This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

Conversational Context-Sensitive Ad Generation With a Few Core-Queries

RYOICHI SHIBATA, Keio University, Japan
SHOYA MATSUMORI, Keio University, Japan
YOSUKE FUKUCHI, National Institute of Informatics, Japan
TOMOYUKI MAEKAWA, Keio University, Japan
MITSUHIKO KIMOTO, ATR, Japan and Keio University, Japan
MICHITA IMAI, Keio University, Japan

When people are talking together in front of digital signage, advertisements that are aware of the context of the dialogue will work the most effectively. However, it has been challenging for computer systems to retrieve the appropriate advertisement from among the many options presented in large databases. Our proposed system, the Conversational Context-sensitive Advertisement generator (CoCoA), is the first attempt to apply masked word prediction to web information retrieval that takes into account the dialogue context. The novelty of CoCoA is that advertisers simply need to prepare a few abstract phrases, called Core-Queries, and then CoCoA automatically generates a context-sensitive expression as a complete search query by utilizing a masked word prediction technique that adds a word related to the dialogue context to one of the prepared Core-Queries. This automatic generation frees the advertisers from having to come up with context-sensitive phrases to attract users' attention. Another unique point is that the modified Core-Query offers users speaking in front of the CoCoA system a list of context-sensitive advertisements. CoCoA was evaluated by crowd workers regarding the context-sensitivity of the generated search queries against the dialogue text of multiple domains prepared in advance. The results indicated that CoCoA could present more contextual and practical advertisements than other web-retrieval systems. Moreover, CoCoA acquired a higher evaluation in a particular conversation that included many travel topics to which the Core-Queries were designated, implying that it succeeded in adapting the Core-Queries for the specific ongoing context better than the compared method without any effort on the part of the advertisers. In addition, case studies with users and advertisers revealed that the context-sensitive advertisements generated by CoCoA also had an effect on the content of the ongoing dialogue. Specifically, since pairs unfamiliar with each other more frequently referred to the advertisement CoCoA displayed, the advertisements had an effect on the topics about which the pairs spoke. Moreover, participants of an advertiser role recognized that some of the search queries generated by CoCoA fitted the context of a conversation and that CoCoA improved the effect of the advertisement. In particular, they assimilated the hang of designing a good Core-Query at ease by observing the users' response to the advertisements retrieved with the generated search queries.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**; **User studies**.

Additional Key Words and Phrases: dialogue, context, advertisement, query generation, mask prediction

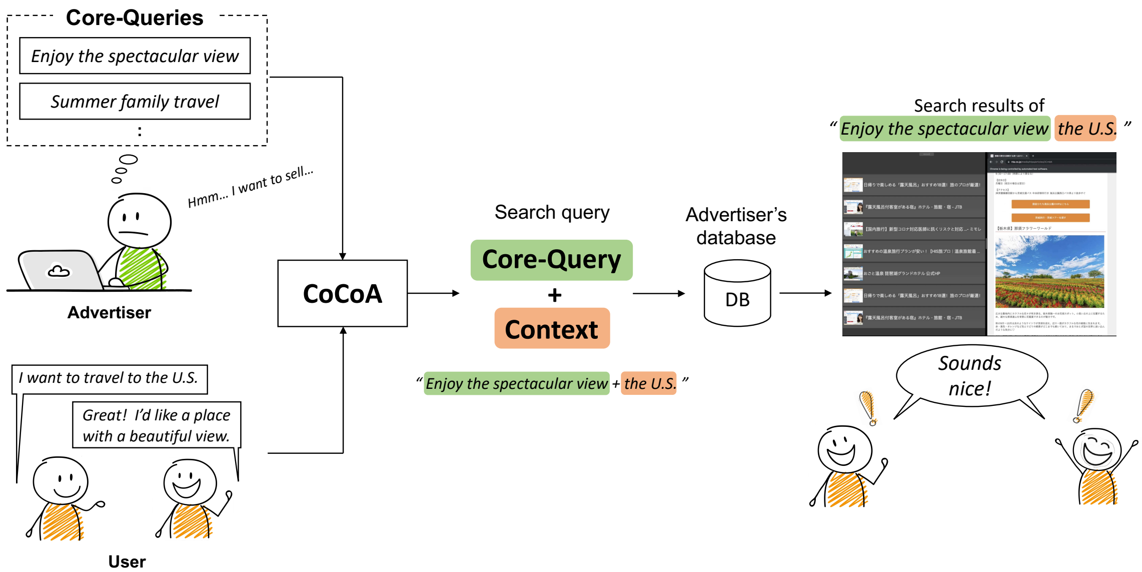

Fig. 1. Example of CoCoA in use. CoCoA refers to the user's dialogue context and selects "Enjoy the spectacular view" from several Core-Queries created by the advertiser because the user's dialogue contains "a beautiful view". In addition, the term "the U.S." in the dialogue is appended to the Core-Query to generate a search query. Then, the database prepared by the advertiser is searched using the search query. The search results relevant to the dialogue context are then presented to the user.

1 INTRODUCTION

Intelligent communication systems such as smartphones, wearable devices, and voice assistants are becoming more and more a part of our daily lives. These systems make it possible to present ads in a context-sensitive manner. Existing systems have utilized time- or user-dependent information as the context, such as weather [7], location information [21, 35], and video viewing history [30]. On the other hand, to the best of our knowledge, there are no systems that utilize the contents of real spoken dialogue as the context or present advertisements relevant to the context in real time. Conversational dialogues contain a variety of topics, and the topic often changes in a relatively short period of time, which makes it extremely difficult to design advertisements that match the content of the dialogues in advance. For example, even in a dialogue limited to the travel domain, various topics such as destination, budget, duration, season, etc. may be discussed as the dialogue progresses. Our key idea in this study is that advertisements can be optimized for individual users if we dynamically transform the advertisements according to the context of the dialogue between two users. To this end, we propose CoCoA, a system that dynamically presents advertisements that match the context of the dialogue. An example of CoCoA in use is shown in Fig. 1. First, the advertiser prepares several Core-Queries, which represent the core messages he wants to convey. When users talk with each other in front of a digital signage on which CoCoA is running, CoCoA selects the appropriate Core-Query given the context of the dialogue and then appends an appropriate word from the dialogue to it to generate a search query that is used to extract advertisements from an advertisement database. This search query is used to advertise the corresponding product from the advertiser's product database. In this way, CoCoA is expected to present effective advertisements by generating search queries that fit the dialogue. Let us consider a scenario in a restaurant to illustrate the benefits CoCoA offers to users. Recently, many

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

restaurants have started using tablet devices on tables to display menus and handle customers' orders. Our system can attract customers' interest by displaying context-aware advertisements on the tablets. From a user's perspective, the advertisements related to the current conversation can be fruitful information to expand their conversation at dining. In addition, offering restaurant coupons, which are funded by arbitrary advertisers, can encourage the use of our system.

The novelty of this paper lies in the following two points.

- **Defining the search query.** We define the search query as the combination of "the phrase that the advertiser wants to convey" and "the keywords in the dialogue". This definition overcomes the difficulty of generating a search query that fits the dialogue. The phrase is called the "Core-Query", since it contains the core message of the search query, and the keyword is referred to as the "complementary word", as it depends on the context in the dialogue. Since there are a variety of candidates for the complementary word, it is difficult to prepare a complete set of search queries applicable for any dialogue in advance. Therefore, we divide the search query into two parts and consider them separately. The appropriate Core-Query according to the context is selected from a set of several Core-Queries prepared in advance, and the appropriate complementary word is selected from the dialogue history by using a Query Completer. The search query is then generated by concatenating the Core-Query and the complementary word. This is how the Core-Query is transformed into a practical search query. Even though the Core-Query itself is an abstract phrase, the generated search query contains the concrete complementary word that is in line with the dialogue. Thus, the search query can come up with an advertisement that effectively attracts the user.
- **Applying mask prediction to information retrieval.** The second point is that the extraction of keywords in the dialogue, which are complementary words, is achieved by using mask prediction. Mask prediction, which predicts words that are hidden in a sentence, is known to be effective in learning language models such as BERT, but it has not been used much except for the purpose of acquiring language models. Our research is the first to implement the mechanism of mask prediction into a framework for information retrieval that takes into account the dialogue context. As a result, CoCoA can continue to dynamically present advertisements that are relevant to the dialogue from a small number of Core-Queries as the dialogue progresses.

We conducted two evaluations to investigate the design of CoCoA. The first was a crowdsourced evaluation utilizing 600 workers to examine the appropriateness of the proposed function of finding context-sensitive complementary words against target Core-Queries in terms of the subjective interpretation of participants. The appropriateness was evaluated from two perspectives. First, we conducted a context sensitivity test to determine whether CoCoA could generate appropriate search queries for a given context. We found that CoCoA outperformed two other methods that either did not have a complementary word or had a word selected by Google Suggest. Second, we conducted a robustness test to evaluate the system's robustness by having each participant create a query to see if the system could adequately generate advertisements for all ad creators and all Core-Queries. We found that it was able to do so regardless of the variations of Core-Queries.

The second evaluation was a descriptive study [48] consisting of two case studies. The objective here was to clarify how pairs of participants interacted with CoCoA while conversing and how the advertisers evaluated the effect of the Core-Queries they created on the actual conversation between users. Usability tests in a laboratory setting provided us with insights into the effect of the advertisements generated by CoCoA on both the conversation and behaviors of the participants. We first conducted a user case study with six people (three pairs) in which we observed how the search query generated by CoCoA in real time during user conversations affected the dialogue. We found that the frequency of system usage changed depending on the level of familiarity between the individuals in a pair, and that the domain of the dialogue changed depending on the system. Second, we conducted an advertiser case study with three groups of one advertiser and two users in which we asked the advertiser to observe the interaction between individuals who were using CoCoA to evaluate its usefulness in actual conversations from the advertiser's viewpoint. The advertiser created several Core-Queries in advance and CoCoA then used them to generate Search queries. These queries were displayed during the conversations of the users and then evaluated by the advertiser. We found that, from the advertiser' s perspective, CoCoA generated search queries that were more contextually relevant and effective than the compared method, which was prepared in the same way as the crowdsourced evaluation. This article constitutes an extension of our original paper [38] published in the Proceedings of the 2022 Conference on Intelligent User Interfaces. The main addition is the "advertiser case study" in Section 5.

This paper is organized as follows. Section 2 describes related work in terms of advertising methods and query expansion. Section 3 proposes our system named CoCoA and explains in detail how it generates search queries from dialog history. Section 4 describes two experiments to evaluate whether CoCoA generates appropriate search queries for given dialogues and discusses the results. Section 5 reports two case studies of face-to-face human conversations using CoCoA and discusses the results. Section 6 discusses future work on our research, and Section 7 concludes the paper.

2 RELATED WORK

2.1 Interactive advertising

Digital displays have been increasing in public spaces over the past several years. One of the advantages of the digital displays is that the content can be changed in real time and therefore information can be presented in a way that is more impactful to the user. Such displays are now frequently being used to present advertisements [34]. Digital displays are attractive to advertisers because they can present video images that are more likely to draw user interest than traditional paper-based, unchanging text and images [16, 31]. Moreover, by introducing interactive advertisements that allow users to not only look at an ad but also touch it, appropriate advertisements can be presented in response to specific user desires, thereby increasing the likelihood of a sale [4, 15]. The interactive ads that have been implemented on TV [5] have demonstrated an improved advertising effect [42, 43]. [12] proposed an interactive mobile ad framework that allows users to filter out ads they are not interested in. [39] developed an interactive coupon-giving robot system for a shopping mall and demonstrated the usefulness of social robots for advertising.

2.2 Context-aware advertising

Systems that utilize the user's context information to present an appropriate advertisement are known as context-aware advertising. The contextual information used by such systems includes the weather [7], location information [21, 35], and video viewing history [30]. However, to the best of our knowledge, a system that uses dialogue content as context and presents advertisements in real time has not been proposed. However, there is another type of system similar to context-aware advertising that "provides appropriate advertisements to users": the recommender system.

Recommender systems estimate user preferences and then use them as a basis for recommending appropriate content and products. They have been used for recommending video content such as in-vehicle systems [14], tourist attractions [53], and even YouTube [8] and Netflix [13] shows. A system that provides higher quality and more accurate recommendations through interaction with the user is called a conversational recommender system (CRS) [6, 22, 41].

[17] created the goal-driven recommendation dialogue dataset (GoRecDial), which can be utilized to estimate the user's preferences and recommend movies by repeatedly asking questions in the dialogue. [54] proposed a new task topic-guided conversational recommendation that focuses on active guidance from the system side, as opposed to the traditional user request-driven dialogue, and built a large high-quality dataset that incorporates topic threads for topic control. In the dialogue, BERT was used for topic estimation, and GPT-2 was used for sentence generation. [26] proposed a new task, dataset, and baseline that can actively and naturally lead from open-domain non-recommendation dialogs such as chitchat to recommendation dialogs in user-system interaction. These studies require interaction and questions between the system and the user to estimate preferences and make recommendations, which is different from our work, in which advertisements are continuously presented on the basis of user interactions without questions.

