H. (2004). Long-term learning of semantic grouping from relevance-feedback. In *Proceedings of the 6th International Workshop on Multimedia Information Retrieval*. <https://doi.org/10.1145/1026711.1026739>
- Zendesk (2017). *The Multi-Channel Customer Care Report: Meeting the Fresh Demands of Multi-Channel Customers*. Retrieved from <https://bit.ly/Multi-channel-Customer-Care-Report>
- Zhai, C., & Lafferty J., (2001). Model-based feedback in the language modeling approach to information retrieval. In *Proceedings of the 10th International Conference on Information and Knowledge Management*. <https://doi.org/10.1145/502585.502654>
- Zhang, Z., & Yang, Q. (1999). Dynamic refinement of feature weights using quantitative Introspective Learning. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.