2.3 Query expansion

Dialogue is a unique event, so a pre-prepared advertisement may not always be a good fit. If the advertisement can be appropriately transformed to fit the context of the dialogue, the effectiveness of the advertisement can be enhanced. The idea of transforming sentences to fit the context (in a broad sense) has been attracting particular attention in the field of information retrieval. This has led to the concept of query expansion (QE), where search results more desired by the user are presented by transforming the search query he or she entered. Consider the case where the user's input query is "senior citizens". In this case, since the synonym "the elderly" is more frequently searched, the user's query is replaced with the latter. Thus, in query expansion, the user's initial query is reformulated as needed. Several methods of query expansion use named entity recognition [9, 49], which classifies named entities such as "Japan" or "Albert Einstein" into appropriately named entity tags such as "Location" or "PersonName", respectively. [9] used the features of named entities and links to Wikipedia and freebase knowledge bases containing structured named entity tags and text for query expansion. However, because the types of queries that can be processed depend on the number of named entity tags, named entity-based query expansion methods have a problem in that the domain is limited when the number of named entity tags is small. In [33], the use of BERT [10] was shown to improve the accuracy for longer queries. These methods utilizing query expansion are aimed at appropriately transforming queries to improve search performance, which differs from our goal of appropriately completing words for Core-Queries.

There are relatively few studies on query expansion methods referring to conversation history. Among them, [50] proposed a method that appropriately extends a query by referring to the conversation history between users when a user inputs a query, and evaluated it using the CAsT dataset. In the CAsT dataset, the queries become more ambiguous with each turn of the conversation. However, the dialogue history included in the dataset is short (one sentence) and does not cover long dialogues where the topic changes dynamically.

Query auto completion (QAC) is a method that takes a part of a query from a user and generates multiple completion candidates in addition to the query. A typical example of query auto completion is Google Suggest, which is used in Google Search: when a query is entered, automatically completed query candidates are displayed at the bottom of the search bar. [2] proposed most popular completion (MPC), which generates a complemented query based on the frequency of searches along with the words entered by the user as a query. [18] achieved faster query completion by referring to a database created in advance. [23, 40] introduced a personalization feature to achieve more accurate query auto completion that adapts to each individual by utilizing user information. Instead of using a database, [47] uses a deep network-based language model to find the most likely complementary words for the user's current input. However, the methods discussed here perform query auto completion using only the user's input query or only the query and user information. Our approach differs in that we perform query auto completion with reference to the

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

conversation history. Research on automatic query completion using dialogue history includes [45], where the authors collected dialogue data and search history data in a conversational task using Web search and then used it as a basis to create a topic model utilizing Dirichlet-Hawkes processes (DHP). When the user performs a Google search, the search history and dialogue history are used to properly re-rank the query candidates in Google Suggest, thereby reducing keystrokes and other search efforts.

[11] has the advertiser create the item keywords in advance, and then extends them. These item keywords are similar to the Core-Queries in our work, but since they are prepared as words, the range of expression is limited. In our system, we enable advertisers to create Core-Queries more intuitively by allowing them to be entered as sentences. Furthermore, [11] does not handle dialogue history, whereas we aim to provide real-time advertisements to users who are talking to each other.

2.4 Mask language models for information retrieval

We focus on BERT [10] as a critical mechanism for developing our system. The reasons for employing BERT are that it is a model built on a large corpus to handle open domains. Moreover, BERT has outperformed existing methods in tasks such as document retrieval [29, 52] and question answering [32, 51]. [1] developed a system that integrates the open source information retrieval toolkit Anserini with BERT to perform end-to-end document retrieval for large document collections. In [37], BERT was utilized to calculate the similarity of unlabeled query-question pairs to increase the QA data size and achieved high-performance retrieval. [19] improved the BERT-based architecture and created a model that allows for real-time document ranking with improved speed and accuracy, and implemented it in a search engine. However, few studies have focused on the function of predicting masked words in BERT to extend queries and perform information retrieval. Because predicting masked words can be utilized to extend queries in any domain, we focus on this function.

2.5 Chat dialogue system

The most basic chat dialogue system uses a sequence-to-sequence framework [44], which can generate simple conversations when given a large conversational training dataset. Although it is suitable for short textual conversational tasks, it is known to generate generic responses such as “I don’t know” in open-domain dialogues because it assumes a one-to-one context-response relationship. To avoid this problem, methods to apply objective functions for one-to-many problems [24] and to control the training of natural language generation from the training data by adding a bias to the corpus during training [27] have emerged. Models that achieve a specific task-oriented dialogue during chatting include those that acquire task-specific knowledge through pre-learning [3], those that extract knowledge from the outside and input it with queries during inference in the context of restaurant reservations [25], and those that learn language processing and recommended entity selection in an end-to-end manner [28]. Although these approaches work well for certain tasks and domains in chat dialogues, they require large datasets to fit the models to those tasks and also demand expertise on the part of the developer, making it difficult to design the tasks. In this paper, we use advertising as a task and aim to enable advertisers without computer science backgrounds to appropriately insert advertisements into open-domain dialogues with only the simple task of preparing phrases.

2.6 Problem setting

We assume that the generated search query is used to search the database prepared by the advertiser to present the information sought by the user. Under this assumption, it is necessary to generate the search query used in the search

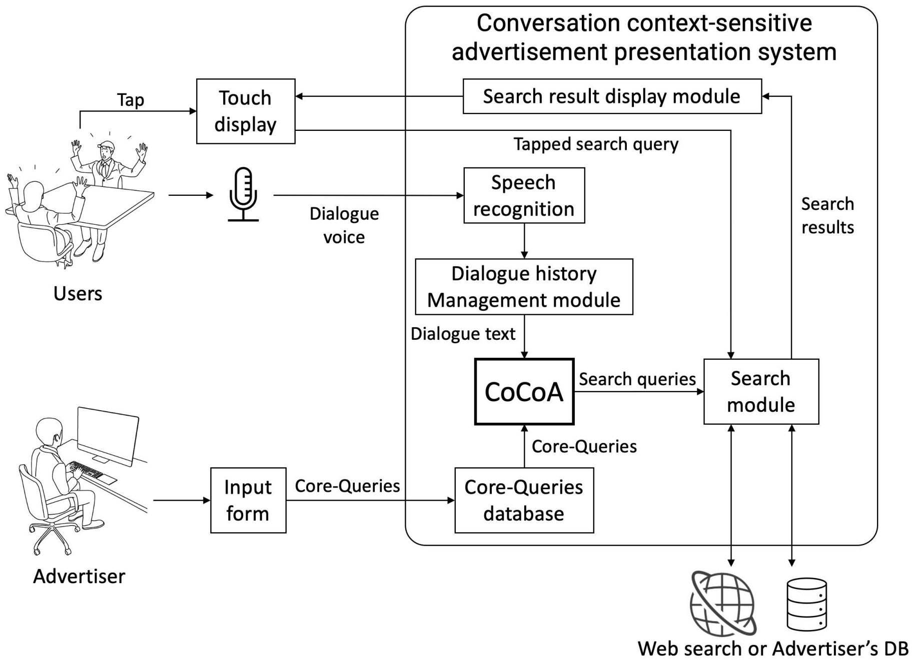

Fig. 2. System configuration diagram

according to the context of user dialogue. In this research, we evaluate whether the context of the user dialogue is properly taken into account when creating advertisements.

We focus on human-human conversations rather than human-computer conversations because we wanted to develop a method that would work well in a variety of contexts. A conversation between a human and a chat system restricts the context to just one specific topic that the system can handle. The specific topic is not appropriate for the target of our method. Furthermore, if the advertisement is presented in face-to-face human communication, it can be shared by both speakers and can facilitate the conversation. These are the reasons we focus on human-human interaction.

Toward the development of technology to generate advertising text in response to user interaction, we will build a system to present advertisements to users during interaction. Figure. 2 shows a diagram of the conversation context-sensitive advertisement presentation system. The system allows advertisers to create Core-Queries and store them in the Core-Query DB. The user's dialogue is recognized by the speech recognition module through the microphone, and the dialogue text is input to the CoCoA proposed in this research through the dialogue history management module. CoCoA selects a Core-Query appropriate for the dialogue text from the Core-Query database, converts the Core-Query into a search query appropriate for the topic of the dialogue, and outputs the search query. Based on the search query generated by CoCoA, the search module retrieves a digest of relevant advertising information from the advertising database or the Internet, creates an advertising list with touch-clickable buttons. The search-result-display module presents the advertising list on the display to the user. The user can view the presented list of ads during a conversation. If the user finds an advertisement in the list that interests him or her, he or she can tap on it. Then the search module retrieves the details of the advertisement information and displays them on the screen for the user. The advertisement

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

lists based on search queries and ads selected by the user with a tap may reversely have an effect on the user' s conversation.

In this research, we study the CoCoA mechanism based on the conversation context-sensitive advertisement presentation system, aiming at the development of a technology that generates advertisements in response to user interaction. In addition, by using this system with CoCoA incorporated, we will support users to actually engage in face-to-face conversation and, at the same time, view advertisements presented in accordance with the topic of conversation. Furthermore, the system will contextualize Core-Queries created by advertisers without effort to consider individual conversational situations according to the content of the user's interaction, and present the user with advertisements that are appropriate to the conversational situation.

2.7 Scope of this research

In this section, we clarify the technical issues that this research aims to address. This study focuses on the mask prediction functionality enabled by the large-scale language model, and challenges the question of how far advertising queries can be created by converting Core-Query text, which is advertising text, within the range of capabilities that can be achieved by contextualizing the text using a language model. The CoCoA module, the core of the system, receives texts of the dialogue and the Core-Query, and then performs a technique to contextualize the Core-Query text using a large-scale language model.

When aiming to present advertisements that appeal to users, it would be highly effective to serve advertisements according to the emotions and preferences of users during the conversation. It is also possible to estimate the mood and preferences of the conversation by conducting sentiment and emotional analysis of the user's conversational text. However, the effect of extending search sentences with a large-scale language model itself is still unclear, so this paper will focus on that aspect of the research.

3 COCOA

In this paper, we propose CoCoA, a method that dynamically selects the Core-Query from a list of candidates prepared in advance in accordance with the context of the dialogue and then appends the complementary word from the ongoing dialogue to the Core-Query. CoCoA combines the "Core-Query Selector", which selects the appropriate Core-Query from a list of candidates, and the "BERT-Query Completer (B-QC)", which determines complementary words, to enable advertisers to generate context-sensitive and effective advertisements in real time in any context with only a few Core-Queries. Since we assume that the generated search query is used to perform a database search, this sentence is considered a search query and does not need to be grammatically correct. The flow of the CoCoA process is shown in Fig. 3.

3.1 Create the Core-Query

The Core-Query is an abstract sentence that forms the core of a recommendation, and is prepared in advance by the advertiser. Examples are "Enjoy the spectacular view" or "Summer family travel" for travel agencies. We chose to make Core-Queries abstract sentences because they are more extensible as advertisements. In the above example, by making the sentence abstract without limiting the destination, it is possible to respond to any context by supplementing keywords from the dialogue in later steps.

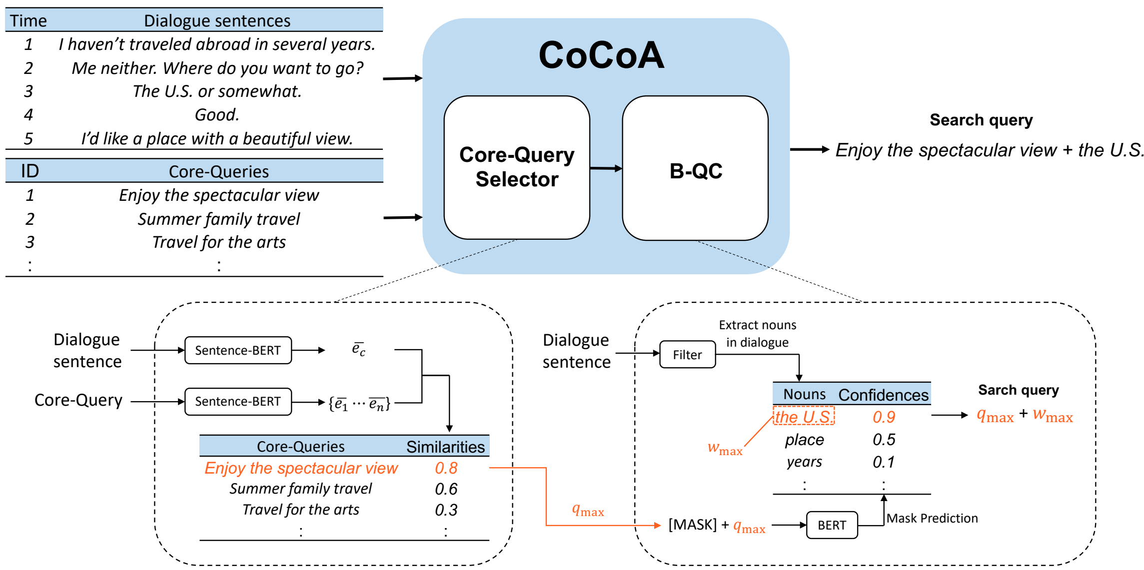

Fig. 3. Process flow of CoCoA.

3.2 Select the Core-Query

Let the set of Core-Queries q be the Core-Query list Q. The Core-Query Selector selects the Core-Query within the context of the current dialogue by calculating the sentence vector similarity between the current dialogue and the Core-Query. Specifically, the similarity γ_q is calculated from the sentence vectors, $\overline{e}_c$ and $\overline{e}_q$, of the last three dialogues $c_{t-2:t}$ and the Core-Query $q \in Q$ at time t,

$$\overline{e}_c = f_s(c_{t-2:t}), \tag{1}$$

$$\overline{e}_q = f_s(q), \tag{2}$$

$$\gamma_q = \mathrm{sim}(\overline{e}_c, \overline{e}_q), \tag{3}$$

where $f_s(\cdot)$ is a function to calculate the sentence vector and $\mathrm{sim}(\cdot)$ is a function to calculate the cosine similarity. We use Sentence-BERT [36] to calculate the sentence vector similarity. Then, for all similarity results $\Gamma_Q = \{\gamma_q | q \in Q\}$, the maximum $q_{\max}$ is output as the Core-Query suitable for the current context.

$$F_{\mathrm{QS}}(\Gamma_Q) = \begin{cases} \arg\max\limits_{q \in Q} \gamma_q, & \text{if } \exists_q \; 0.5 \leq \gamma_q \\ \varphi, & \text{otherwise} \end{cases} \tag{4}$$

Here, if all γ_q are below the threshold value of 0.5, it is determined that there is no Core-Query suitable for the current context, so no Core-Query is output and no final search query is output. This prevents the output of search query at every turn, which would probably irritate the user, and makes it possible to output search query only in contexts close to the Core-Query.

3.3 Select complementary word

B-QC takes as input the Core-Query $q_{\max}$ selected by the Core-Query Selector and all dialogue histories $c_{1:t}$ up to time t and then selects the word $w_{\max}$ from the dialogue history that matches the Core-Query. The final search query is output by adding the word to the Core-Query. Since the Core-Queries are abstract sentences and lack context-sensitive words, they need to be supplemented with a keyword contained in the dialogue history. We felt that the masked language model, which predicts the words in masked areas on the basis of the context, would be effective. Therefore, B-QC uses BERT's mask word prediction to predict context-dependent keywords. Since BERT includes masked language models in its pre-training, it can accurately complete words in sentences where some words are missing. For the Core-Queries, we assume that the context-dependent word is missing, and designed the B-QC to estimate the missing word using BERT.

In B-QC, the complementary word is determined by the following procedure, where the Core-Query selected by the Core-Query Selector is denoted as $q_{\max}$ and the part of BERT that predicts words is denoted as [MASK]. We assume that "+" denotes the conjunction of letters and sentences (e.g., this+is+an+apple $\longrightarrow$ this is an apple), and let the input sentence S to BERT be $S = [\text{MASK}] + q_{\max}$. Then, the words are predicted in [MASK] by

$$P = f_{\mathrm{B}}(S), \tag{5}$$

where $P = \{p_w | w \in \mathcal{W}\}$ denotes the predicted probability for any word w. From among the nouns $N_w(c_{1:t})$ in the dialogue history $c_{1:t}$, the word $w_{\max}$ with the maximum p_w is obtained as the complementary word.

$$w_{\max} = \underset{w \in \mathcal{N}_w(c_{1:t})}{\arg\max}\ p_w \tag{6}$$

Finally, the complementary word $w_{\max}$ is combined with the Core-Query $q_{\max}$ to output the final search query $\{w_{\max}, q_{\max}\}$.

4 EVALUATION EXPERIMENTS

4.1 Overview of experiments

We conducted two experiments: a "context sensitivity test" and a "robustness test".

In the context sensitivity test, the system generated search queries using core queries that we prepared in advance, and the participants answered the questionnaire by seeing them inserted in the dialogue.

In the robustness test, the system generated search queries using core queries created by each participants, and the participants answered a questionnaire by seeing them inserted in the dialogue.

The purpose of each is described below.

Context sensitivity test Conduct experiments using the common Core-Queries predetermined by the participants, and evaluate whether the Core-Query selected by the system and the complementary word are appropriate for the context.

Robustness test Evaluate the robustness of the system by having each participant create a query to see if the system can properly generate advertisements for all ad creators and all Core-Queries.

The quality of the advertising system by CoCoA depends on the performance of the part of the system that transforms Core-Query. In this experiment, we evaluated the performance of text processing related to core queries in a cloud experiment. In particular, since the purpose of CoCoA is to make it easier for advertisers to prepare the Core-Query itself, the evaluation of the contextualization of the Core-Query is important. In addition, the type of search results

Table 1. Details of the dialogue dataset.

Conversation	Trip	Food	Sports	Open-domain
No. of data	5	5	5	5
No. of turns in dialogue	32.4	43.2	33.2	107.6
Average no. of words	492.6	622	475.2	1789.8

that can be obtained from the Core-Query itself depends on the results of the Internet search, and since the uncertainty of the results is expected to be large, only the text was evaluated in the cloud experiment.

4.2 Datasets

We used the Osaka University Multimodal Dialogue Corpus (Hazumi) provided by Osaka University via the IDR Dataset Service of the National Institute of Informatics [20] as the dialogue data to present advertisements in both experiments. Hazumi is a multimodal corpus of dialogues between agents and people using Wizard-of-Oz. Each dialogue lasts about 15 minutes and contains data for 89 people. The dialogues of each experimental participant contain conversations on multiple themes. From these, we collected dialogues in three domains: "trip", "food", and "sports". In the dialogue data, the agent starts a dialogue by specifying a theme, such as "Let's talk about a trip!", so this is what we used to determine the domain. This resulted in 67, 72, and 29 sets of dialogs for the "trip", "food", and "sports" domains, respectively. The five dialogues with the longest number of turns were selected as "trip conversation", "food conversation", and "sports conversation", respectively. Next, from the dialogues of the experimental participants that were not selected for the above dialogues, we selected the five dialogues that contained multiple domains and had the longest number of dialogue turns, and designated these as "open-domain" dialogues. We obtained a total of 20 dialogue datasets containing five dialogs each and four types. Finally, the agent's utterance "Let's talk about XX!" indicating a change in the dialogue theme was removed to prevent the system from easily guessing the context of the dialogue and to make the dialogue more natural. The details of the obtained dialogue data are shown in Table 1.

4.3 Experimental setup

4.3.1 *Common setup for the two experiments.*

Design: In the two crowdsourced experiments, we presented information in a web form for crowd workers to evaluate online whether the search queries generated from Core-Queries were appropriate for the context of the conversation. Specifically, Core-Queries, conversational text, and system-generated search queries were presented in text form, and the crowd workers were asked to rate the system-generated search queries using a web form questionnaire.

Experimental environment: Since the crowdsourcing experiment employed a text-based evaluation, we could not display advertisements dynamically. Thus, the generated search query was evaluated by the participants as an "advertising sentence". This evaluation method enables evaluation of the text processing that CoCoA transforms Core-Queries into contextualized search queries. We assumed that web search would be used as a general-purpose database instead of the database prepared by advertisers.

Since we recruited Japanese participants, CoCoA had to deal with Japanese texts. CoCoA can be used with any language for which a masked language model is available (English, Spanish, French, German, Chinese, Japanese, etc.), but in this experiment, we used Japanese. In order to generate advertisements in Japanese, the following processes were performed in B-QC. We define the post-positional particle in Japanese as $j \in J$, where $J = \{$"ga","wo","ni","he","to","$yori$","$kara$","de","ya","no"$\}$. The Core-Query selected by the Core-Query Selector is q, the part of BERT that predicts the word is [MASK], and the input sentence S to BERT is S =[MASK]+$j + q$, after which the word is predicted in [MASK]. The confidence of the predicted word is p, and the word with the largest p among the nouns in the dialogue history is denoted as the candidate $w(j)$ for the complementary word predicted using the particle j. The above operations are performed for all j. $w(j)$ is added to the Core-Query as a complementary word in the order of frequency of occurrence among the set of $w(j)$, and the result is output as the final advertisement.

Experimental conditions: Three experimental conditions were set up for the two crowdsourcing experiments. The following three methods were used to generate search queries from Core-Queries, and we investigated the superiority of the CoCoA condition through evaluation of the search queries.

1. CoCoA condition: Create a search query using the CoCoA method. Select a Core-Query that is appropriate for the dialogue, and attach a complementary word that is appropriate for both the context of the conversation and the Core-Query.
2. Core condition: Use only the Core-Query Selector. Select a Core-Query appropriate for the dialogue and use it as the search query. No complementary word is attached.
3. Google condition: Use Google Suggest to select a complementary word. Select a Core-Query that is appropriate for the dialogue, and attach a complementary word that is the most frequently searched word among the words suggested by Google Suggest.

Since the Core condition uses a model that outputs only Core-Query as a search query without a complementary word, the output search query should be less context-dependent than the CoCoA condition. The Google condition uses Google Suggest for the complementary word, so the output search query is more specific than the Core condition. But the complementary word is not related to the context of the dialogue, so it is expected to be less context-dependent than the CoCoA condition.

Participants: All participants in both experiments were recruited through a crowdsourcing website.

4.3.2 Context sensitivity test.

Design: Using the web form prepared for the crowdsourcing experiment, we prepared to investigate which method is better by presenting search queries made under the CoCoA, Core, and Google conditions, along with dialog text. We conducted a context sensitivity test in which the Core-Query was the same for all participants. We extracted catchphrases consisting of two or more words related to "trip", "food", and "sports" from travel sites, restaurant search sites, and sports facility sites, respectively, and randomly selected ten catchphrases each for use as the Core-Queries in the "trip", "food", and "sports" domains. The Core-Queries used for each domain are listed in Table 2.

Experimental environment: To prevent bias in the evaluation of Core-Query by conversational domain, we prepared two different corpus: one is a conversational corpus with the same domain as the domain of Core-Query. The other is an open-domain conversation corpus, independent of the Core-Query domain. The first one was the advertisements generated in the "trip" Core-Query for "trip conversation", the "food" Core-Query for "food conversation", and the "sports" Core-Query for "sports conversation". The average number of advertising sentences generated by each dialogue was 24.8. The other was the advertising sentences generated in the "open-domain dialogue" in the Core-Query for each domain. Participants were able to see the dialogue data and the advertising sentences created by each system that were inserted in the dialogue data. The order in which the systems were displayed was randomly switched for each participant.

Participants: We used crowdsourcing to recruit 300 participants for the experiment: 128 men and 172 women aged between 18 and 73 years with a mean age of 40.7 years. We divided the participants into six groups of 50 people each. The first three groups were asked to respond to search queries where the topic of the Core-Query corresponded to the topic of the conversation. The other three groups were asked to respond to search queries for open-domain conversational text.

Procedure: In the evaluation, the following instructions were given. The variable {DOMAIN} contains the values "trip", "food", and "sports" according to the above domains.

> An advertiser decided to run an ad that would encourage {DOMAIN} after COVID-19 converged. So, she came up with catchphrases that would make users think "What kind of things are there?", search for it, and then want to go to {DOMAIN}. Evaluate the advertising sentences created by the robot based on the catchphrase.

Observation and data analysis: We adopted a within-subjects design between CoCoA, Core, and Google condition because dynamically providing advertisements according to a conversation is a novel experience for users, which leads to large individual differences in their impression and makes it difficult to draw a clear conclusion. The within-subjects design enables participants to evaluate the results in a unified measure. The questions used in the evaluation are shown in Table 3. The survey results from the experiments with conversational data from the same domain as Core-Query and the open-domain conversational text were mixed and the mixture of data eliminated the effects of differences in the conversational domain.

4.3.3 Robustness test.

Design: The second experiment was a robustness test in which participants created the Core-Queries. At the beginning of the web form, participants were asked to create ten Core-Queries. After input, the web form proceeded to the web form screen prepared for the crowdsourcing experiment, where search queries created based on the Core-Queries created by the participants were presented along with conversational text, using the CoCoA, Core-Query, and Google conditions. Participants were then surveyed to determine which method they preferred.

Experimental environment: As in the Context sensitivity test, we prepared conversational corpus of the same domain as the Core-Query for one group and the open-domain dialogue for the other.

Participants: We recruited 300 participants through crowdsourcing: 136 men and 164 women aged 15 to 73 years with a mean age of 39.1 years.

Procedure: The participants were divided into three groups of 100 people each, and each group was asked to create ten Core-Queries for each of three domains ("trip", "food", and "sports"). In the evaluation, the following instructions were given. The variable {DOMAIN} contains the values "trip", "food", and "sports" according to the above domains.

Table 2. Core-Queries used in context sensitivity test (originally in Japanese; translated into English for inclusion in this paper).

Trip	Food	Sports
A must-visit tourist destination	A place where we can drink craft beer	Health maintenance
A trip to experience something extraordinary	A place for entertaining	Increase basal metabolism and lose weight
Relaxing trip to an open-air bath	Cheap BBQ restaurant	Beginners are welcome
Enjoy the spectacular view	A place to eat sweets	Helps to combat lack of exercise
Experience nature	Famous Italian restaurants	Continue to have fun
Summer family travel	Bars with all-you-can-drink	Regain your strength
First cruise trip	Recommended for lunch	Fitness in your spare time
Tour to enjoy local food	Enjoy local cuisine	Become an athlete for the weekend
Women's travel by bullet train	Popular sports bar	Comfortable tiredness
A trip to enjoy the arts	All-you-can-drink course for more than 3 hours	Enjoy with friends

> You work for an advertising agency, and you've been assigned to run an ad that encourages people to {DOMAIN} after COVID-19 converged. So, you need to come up with catchphrases that would make users think "What kind of things are there?", search for it, and then want to go to {DOMAIN}.

In addition, we instructed them to pay attention to the following two points when creating Core-Queries, and allowed them to use a web search if necessary.

- The catchphrase should support various genres of {DOMAIN} in various regions.
- The ten catchphrases should not be too similar to each other.

Each group was divided into 50 participants, and each participant evaluated the advertising sentences generated by each system using the same procedure as the context sensitivity test, using the same domain as the Core-Query for three groups and the open-domain dialogue for the other three groups. The questions used in the evaluation are shown in Table 4.

Observation and data analysis: The questions used in the evaluation are shown in Table 4. Here, as in the Context sensitivity test, the questionnaire results from the experiment with conversational data from the same conversational domain as Core-Query and the experiment with open-domain conversational texts were mixed.

4.4 Experimental results

4.4.1 Context sensitivity test. The evaluation of each system by the experiment participants is shown in Fig. 4. We used ANOVA to compare each system and found that the main effect of the system was significant for all questions. Q1: $[F(2, 897) = 33.3, p < .001, \eta^2 = 0.069]$, Q2: $[F(2, 897) = 22.2, p < .001, \eta^2 = 0.047]$, Q3: $[F(2, 897) = 15.6, p < .001, \eta^2 = 0.034]$, Q4: $[F(2, 897) = 30.4, p < .001, \eta^2 = 0.063]$, Q5: $[F(1, 598) = 239, p < .001, \eta^2 = 0.285]$, Q6: $[F(2, 897) = 17.6, p < .001, \eta^2 = 0.038]$, Q7: $[F(2, 897) = 18.4, p < .001, \eta^2 = 0.039]$. Q8, which was an open-ended question, is described in the Discussion.

4.4.2 Robustness test. The evaluation of each system by the participants is shown in Fig. 5. We used ANOVA to compare each system and found that the main effect of the system was significant for all questions. Q1: $[F(2, 897) = 3.26, p = 0.039, \eta^2 = 0.007]$, Q2: $[F(2, 897) = 6.33, p = 0.002, \eta^2 = 0.014]$, Q3: $[F(2, 897) = 4.14, p = 0.016, \eta^2 = 0.009]$, Q4: $[F(2, 897) = 4.33, p = 0.013, \eta^2 = 0.010]$, Q5: $[F(2, 897) = 8.88, p < .001, \eta^2 = 0.019]$, Q6: $[F(2, 897) = 3.09, p = 0.046, \eta^2 = 0.007]$. The average creation time of the Core-Queries was 569 seconds, and we obtained 2955 different queries after duplicates were removed. The average number of advertisements output in each dialogue was 12.8.

Table 3. Questions for the context sensitivity test. A and B are the two characters in the dialogue.

Metric	Questionnaire items (Q1–Q7 were responded with seven-point scales.)
Understanding	Q1. Did the robot have a clear understanding of the conversation?
Effectiveness	Q2. Is it likely that the advertisement will have an effect on A and B?
Appropriate timing	Q3. Was the robot able to present the catch phrases created by the advertiser at the appropriate time?
Ad context awareness	Q4. Were the advertising sentences created by the robot based on the catch phrases created by the advertiser appropriate for the context?
Word context awareness	Q5. Are the words added by the robot to the catch phrases created by the advertiser appropriate for the context?
Expected timing	Q6. Did the robot display the advertising sentences at the expected time?
Unexpectedly good timing	Q7. Were there any advertising sentences that you did not anticipate when you saw the catchphrase, but which were displayed at the appropriate time?
Best ad	Q8. Which advertising sentence did you think was the best? (free description)

Table 4. Questions for the robustness test. A and B are the two characters in the dialogue.

Metric	Questionnaire items (each is rated on a 7-point Likert scale)
Understanding	Q1. Did the robot have a clear understanding of the conversation?
Effectiveness	Q2. Is it likely that the advertisement will have an effect on A and B?
Appropriate timing	Q3. Was the robot able to display the appropriate advertising sentence at the appropriate time?
Expected timing	Q4. Did the robot display the catchphrase at the timing you had in mind when you created it?
Unexpectedly ad	Q5. Did you receive an advertising sentence that you did not anticipate when creating your catchphrase?
Good timing	Q6. Did you see any advertising sentences that were displayed at the appropriate time?

4.5 Discussion

4.5.1 Context sensitivity test. First, we compare the proposed method with the other two methods. In all questions, the proposed method outperformed the others. This demonstrates that the proposed method generates appropriate advertising sentences in accordance with the context of the dialogues, which makes the advertisement more effective.

Next, we discuss the Google Suggest's overall lower ratings for Q1–Q7 compared to the Core-Query Selector-only model, even though Google Suggest was used to complete the words. A Core-Query is an abstract advertising sentence that forms the core of what the advertiser wants to recommend. Because of its abstraction, the minimum requirements for an advertisement can be satisfied as long as the Core-Query that matches the context is selected by the Core-Query Selector. Google Suggest, on the other hand, makes queries concrete by performing completion on the Core-Queries. In this case, Google Suggest is not able to refer to the dialogue history, so if the complementary word is inappropriate, it will result in a specific advertisement that does not match the context, and the evaluation will be low.

Finally, we discuss Q8, where the respondents were asked to select the best advertising sentence. Only those selected as the best created by the proposed method are discussed. First, for "trip", there are two Core-Queries included in the advertising sentences that are considered to be significantly better only in the proposed system: "A must-visit

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

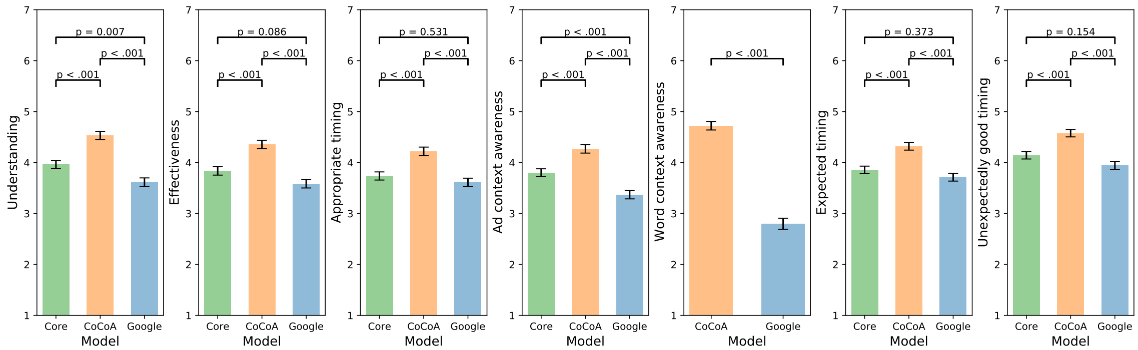

Fig. 4. Analysis results for context sensitivity test. The system with only Core-Query Selector is called "Core", the proposed system is called "CoCoA", and the system with Google Suggest is called "Google". As for Q5, the data are only for "CoCoA" and "Google", where the word is completed.

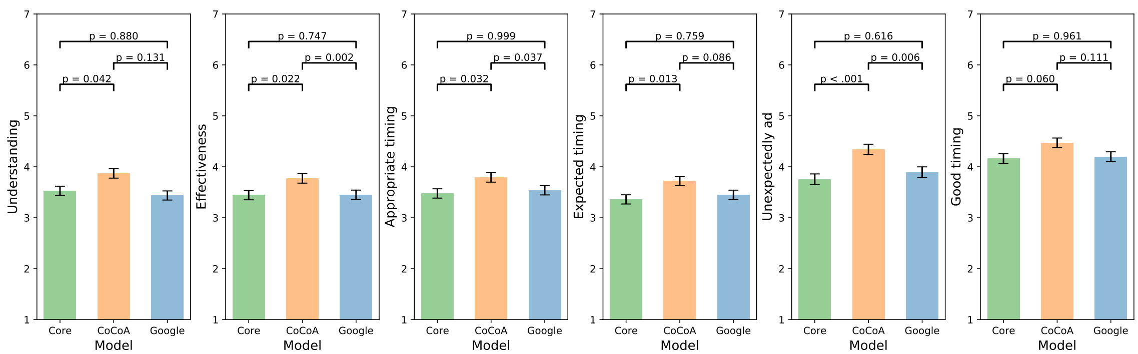

Fig. 5. Analysis results for robustness test.

tourist destination" and "Enjoy the spectacular view". For these, the advertising sentences with complementary words that specify the location received good evaluations. Specifically, participants mentioned the advertising sentences "A must-visit tourist destination Barcelona" in the context of a dialogue about Barcelona, "A must-visit tourist destination foreign country" in the context of a person who has not been abroad recently, and "Enjoy the spectacular view a hot spring" in the context of a dialogue about a person who wants to go to a hot spring. Next, for "food", there are two Core-Queries included in the advertising sentences that are considered to be significantly better only in the proposed system: "Enjoy local cuisine" and "Popular sports bar". In the former case, the advertising sentence "Enjoy local cuisine Hokuriku" (Hokuriku is a place name in Japan), which was complemented with place names as in the case of "trip", was highly evaluated. In the latter case, sports-related words were complemented instead of place names, and advertising sentences such as "Popular sports bar Olympics" when the topic was the Olympics or "Popular sports bar snowboarders" in the context of snowboarding as a hobby were highly evaluated. Finally, for "sports", the Core-Query included in the advertising sentences that was significantly better only in the proposed system was "Fitness in your spare time". This was highly appreciated when the word "housework" was complemented, which is interesting because, unlike with "trip" and "food", it is not a place name or a named entity. Specifically, in the context of the fact that housework is hard

Table 5. Core-Queries with the highest number of occurrences in each domain (originally in Japanese; translated into English for inclusion in this paper.)

Trip
Did you know that traveling is fun?
Popular hot spring resorts
Enjoy famous sightseeing spots
Hot spring lovers!
A tourist spot full of attractions
A chance to visit popular tourist spots
Rediscovering the charms of Japan
Travel to enjoy the hot springs
Treat yourself to a trip of luxury
This is the best time of year to visit

Food
Drink, eat, and be happy!
Daily lunch specials are recommended
Delicious snacks to go with drinks
A place with good drinks
A place to eat and relax in peace
Cafe with great sweets
A Japanese restaurant so delicious you'll want to keep coming back
Cafes with lots of delicious sweets
Women's favorite sweets shop
A cheap and tasty restaurant

Sports
Recommended sports for hobbies
If you want to eat what you love, play sports! Keep moving your body and eating what you like!
Continue to enjoy sports more easily and happily
If you watch the competition and want to do the same
It can be fun
Fun sports
Let's experience it! Olympic Games
Those who were inspired by the Olympics and want to play sports
If you want to lose weight, you need to do it in a fun and sweaty way.
It could be an Olympic sport

Table 6. Complementary words and number of times completed per domain for the Core-Queries (originally in Japanese; translated into English for inclusion in this paper.)

Trip		Food		Sports	
Word	Number of completions	Words	Number of completions	Words	Number of completions
Japan	173	Japan	126	sports	215
trip	95	cake	118	health	130
world	56	like	101	Japan	75
foreign countries	55	chocolate	100	Olympics	51
convenience store	28	confectionery	94	TV	47
Hokkaido	26	sake	91	tennis	24
Hokuriku	24	convenience store	73	friends	24
sea	23	department store	70	movie	21
sake	23	food	69	food	19
hot spring	22	coffee	67	baseball	19

because there are many things to do, the advertising sentence "Fitness in your spare time: housework" was generated to promote physical fitness by making efforts during housework.

4.5.2 Robustness test. First, we compare the proposed method with the other two methods. In all questions, the proposed method outperformed the others. Although the Core-Query in the robustness test was different for each participant, unlike in the context sensitivity test, we found that the system was effective regardless of the variation of the Core-Query. At the same time, we found that the evaluation from the Core-Query creator side, i.e., the advertiser side, was also high. In addition, although the Core-Queries alone can only present advertisements within the expected range, the responses to Q5 demonstrate that word completion using B-QC produced advertising sentences that were not expected by the advertisers.

Next, we discuss the obtained Core-Queries. Table 5 shows the Core-Queries with the highest number of occurrences in each domain. The total number of Core-Queries obtained was 2955, of which 1133 were displayed at least once as advertising sentences. The average number of characters in the Core-Queries that were displayed at least once was 12.1, and the average number of characters in the Core-Queries that were never displayed was 10.7. The results of ANOVA testing showed that there is a significant difference $[F(1, 2953) = 75.9, p < .001, \eta^2 = 0.025]$, which means that even in this limited dataset, the Core-Queries with a large number of words have a large number of occurrences. The reason may be that the sentence-vector could not be calculated accurately for the Core-Queries with a small number of characters, i.e., short sentences, because Sentence-Bert was used to select the Core-Queries.

Finally, we consider complementary words. Table 6 shows the complementary words with the highest number of occurrences in each domain in the advertising sentence created by the proposed system. In the trip domain, words describing destinations such as “Japan”, “world”, “overseas”, “Hokkaido”, “Hokuriku”, “sea”, “hot spring”, etc. were the most complemented. In the food domain, words indicating specific foods or drinks, such as “cake”, “chocolate”, “confectionery”, “sake”, and “coffee”, were frequently supplemented. In the sports domain, a wide range of complementary words related to sports were complemented, such as “tennis”, “baseball”, “Olympics”, “health”, “friends”, and so on. These findings, along with the evaluation results of the experiment, demonstrate that the system was able to complete the words appropriately.

5 CASE STUDIES

5.1 Purpose of case study

We conducted two case studies to examine how participants use CoCoA in an actual conversation by providing a preliminary CoCoA system to advertisers and users. These studies provide insights into the effects of CoCoA that cannot be obtained from the crowdsourced experiment alone. We have the following objectives in this case study.

1. How often do users refer to the advertisements CoCoA generates while conversing?
2. Do the advertisements from CoCoA have any effect on the topics of the users’ conversation?
3. How did advertisers evaluate the generated search queries?
4. Can the advertisers create the Core-Queries easily?

The first is a “user case study” and the second is an “advertiser case study”. The user case study investigates objectives 1 and 2 by having two participants acting as users conduct a conversation in front of the experimental system displaying generated advertisements during the conversation in real-time. The advertiser case study mainly investigates objectives 3 and 4. It consists of three participants acting as an advertiser and two users. The advertiser creates original Core-Queries and evaluates how the experimental system selects and modifies them to generate the search queries during conversations between the two users. As this case study includes participants playing user roles, we also investigate objectives 1 and 2 here.

5.2 Setup of case study

This subsection describes the policy, system and environment, experimental participants, procedures, observation methods, and data analysis utilized for the user case study (5.2.1) and advertiser case study (5.2.2).

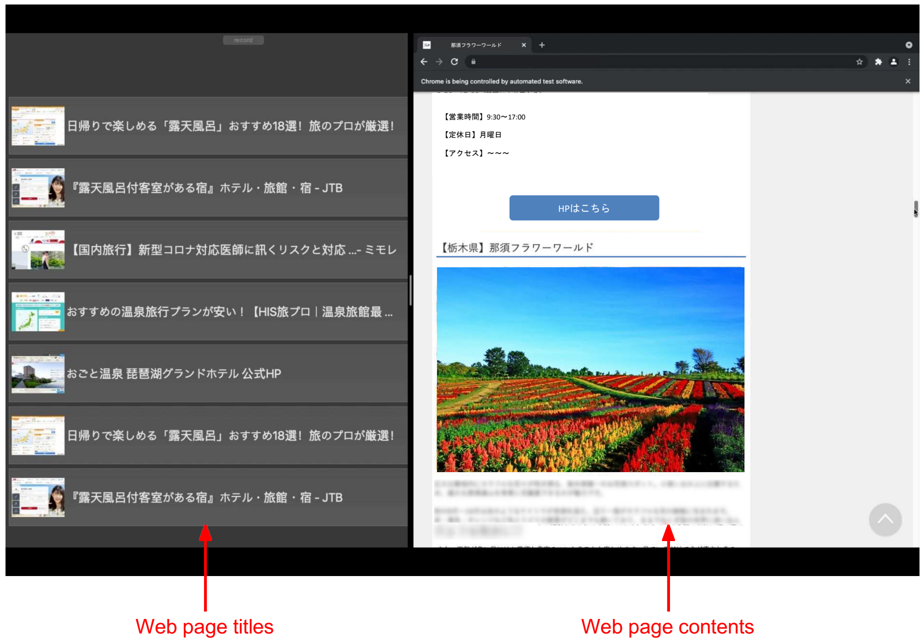

Fig. 6. System screen used in experiment.

5.2.1 User case study.

The policy of the user case study design: We prepared an interaction environment in which participants acting as users have a casual conversation in front of CoCoA, which allows us to observe how the system is utilized during conversation. The system, interaction environment, and experimental procedures were designed to satisfy objectives 1 and 2 discussed in the previous subsection.

System and environment for the user case study: We conducted the experiments for this case study in our laboratory. A screenshot of the system is shown in Fig. 6. The proposed system was implemented using the Core-Queries for "Trip" that had been derived from the context sensitivity test (Sec. 4). In these experiments, we used web search as a general-purpose database. Specifically, search queries were generated along with the dialogue input, and after performing a Google search using the search query, the titles and thumbnails of the top five pages of the search results were displayed as a set of buttons on the left side of the screen. Participants could touch one of the buttons to display the page contents on the right and then freely scroll and explore. A microphone placed on the desk was used to collect the dialogue audio. Figure 7 shows a photograph of the actual experimental environment, where participants were seated in pairs in front of a CoCoA-equipped display.

Participants for the user case study: A total of six participants (two men and four women, average age 23.0 years) took part in this experiment. They were divided into three pairs (Pairs A, B, and C), and each participant in the pair had an individual number (A-1 and A-2, B-1 and B-2, and C-1 and C-2) that we utilized when creating transcripts of their interaction.

Fig. 7. Experimental environment.

Procedure for the user case study: We asked participants to sit in pairs in front of the CoCoA-equipped display, gave instructions on how to use the system, and then encouraged them to chat for ten minutes. We did not provide any topics for the conversation. We told the participants they could use the system if they needed to, but we did not force them to do so. During the conversation, participants were free to view the advertisements on the display, which were updated in real time. After the experiment, we interviewed the participants to determine how they felt about using CoCoA.

Observation and data analysis for the user case study: We recorded the interaction between participants on a video to count how often they referred to advertisements appearing on the display and to observe any effect the advertisements from CoCoA had on the topics of conversation. We also created transcripts of the conversations from the video to grasp how participants behaved with reference to the advertisements and how they changed the topic of conversation in response to the generated advertisements.

5.2.2 *Advertiser case study.*

The policy of the advertiser case study design: We prepared the same system as the user case study, where two participants acting as users have a casual conversation. There was also an additional participant acting as an advertiser, who we asked to provide original Core-Queries and to evaluate the effectiveness of the search queries generated from them. This design allows us to observe how the advertiser creates Core-Queries as well as how the users utilize CoCoA during the conversation. We designed the system, interaction environment, and experimental procedures to mainly investigate objectives 3 and 4, as discussed in subsection 5.1. We also acquired some knowledge related to objectives 1 and 2.

System and environment for the advertiser case study: The system for the participants playing user roles was the same as that in the user case study. They had a conversation in the same environmental configuration shown in Fig. 7.

As for the participants playing advertisers, we prepared a web-based form for them to input the Core-Queries they created. This form was separated from CoCoA so that we could conduct the session of preparing the Core-Queries independently from the session of users' interaction. Before starting the users' interaction session, we input the created Core-Queries into CoCoA.

Experimental conditions for the advertiser case study: We set two experimental conditions: CoCoA and Core. The CoCoA condition generated search queries using the CoCoA model, which selected a Core-Query appropriate to the current users' conversation and found a complementary word to make the search query more suited to the context. The Core condition determined a Core-Query using the Core-Query Selector only and did not add any complementary word when outputting a search query. The displayed advertisements included the search queries created under both conditions at the same time. Since both search queries were based on the same Core-Query, the experimental system might show similar advertisements, which could confuse the users, so we attempted to mitigate this by having the system randomly shuffle the advertisement list. Ultimately, the users had conversations while watching advertisements generated by two conditions, and the advertisers could compare the effects of both.

Participants for the advertiser case study: A total of nine participants split into three groups took part in this experiment. In each group, one person played the role of advertiser and the other two played the role of users. The advertisers were participants from the user case study (one from each pair) and included one man and two women with the average age of 23.7 years. The users were six men with the average age of 22.5 years. The three groups were named D, E, and F, where D-1, E-1, and F-1 were the advertisers and D-2 and 3, E-2 and 3, and F-2 and 3 were the users.

Procedure for the advertiser case study: The advertisers chose their favorite domain from among "Trip", "Food", and "Sports" and then created ten Core-Queries using the same procedure as the Robustness test in Section 4. We sent the Core-Queries to CoCoA. Advertiser D-1 selected "Trip", E-1 selected "Food", and F-1 selected "Trip". We asked the users to converse in front of the CoCoA-equipped display in the same manner as the user case study. Also, the users in each group were not informed of the presence of the advertiser role. After the experiment, we interviewed the users to determine how they felt about using CoCoA. Advertisers were also interviewed after the end of each experiment. We asked the advertisers to evaluate the generated search queries after we had finished preparing the recorded video and transcripts of users' interactions (i.e., on another day).

Observation and data analysis for the advertiser case study: We recorded user interactions on a video that the advertisers could watch to assess their satisfaction with the search queries (objective 3). Advertisers were requested to respond to a questionnaire while referring to the user interaction video, dialogue transcript, generated search queries, and visited websites. The generated search queries and visited websites were aligned with the transcribed dialogue so that advertisers could follow the timing of each.

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

Table 7. Items evaluated by advertisers.

Metric	Questionnaire items (Q1–Q2 were answered using 7-point Likert scale).
Context awareness	Q1. Did the robot choose the appropriate catchphrase from among the prepared candidates, and/or transform the catchphrase into a more contextualized form?
Effectiveness	Q2. How effective was the advertisement?
Clicked website	Q3. Why do you think users clicked on the web site at this time? (free description)
Good catchphrase	Q4. Which of the catchphrases you created was the best, and why? (free description)
Bad catchphrase	Q5. Which of the catchphrases you created was the worst, and why? (free description)
Other Core-Query	Q6. Based on the results, what catchphrases other than the ones you prepared might would have generated more effective advertisements? (multiple answers allowed)
Best ad	Q7. Which was the best search query the system generated? (free description)
Impression	Q8. Please provide your overall impressions. (free description)

The items evaluated by advertisers are listed in Table 7. Q1 and Q2 were assessed in terms of each search query generated by the CoCoA condition and the Core condition, respectively, in the dialog. Q3 was answered only if the users visited the website. We utilized descriptive statistics to grasp the overall tendency of the responses from the relatively small number of participants: specifically, the median, maximum, and minimum values, the interquartile range, and the smoothed probability density of the data of the answers. The advertisers also responded to items Q4 to Q8 in Table 7. We interpreted the free descriptions and multiple answers to these questions to determine whether the advertisers were able to easily create the Core-Queries (objective 4). To cover objectives 1 and 2, we also counted how often users referred to advertisements in the advertisement case study and observed the effects of the search queries on the topics of the users' conversations.

5.3 Results

5.3.1 Overview of results. Subsection 5.3 reports the case study results to clarify the effectiveness of CoCoA in terms of the four objectives listed in subsection 5.1. We discuss two main results: the interaction that occurs between the pair of users and CoCoA, and the quality of the Core-Queries.

One notable result regarding the interaction between users is that CoCoA could make advertisements influence the user's conversation by presenting ads that match the context of the dialogue. In particular, the less familiar users were with each other, the more frequently they used CoCoA to search for information (see 5.3.2). The other results suggest that the advertisements CoCoA continued to display influenced the user's topic of conversation, even if the user did not tap on the ad (see 5.3.3).

The results regarding the quality of Core-Queries yields two findings. The first is that the search queries generated by CoCoA reflected the dialogue context to a remarkable extent. The other is that advertisers could create Core-Queries quite easily. The advertiser case study confirmed that the search queries generated by CoCoA were highly context-aware and accurately reflected the users' dialogues (see 5.3.4). Moreover, the proposed way of providing Core-Queries was easy for advertisers to learn and did not require them to have any specific knowledge about CoCoA (see 5.3.5). Advertisers could learn how to create more effective Core-Queries after just one trial of creating initial Core-Queries for CoCoA and observing how users behaved in response to the CoCoA-generated advertisements.

5.3.2 How often do users refer to advertisements generated by CoCoA while conversing? The number of times users tapped on the system screen during a ten-minute interaction was broken down as follows: case with the most taps (Pair A: nine times), cases with few taps (Pair C: one time, Group D: one time, and Group F: two times), and cases with no taps (Pair B: zero times (three times when the experiment was extended for another ten minutes) and Group E: zero times).

Case with the most taps: Pair A consisted of two casual acquaintances who had never chatted before. The fact that they tapped on the advertisements nine times in ten minutes indicates that they heavily relied on the system to keep the conversation going. We obtained essentially the same insight when we interviewed them: after the experiment, they stated that they had not been able to think of what to say, so they used the system to find topics for conversation.

Cases with few taps: Pair C consisted of two individuals who were meeting for the first time. They began the conversation with self-introductions, and after four minutes and fifty seconds, C-1 mentioned that she had studied in France. The system displayed a search result related to France and the remainder of the dialogue continued as they scrolled through the search results. Although the number of taps was just one, they used the system the entire time. D-2 and D-3 were friendly acquaintances who had conversations daily. Transcript 1 indicates that they tapped on the advertisement generated in response to a conversation about a TV travel program. The generated search query was "No Worries about Rain! Enjoy your holiday with indoor activities on a one-day tour" "definitely (a complementary word)".

Transcript 1 from Group D

D-2	The system is all about rainy day outings.		
D-3	Maybe it's because it's raining right now. You want to see something about a rainy day out?		
tapped query	"No Worries about Rain! Enjoy your holiday with indoor activities on a one-day tour" "definitely"	Q1: 3	Q2: 3
Web site	"https://www.jalan.net/news/article/231633/"		

"Q1: 3 and Q2: 3" on the right side of the tapped query in Transcript 1 indicate the advertiser's answers to questionnaire items Q1 and Q2 in Table 7. The value of 3 given to both indicates that advertiser D-1 felt the tapped search query was inappropriate to the conversation context and did not have enough of an effect for the advertisements. This opinion was confirmed in the interview when we asked D-1 to explain the reason for the tapping. D-1 answered that the users were surprised to see a website with content different from the context of their conversation at that time.

Users in Group F were friendly acquaintances who had conversations daily. In the case study, when they were talking about lack of exercise, they tried to tap on the search query consisting of "mountain climbing trip for exercise" "hiking". However, the system updated the query at the moment of the tap, and they accidentally tapped the query of "Round-the-Prefecture Bus Tour" "sport" by mistake.

Transcript 2 from Group F

F-2	I'm sure I'll get good at it if I do it a lot, but I don't have the opportunity to do it a lot. It's not something I go to every week.		
F-3	I know what you mean. People who score over 150 points are amazing.		
F-2	The system was showing something like "hike for exercise", but it's changed.		
tapped query	"Round-the-Prefecture Bus Tour" "sport"	Q1: 1	Q2: 1
website	"https://www.hatobus.co.jp/"		

After that, they talked about cooking and tapped the search query "Summer swimming tours to beat the heat" "processing".

Transcript 3 from Gropu F

F-3	Washing is the most annoying part.		
	Every time I worry about processing the grease that remains after cooking meat, I put it off.		
F-2	You have a lot of dishes to wash.		
F-3	By the way, I make curry a lot.		
tapped query	"Summer swimming tours to beat the heat" "processing"	Q1: 1	Q2: 1
web site	"https://www.city.fukui.lg.jp/kyoiku/kyusyoku/ncenter/natuyasaikare-.html"		

In this scene, the tapped query was not related to the conversation topic of cooking curry. However, the search resulted in the web page "Blow off the heat! Nutritious [Summer Vegetable Curry]", and since the webpage matched the curry conversation by chance, users F-2 and 3 tapped on it.

The low values of Q1 and Q2 indicate that advertiser F-1 did not recognize the appropriateness of the generated search query. We explain the reason in subsection 5.3.4.

Cases never used: The individuals in Pair B were acquainted with each other, had a chatting relationship, and did not use the advertising system at all during the first ten minutes. The reasons given were "I thought it would be rude to the other person because I had to move my eyes a lot to look at the screen while the dialogue was going on." There were also comments such as "It would be easier to use a screen the size of a computer on a desk" and "I was curious but didn't want to look away from my partner." After that, Pair B wanted to use the system a little more, so the experiment was extended another ten minutes and observation was continued. They checked the advertisements three times during the latter ten minutes of the interaction.

Users E-2 and E-3 were acquainted with each other and had a chatting relationship. The topic of conversation began with E-3's hobby of houseplants and ended with E-2's hobby of golf. They never tapped on the advertisement throughout their interaction. User E-2 mentioned in the interview that "I knew that when we were talking about health, there were web pages about health-conscious eating, but the rest of the time we weren't talking about restaurants, and I thought the topic of the conversation was different."

The system used for Groups D, E, and F generated two search queries based on the conditions of CoCoA (our method) and Core. Although the number of taps was small, all tapped queries came from the CoCoA condition.

5.3.3 Do the advertisements from CoCoA affect the topics of the users' conversation? The case studies featured some scenes where the topic of the users' conversation changed in accordance with the displayed advertisements. We discuss here all groups except Group E, as the users in this group did not look at the display during their conversation.

Pair A. The system displayed a search query about the hometown of one of the participants, and from there, the users started a dialogue about tourist attractions in that town. Since they did not have a topic for the conversation at first, this advertisement was able to initiate their interaction.

Pair B. Users did not use the advertising system at all during the first ten minutes. When the dialogue about going to Disneyland stopped, B-2 looked at the travel advertisement on the left half of the screen and then asked, "If you were going on a trip, where would you go?" In other words, they changed the conversation topic to a trip in response to the advertisement. In the latter ten minutes, the conversation proceeded in accordance with the contents of the displayed web page, and even when the conversation turned into a chat that had nothing to do with travel, an advertisement for a

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

cruise trip as the Core-Query was displayed. The context was then changed to a discussion about the price of going on a cruise.

Pair C. When one of the participants mentioned he had studied in France, the other asked about famous places there. This dialogue led to the generation of the search query “A must-visit tourist destination” “France” and a button about the beautiful scenery of France was displayed, which the user tapped. This would not have been possible if the system did not refer to the context of the dialogue. The button was only tapped once, but after that, the dialogue was inspired by the displayed web page until the end of the experiment. Specifically, the participant who had gone to France explained the details of the displayed tourist attractions, while the other participant asked several questions. In addition, there were times when tourist attractions that were not known to either of them were displayed, which eventually led to the utterance “I want to go to France.”

Group D. D-2 began by talking about his hobby of strength training, and from there, the topic shifted to exercise and then to travel with CoCoA as the trigger. D-2 commented on the topic shift that “I was talking about muscle training, but only ads about travel were shown. I found it hard to connect with the conversation.” D-3 said in the interview, “The system gave a good opportunity to change the topic and initiate a conversation.”

Group F. User F-2 had a conversation about not getting enough exercise. However, F-2 tried to tap on the search query “Round-the-Prefecture Bus Tour” “sport”, which was different from the topic of the ongoing conversation. Advertiser F-1 gave the following reasons for the tap: “The user had been feeling a lack of exercise recently, and wanted to see an advertisement for a trip that would help him be more active. As a result, the ad changed to a different ad, so it is not that they tapped on the ad they actually wanted to see, but I think their interest in the ad was strengthened.” In this situation, the user was interested in a search query about hiking and tried to tap on it, but the moment he did so, it changed to another search query, and as a result, he tapped on a search query that was not the one he wanted. However, at the time of the tap, the user’s interest in the ad was piqued and the conversation continued on the basis of the web site he opened. They began talking about tourist attractions in Tokyo that were listed on the web site.

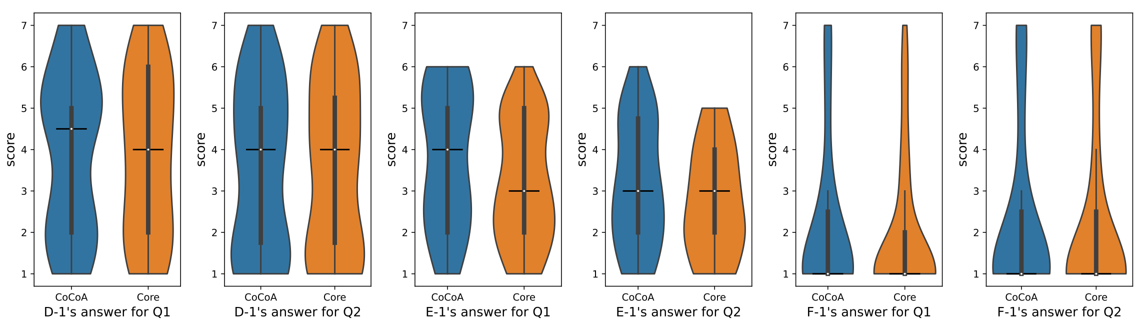

Fig. 8. The violin plots express the median, maximum and minimum values, interquartile range, and the smoothed probability density of the data of the answers to Q1 and Q2 obtained from Advertisers D-1, E-1, F-1.

5.3.4 *How did advertisers evaluate the generated search queries?*

The violin plots in Fig. 8 depict the descriptive statistics of each advertiser’s answers to Q1 and Q2 in Table 7. This formatting is intended to clarify the overall tendency of how the advertisers perceived the effect of the generated search queries.

We found from D-1's and E-1's answers to Q1 that the advertisers (except for F-1) felt the search queries generated by CoCoA were more suitable for the conversation context than the search queries without complementary words. These results are compatible with those of the crowdsourced evaluations.

In contrast, D-1's and E-1's answers to Q2 indicate a difference from the results of the crowdsourced evaluation (Q2: Effectiveness). Advertisers in the case study (except for F-1) did not identify a difference in the effect of two queries on the advertisement to the users, even if they recognized the impact of the complementary word to make the search query contextual. On the other hand, there were significant differences between the methods in the crowdsourced evaluations.

Two typical interactions we observed help us to infer why there was no tendency of difference. In the first, the query generated by CoCoA had the same weak effect on the advertisement as the one without a complementary word, even if CoCoA's query was appropriate to the conversation context. In the second, a selected core query was enough for giving context to the advertisement regardless of whether there was a complementary word.

The following transcript shows a case where the contextualized search query had only a weak advertisement effect (the same as the one without a complementary word).

Transcript 4 from Group E

E-2	I bought a flowerpot at IKEA and bought soil at Home Depot.		
E-3	It's authentic.		
E-2	How many do you have?		
E-3	About 10 now.		
E-2	What size?		
E-3	Small sizes.		
CoCoA	"Fully private rooms for peace of mind" "size"	Q1: 5	Q2: 1
Core	"Fully private rooms for peace of mind"	Q1: 2	Q2: 1

Users E-2 and E-3 talked about growing houseplants. Since CoCoA gave the Core-Query the complementary word "size", it acquired a higher Q1 score (5 points) than the Core-Query without it (2 points). However, the selected Core-Query "Fully private rooms for peace of mind" did not have enough attractiveness as an advertisement regardless of having/not having the complementary word. The Q2 scores of both generated search queries also received the same low point value because the selected Core-Query "Fully private rooms for peace of mind" did not have enough attractiveness as an advertisement for the current conversation, even though it contained an appropriate contextual word. The following transcript demonstrates that the Core-Query had a sufficient advertisement effect regardless of whether it contained a complementary word or not.

Transcript 5 from Group D

D-3	Yes, the overall concept was to create something like the bathhouse in "Spirited Away". So it's like a festival with lots of food stalls and stuff, so the kids are basically never bored.		
D-2	So it's family fun there then.		
D-3	That's the impression I got when I went there.		
CoCoA	"Three-day trip for you and your children" "family"	Q1: 7	Q2: 6
Core	"Three-day trip for you and your children"	Q1: 6	Q2: 6

CoCoA chose the Core-Query "Three-day trip for you and your children" in response to the conversation about a trip to a hot spring. Since the selected Core-Query fit the conversation topic, advertiser D-1 gave high scores for Q1 to both search queries: 7 for the query with a complementary word and 6 for the one without. In addition, the complementary word "family" improved the Q1 value of the search query.

Advertiser D-1 gave the same evaluation (6) for Q2 for both search queries. The complementary word "family" did not give new contextual information to the selected Core-Query, which included the phrase "your children", but the case had a sufficient advertisement effect due to the initially suitable design of the Core-Query.

Although the violin plots indicate that the advertisers seemed to consider the effect of complementary words on the advertisement to be insignificant (D-1's and E-1's answers for Q2 in Fig. 8), there was one search query among those generated by CoCoA that advertiser E-1 gave 6 points to in contrast to 5 points for the one without the complementary word. A partial transcript from this interaction is as follows.

Transcript 6 from Group E

E-2	That's good. I'd like to try it in the hot summer too.		
E-3	I want to go to the beach in the summer.		
CoCoA	"Bar with beautiful night view" "beach"	Q1: 6	Q2: 6
Core	"Bar with beautiful night view"	Q1: 5	Q2: 5

CoCoA added the complementary word "beach" to the Core-Query "Bar with beautiful night view", which was related to the conversation between the users. Advertiser E-1 mentioned the difficulty of giving advertisements to the users who talked about topics utterly different from the target domain that the Core-Query designated. Since the complementary word in the above example had the effect of contextualizing the Core-Query, the 6 points awarded to Q2 against the search query indicates that advertiser E-1 recognized the effect of the advertisement.

The violin plots for advertiser F-1's answers indicate that she evaluated the search queries severely. She stated during the interview that it was difficult to cover a lot of conversation topics by utilizing Core-Queries prepared for a single domain. Moreover, she was extremely strict when evaluating the search queries in terms of whether the users had been talking about a trip. This style of evaluation resulted in lower scores for Q1 and Q2.

5.3.5 Can the advertisers create the Core-Queries easily? All advertisers in these experiments were creating Core-Queries for the first time. The violin plots in Fig. 8 indicate that these Core-Queries included both "good" and "bad" ones. Although relatively few queries were created overall, advertisers D-1 and E-1 felt they could provide appropriate queries on average, and F-1 also created three good Core-Queries. We investigated whether the advertisers were able to make good Core-Queries by analyzing how well they grasped the created queries from questionnaire items Q4–6 and the interview responses.

The responses of the advertisers to Q4 (good catchphrase) and Q5 (bad catchphrase) after the experiment indicated they were able to judge the Core-Queries as good or bad simply by watching ten minutes of conversations conducted on the basis of the Core-Queries. The specific answers to Q4 and Q5 are as follows.

For Q4 (good catchphrase), advertiser D-1 selected "A little luxury! All-you-can-eat local gourmet tour." The reason for this answer was "because it can be used in a variety of situations and is very versatile." Advertiser E-1 chose "Bar with beautiful night view" and explained her reasoning by stating that "If the topic includes a place or scenery, users can relate to it even if they are talking about something other than food." Advertiser F-1 chose "Tour to enjoy seafood,"

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

commenting that "it would be easy to advertise because it can be displayed not only in trip conversations but also in meal conversations."

For Q5（bad catchphrase), advertiser D-1 selected "Enjoy a slightly different kind of holiday on a sleeper train." The reason given was "the limited number of places that can be reached by sleeper train and the lack of versatility." Advertiser E-1 chose "Fully private rooms for peace of mind" because "while it was generic, the fact that it did not refer directly to the restaurant made it a catchphrase with unintended connotations, depending on the words that followed." Advertiser F-1 selected "Summer swimming tours to beat the heat." The reason given was that "it would be summer-only or it would not appear unless it was a sea-related conversation."

Watching the users' 10-minute conversations helped the advertisers get the hang of creating Core-Queries. They evaluated Core-Queries as good if they did not restrict information and could be handled in conversations in other domains, while Core-Queries that could only be used in more limited situations were judged as less versatile and therefore as less desirable catchphrases.

In Q6, we asked the advertisers to create additional Core-Queries based on their experience in the advertiser case study. Advertiser D-1 came up with "For the active! Full-day activity tours", "No worries, even for an indoor person! The hottest indoor tourist spot", and "Student Discount: Cheap 2-day tours". E-1 came up with "Stop by and enjoy gourmet food", and F-1's answer was "Luxurious hot spring trip to enjoy seasonal ingredients". These responses indicate that the advertisers were able to learn how to create more sophisticated Core-Queries after just one experience of using the CoCoA system.

5.4 Discussion

5.4.1 Overview. Subsection 5.4 discusses the usage scenarios of CoCoA, suggestions for improvement, and limitations of the present setup based on the results of each case study.

We can clarify the actual usage scenarios of CoCoA by taking into account the relationship between CoCoA users, which also helps us to consider ways of improving CoCoA (see 5.4.2). From the user interviews, we found that there was a clear advantage in terms of dialogue support in that it was possible to provide conversation topics to users with low familiarity with each other. Furthermore, it is possible to improve the system to attract a broader range of users by adjusting the frequency of displaying advertisements based on the familiarity and relationships between users. In particular, we found that the system had an effect on user interaction even when the topic of the dialogue was beyond the scope of what the Core-Query and the complementary word could cover (see 5.4.3). This means that if the advertisements displayed by CoCoA are in domains unrelated to the conversation, the unexpectedness of the ads could prompt users to tap them, thus changing the dialogue to the domain of the advertisements and activating user conversation.

We also investigated appropriate complementary words, what a better Core-Query is, and how to create a Core-Query. We found that proper complementary words were related to the context of the user's dialogue, as is the goal of CoCoA. The discussion referring to the transcriptions of actual interactions reveals that advertisers gave high evaluations to search queries that CoCoA generated by contextualizing the Core-Queries with appropriate complementary words from the users' dialogues (see 5.4.4). In addition, the case study results indicate that better Core-Queries covered a wider range of topics (see 5.4.5). Moreover, based on the results of the questionnaires for advertisers, we identified three standard features that better Core-Queries should possess. The criteria provide a guideline for advertisers in creating Core-Queries in the future (see 5.4.6).

We emphasize the need to improve the user interface of CoCoA when implementing it in actual stores (see 5.4.7). The critical points of improvement are the frequency of advertisement presentation and the user's gaze behaviors. The reason for choosing these points is that the frequency of advertisements provided by the UI utilized in the case study was too high for users to read, and the users felt it was inconvenient to turn their gaze to check the display during the conversation.

In addition to the above points, we also discuss possible biases and limitations in the case studies (see 5.4.8 and 5.4.9).

5.4.2 Effect of the relationship between users on the usage of CoCoA. We found that the less familiar the relationship between two users, the more they used the system. Once they started using the system, they continued to interact by keeping the displayed domains open and did not leave. For example, the number of conversational turns, displayed queries, and taps was 42, 44, and 3, respectively, for Pair B, which indicates a high degree of familiarity between users. In contrast, these numbers were 32, 37, and 9, respectively, for Pair A, which indicates a low degree of familiarity. Thus, the lower the familiarity, the more difficult it is to find a topic for conversation. The system thus played a key role in providing a topic of conversation for people who were not particularly close to each other.

For example, A-1 commented after the experiment, "I couldn't think of a topic, so I used the system to find out", and E-3 stated, "it may be useful in assisting communication between people who meet for the first time". F-2 also commented that "I found this helpful when there was a pause in the conversation. It might be even more helpful if it was a first meeting." Advertiser F-1 commented on the usefulness of CoCoA, saying "users who didn't know each other well could use this system to expand the conversation."

Since the low-familiarity pairs had fewer dialogue turns, they had more time to use the system to find topics when both users were silent, which increased the number of taps. In contrast, since the close-familiarity pairs had more dialogue turns, they paid less attention to the system. This is the reason for the low attention to the system in the close-familiarity pairs.

The case study results indicate that CoCoA has a distinct advantage in terms of supporting dialogue by giving a topic of conversation to users with low familiarity. Furthermore, we believe the system can adjust the frequency of displaying advertisements depending on the user's relationship to include a broader range of users.

Finally, we should point out that even close-familiarity pairs found it difficult to find a conversation topic when talking for a long time. The results from the low-familiarity pairs indicate that CoCoA can also support the close-familiarity pair when they run out of topics.

5.4.3 Advertisements for domains unrelated to the conversation trigger new topics of conversation. The domain of the ongoing dialogue is not necessarily the same as the domain of the Core-Query that the advertiser wants to advertise. For example, users in Group E did not tap the advertisement for Food generated from the Core-Query related to the "Food" domain because the content of their dialogue was not in the Food domain. Furthermore, if the domain of the Core-Query differs from the domain of the dialogue, CoCoA may not select the appropriate complementary word. For example, if the user is talking about travel and CoCoA selects a Core-Query related to sports, CoCoA chooses a complementary word that is a place name, which is not necessarily appropriate for a sports Core-Query. As a result, the generated search query may not be able to draw users' attention.

On the other hand, these search queries may be eye-catching to the user. Users in Group D selected a search query unrelated to the dialogue because of the unexpectedness of the search results by tapping it. Also, users in pair B, Group D, and F tapped on the search query unrelated to their conversation. While they were looking at an advertisement shown by the search query, the users shifted the topic of their conversation to the domain that the advertisers had

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

originally wanted to advertise by the Core-Query. Advertiser E-1 also commented, “The catchphrases created by Robot 1 (CoCoA) were often strange word combinations, but I found them to be interesting for the accidental ways they went together.” The results of the case study suggest that CoCOA has the potential to attract the user’s interest by continuing to present information even in domains that are irrelevant to the context. Furthermore, once the topic of conversation shifts to the same domain as the Core-Query over the course of continuous presentation, CoCoA can more accurately select complementary words and present more effective advertisements.

An earlier study on creativity support utilized a method of presenting information unrelated to the topic to activate user interactions and stimulate new ideas [46]. The usage scenarios of CoCoA include similar interactions of giving users topics they may not expect and activating new conversations. For example, F-3 stated, “I enjoyed learning about new things. For example, I have never been to Tokyo Sky Tree, but it appeared on a web site I happened to be looking at. I thought that was good in that it was not 100% related to what we were talking about, but slightly off, leading to a new topic. It made me want to go to places I didn’t know existed.”

The main purpose of CoCoA is to advertise in a way that is relevant to the topic. However, the results of the case study suggest that CoCoA can also benefit users by activating their conversation even beyond the scope of topics that Core-Query and complementary words can cover. We did not envision this advantage when we originally designed CoCoA.

5.4.4 Advertiser’s evaluation of contextualizing the Core-Queries. This section discusses the context awareness and effectiveness of the advertisers’ responses to the search queries generated by CoCoA. A close examination of the search queries that the advertisers cited in Q7 (best ad) of the questionnaire demonstrates the effectiveness of CoCoA, since the search queries that the advertisers in each group cited here were all generated by CoCoA. For example, the best ad in Group D, “A little luxury! All-you-can-eat local gourmet tour” “Okinawa”, was generated by CoCoA. The adjacent conversation is as follows.

Transcript 7 from Group D

D-2	Yes, I went to Okinawa.		
D-3	What were you doing there? Did you just want to go to Okinawa?		
CoCoA	“A little luxury! All-you-can-eat local gourmet tour” “Okinawa”	Q1: 6	Q2: 4
Core	“A little luxury! All-you-can-eat local gourmet tour”	Q1: 5	Q2: 4

The reason this was considered the best ad is presumably that words representing place names have a high affinity with “Trip”, the domain of Group D’s advertisements, and the word “Okinawa”, which represents a place name that appeared in the conversation immediately before the word, was complemented, thus increasing its suitability as the search query. Therefore, in the trip domain, ratings were often higher when a word representing a place was chosen as a complementary word in the dialogue. For example, the conversation about sports continued and turned to E-2’s hobby of golf. In transcript 8, the place name “Saitama”, which appeared in the conversation as a place where a golf course is located, was chosen as a complementary word.

Transcript 8 from Group E

E-2	There are not many golf courses near here to begin with, but golf courses in Kawasaki and Tama are really expensive.		
E-3	I see.		
E-2	But if you go to the Chichibu area in Saitama Prefecture, or Yamanashi, Ibaraki, or Tochigi, it is a little cheaper.		
CoCoA	"Old-fashioned retro coffee shop" "Saitama"	Q1: 6	Q2: 5
Core	"Old-fashioned retro coffee shop"	Q1: 3	Q2: 3

This narrowed down the location of the Core-Query "Old-fashioned retro coffee shop", which is why the CoCoA search query was rated higher in terms of both Context awareness and Effectiveness than the one by Core.

As the above examples show, CoCoA is able to generate good search queries by contextualizing the Core-Queries with contextual complementary words.

While CoCoA can generate good search queries, the overall results of the advertiser case study indicate that advertisers often did not perceive a clear difference in the advertising effectiveness (Q2: Effectiveness) of CoCoA's contextualized search queries from the original Core-Queries. Unlike the crowdsourced experiment in Section 4, where the search queries were evaluated on the conversational text, in the advertiser case study, where the search queries were evaluated on the live interaction, there were multiple factors that may affect which advertisements appeal to users, and this was the main reason there was little difference in the evaluation in Q2. Specifically, in Section 4, we found that CoCoA was significantly superior to the other systems with respect to Q2 in both the Context sensitivity test and the Robustness test (Effectiveness in Figs. 3 and 4).

Although the advertiser case study cannot be directly compared with these results because they are descriptive statistics of the individual evaluations by each advertiser, there is no noticeable difference among all of the violin plots for Q2 in Fig. 8, and the results of the advertiser case study reveal a different trend from the Effectiveness results of the context sensitivity test and robustness test.

The most significant reason unique to the case studies is that advertisers may rate the system less highly if they use it less frequently. For example, advertiser E-1 responded in Q8 (Impression) that "I felt it was difficult to advertise when we were having a conversation with a completely different domain," and the evaluation was lower overall than that of Group D. In fact, while the domain of Group E's Core-Query was "Food", there was no dialogue in which "Food" was the domain, and no taps were made, so Q2 (Effectiveness) was also rated low overall. On the other hand, the maximum score of Q1 may have been higher for CoCoA than for Q2 in terms of the combination of contextualized complementary words. Thus, it seems likely that Q2 was lower overall because it was used less frequently owing to the dialogue not always transitioning to the dialogue in the domain of the Core-Query in a short 10-minute dialogue.

5.4.5 Appropriate breadth of topic domain for designing Core-Queries. The results in subsection 5.3.3 and corresponding discussion in 5.4.3 and 5.4.4 suggest that Core-Queries should have a wide domain that can cover a variety of topics. Since CoCoA selects Core-Queries that are appropriate for the user's conversation and at the same time selects a complementary word that is appropriate for applying the Core-Query to the context of the conversation, even if the domain of the Core-Query is broad, we can expect the search query to be contextualized and on-topic. To increase the chances of appropriate Core-Queries appearing and thereby increase the effectiveness of advertising, it is important to design Core-Queries with an awareness of the breadth of the domain.

This is a preprint of an article published in ACM Transactions on Interactive Intelligent Systems.
The final authenticated version is available online at: https://doi.org/10.1145/3588578

In this case study, each group created ten Core-Queries with one domain, so there were situations where the domain of the Core-Query did not match the domain of the dialogue, making it difficult to display the advertisement. In fact, as advertiser D-1 stated, “I felt there were situations that could not be handled with the ten catchphrases I had prepared.” Advertiser E-1 similarly stated that “I found it difficult to advertise when users were having a completely different domain conversation.” She then added, “I felt that I needed to either create catchphrases that connect the different domains or prepare ads for multiple domains.” This suggests that, if multiple advertisers were simultaneously creating Core-Queries for different domains, it would be possible to present appropriate advertisements based on the domain of the conversation and on the user interests.

As indicated by the above comments from the participants playing the role of advertisers, a method of preparing Core-Queries for each of multiple different domains should be considered. At the same time, considering CoCoA’s ability to contextualize Core-Queries, it would also be effective to design Core-Queries that can span multiple domains to some extent.

5.4.6 How to prepare Core-Queries. In the advertiser case study, participants in the advertiser role gained a better understanding of how to design Core-Queries after reflecting on the Core-Queries they created in a single experiment. This relatively easy learning curve suggests that only a few cycles of creating Core-Queries and using the CoCoA are required for helping advertisers figure out the appropriate way to create Core-Queries. Specifically, advertisers were able to understand that it is important to create Core-Queries that clearly represent the domain they want to advertise but that can also be used for a wider range of domains. The three advertiser participants in this case study created Core-Queries for the first time, and many of the search queries created on the basis of these prepared Core-Queries were found to be effective for users, as indicated by the violin plot in Fig. 8. We found that even advertisers with no experience in creating Core-Queries could present more effective advertisements to users by using CoCoA. In the following, we present some basic considerations for how to create effective Core-Queries.

The responses to Q4, Q5, and Q6 (subsection 5.3.5 Can the advertisers create the Core-Queries easily?) indicate that the following three points are common to effective Core-Queries.

1. Core-Queries that contain the elements of the domain you want to advertise.
2. Core-Queries that can be used for other domains.
3. Core-Queries that are not too specific.

Regarding 1, advertiser E-1 chose the domain of “Food”, created Core-Queries, and then indicated after the experiment that the Core-Query “Fully private rooms for peace of mind” was least effective (Q5: bad catchphrase). As a reason, E-1 stated that “this Core-Query did not explicitly include the domain of food that I wanted to advertise, and as a result, search queries for the domain of food were not created.” This comment indicates that if the Core-Query does not adequately represent the target advertising domain, it may generate search queries that are generic but do not accurately advertise.

Regarding 2, advertiser D-1 selected “A little luxury! All-you-can-eat local gourmet tour” as a good catchphrase (Q4), and advertiser F-1 selected “Tour to enjoy seafood”. In both cases, it was possible to take a place name as a complementary word in the trip domain, and since the Core Query also included an element of “Food”, advertisements in a wider range of domains could be presented. Another search query, “A morning menu to energize you for the day”, which was created by E-1 and displayed during a conversation about sports, was similarly an effective Core-Query for both “Food” and “Sports”. Thus, Core-Queries that can be associated with multiple domains have the potential to generate appropriate search queries in a broader domain dialogue.

Regarding 3, for Q5 (bad catchphrase), advertiser D-1 selected the Core-Query "Enjoy a slightly different kind of holiday on a sleeper train", which included a very specific mode of transportation ("sleeper train"). While this specific Core-Query represented the travel domain well, it was less versatile and actually used only once as a Core-Query.

We believe that by presenting the above three characteristics of effective Core-Queries to advertisers as design guidelines, they will get the hang of preparing Core-Queries more quickly. We should point out of course that characteristics 1 and 2 are contradictory and neither can be fulfilled perfectly. However, by designing Core-Queries while at least keeping 1 and 2 in mind, advertisers can be motivated to adjust the size of the domain to the extent possible.

5.4.7 How to display search results and where the monitor should be. The proper placement and UI of the display showing advertisements based on the search query created by CoCoA needs to be improved. Our objective in the case study was to automatically generate search queries tailored to the content of the user's dialogue, but if the system is to be used in a real environment, it needs to be provided in a form that is more user-friendly and that places less of a cognitive load on the user. We therefore asked participants for their opinions about the UI after each experiment. User A-2 commented that "there were so many buttons to be updated that they would disappear the moment we tried to tap them", so we changed the number of buttons to be updated from five to two in subsequent experiments.

Regarding Group D, when we asked if they had noticed the search query about "Okinawa", as they did not seem interested in it when they were talking about "Okinawa", D-2 said, "I hadn't noticed. The screen was right next to where we were talking, so I didn't have a chance to see it. We were both looking right at the other person, so I knew something was on the side screen, but it was hard to see." D-3 mentioned, "I think it would be better if the web site was displayed automatically, instead of just having a button waiting on the left screen to open the web site. Although the button displays the title of the web site, it is just a list of characters and has a lot of information. It would be better to have the page open suddenly." In addition, D-2's comment that "It is also difficult to see that the same things are being displayed" suggests that the same advertisement with similar queries increases the cognitive load on the users. In Group E, E-2 commented that "It would be easier to see the system if it were placed on the table between us, because the direction of eye movement would be up and down." E-3 said "I would like to see animations that move up and down when the list on the left side of the screen is updated. This would make it easier to see which button is the new one, and the movement would draw the eye."

Thus, our future work will focus on determining the appropriate display positions and how to display search queries in ways that facilitate eye guidance during dialogue.

5.4.8 Bias in the use of CoCoA. In any case study conducted in a laboratory setting, biases that encourage participants to use the target system naturally exist, so care must be taken in interpreting the results regarding whether participants used the system or not. In our case study, participants sat across from each other at a desk in a laboratory equipped with a microphone and chatted with each other while interacting (or not) with a display. This environment entails a possible bias in that it is designed for usage of the system.

We attempted to minimize this potential bias by telling participants that they could use the system if they wanted to but by no means were they required to. In fact, some of the participants never used the system, so we believe the bias was effectively minimized.

The instructions we provided to users brought them closer to the actual situation in which they use applications on websites and mobile phone, i.e., with the understanding that advertisements will be displayed on them. Specifically, when using such applications, it is up to the user's free will whether or not to click and view the presented advertisements. In this experiment, we tried to create a usage scenario that mimics this situation as closely as possible.

Overall, although there was some degree of bias regarding CoCoA use in this case study, we believe it was fairly close to the actual use scenario of the system and that the bias was minimized as much as it could be.

5.4.9 *Limitations of the case study.* This case study has the following limitations

Dependence on BERT model: Since the processing of dialogue text depends on the BERT model, CoCoA may not perform correctly for words that were not included when the BERT model was trained, for dialogue texts in completely new domains, or for grammatically incorrect sentences. CoCoA may not have performed as designed depending on the topic of the conversation, the user's utterances, or Core-Query.

Limitations of speech recognition: Since CoCoA utilizes speech recognition technology to convert user dialogue into dialogue text before processing it, this case study was conducted within the performance limits of speech recognition. In our case studies, some search queries were rated lower due to speech recognition errors in which the words uttered by the user were not recognized correctly, resulting in words that were not included in the dialogue becoming complementary words. In addition, in speech recognition, failure to detect speech segments and misrecognition can occur due to surrounding noise. Since this case study was conducted in a laboratory setting, the effects of environmental noise were eliminated.

Limitations of search databases: The database used in this experiment was not a dedicated database for advertisements, so there may be no search results at all, or it may happen to show something like it in an online search.

6 FUTURE WORK

Although the experimental results demonstrated that CoCoA can generate advertisements in accordance with the conversational context, challenges still remain in applying this system to the real world.

Resolving the privacy issue is critical if people are going to accept our approach. At the very least, we need to ask users to provide their consent prior to use. Then, after obtaining consent, there should be some sort of benefit involved, such as the user getting a coupon or using the facility at a lower price, as in the scenario described in Section 1. At present, our research is focused on demonstrating the application of masked word prediction and we have not addressed the privacy issue yet. In putting CoCoA to practical use, it is also important to resolve the privacy issue on a technical level. For example, we need to identify proper nouns extracted from dialogues that lead to personal information or domain words that users feel uncomfortable with, and to improve the system so that effective advertisements can be provided while protecting privacy at the same level as other personalized advertisements.

As pointed out by the experiment participants in the case study, the current method of presenting advertisements in CoCoA requires various minor improvements. On the other hand, the interaction scenes that can be supported by the CoCoA technology are not limited to human interaction. For example, it is possible to extend the technology to present advertisements linked to text chats such as Messenger, or to present information on special offers. When actually developing applications, support considerations and user interface designs that are different from those of CoCoA will be necessary. Further development of the method constructed in this paper will be possible through new research and development.

One significant development issue for the practical application of CoCoA in the future will be the resolution of problems related to speech recognition. Besides improving speech recognition technology, we believe that CoCoA's ability to present advertisements can be improved. To solve the problems caused by misrecognition of speech, consider first that the attributes of words that are compatible with a Core-Query will vary from domain to domain. For example, in the travel domain, complementary words are compatible with words that describe places. Thus, when the advertiser creates a Core-Query for "Trip", if the system can automatically select only words with the attribute of location as complementary words, it can create search queries for the domain the advertiser wants to advertise. At the same time, if the system can filter out complementary words that are irrelevant to that domain or are generic (e.g., "regular"), advertisers will be able to provide more effective advertisements, and users will see advertisements that are relevant to the context of the interaction.

Improving the language model used in CoCoA is also important for system improvement. The BERT used by CoCoA is based on an existing Japanese model trained using the Japanese Wikipedia corpus, but it is also possible to fine-tune the BERT model with dialogue data, to fine-tune it with the assumption that Core-Queries are used, and to use other natural language models to generate effective advertisements that are more appropriate to the dialogue context from a wider range of Core-Queries. Furthermore, CoCoA currently adds only one complementary word to a Core-Query, but by increasing the number of complementary words, more specific search queries can be generated. Furthermore, instead of combining Core-Queries and complementary words, Core-Queries can be converted into a single sentence that matches the context based on the Core-Query, and that sentence can be presented as an advertisement. This could then be presented to the user in a way that feels more intuitive for them.

7 CONCLUSION

In this paper, we proposed CoCoA, a method for dynamically generating search queries for generating context-sensitive advertisements by utilizing mask prediction. CoCoA consists of a Core-Query Selector and a BERT-Query Completer (B-QC). The Core-Query Selector calculates the similarity of sentence vectors between user utterances and Core-Queries, abstract advertising sentences prepared by the advertiser, and selects one that fits the context of the dialogue best. In the B-QC, we used masked word prediction to generate concrete search queries in real time by complementing Core-Queries selected by the Core-Query Selector with an appropriate word from the dialogue history. We can extract context-sensitive advertisements by utilizing word-complemented search queries. We conducted experiments in which users evaluated the advertisements presented by CoCoA against dialogues in several domains prepared in advance. The results showed that CoCoA was able to present more contextually relevant and effective advertisements than Google Suggest or a method without word completion. In addition, we found that CoCoA could generate high-quality advertisements that advertisers had not expected when they created the advertisements. In addition, we conducted case studies involving users and advertisers in using CoCoA. The results indicated that pairs unfamiliar with each other referred to the advertisement CoCoA showed on display more frequently, which increased the chance for the advertisement to affect their conversation topics. Moreover, the advertisers highly evaluated the quality of the search queries. They easily grasped the way of designing a good Core-Query by observing the response of the users to the advertisements retrieved with the generated search queries.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant number JP21J13789 and JST CREST Grant number JPMJCR19A1, Japan.

REFERENCES

[1] Zeynep Akkalyoncu Yilmaz, Shengjin Wang, Wei Yang, Haotian Zhang, and Jimmy Lin. 2019. Applying BERT to Document Retrieval with Birch. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference

on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations (Hong Kong, China). Association for Computational Linguistics, 19–24. https://doi.org/10.18653/v1/D19-3004

[2] Ziv Bar-Yossef and Naama Kraus. 2011. Context-Sensitive Query Auto-Completion. In Proceedings of the 20th International Conference on World Wide Web (Hyderabad, India) (WWW '11). Association for Computing Machinery, New York, NY, USA, 107–116. https://doi.org/10.1145/1963405.1963424

[3] Antoine Bordes, Y-Lan Boureau, and Jason Weston. 2017. Learning End-to-End Goal-Oriented Dialog. In International Conference on Learning Representations. https://openreview.net/forum?id=S1Bb3D5gg

[4] Salem Mohamed S Busen and Che Su Mustaffa. 2014. The role of interactive advertisements in developing consumer-based brand equity: A conceptual discourse. Procedia-Social and behavioral sciences 155 (2014), 98–103. https://doi.org/10.1016/j.sbspro.2014.10.263

[5] Verolien Cauberghe and Patrick De Pelsmacker. 2006. Opportunities and Thresholds for Advertising on Interactive Digital TV. Journal of Interactive Advertising 7, 1 (2006), 2–23. https://doi.org/10.1080/15252019.2006.10722122

[6] Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. 2019. Towards Knowledge-Based Recommender Dialog System. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (Hong Kong, China). Association for Computational Linguistics, 1803–1813. https://doi.org/10.18653/v1/D19-1189

[7] Keith Cheverst, Nigel Davies, Keith Mitchell, and Adrian Friday. 2000. Experiences of Developing and Deploying a Context-Aware Tourist Guide: The GUIDE Project. In Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (Boston, Massachusetts, USA) (MobiCom '00). Association for Computing Machinery, New York, NY, USA, 20–31. https://doi.org/10.1145/345910.345916

[8] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In Proceedings of the 10th ACM Conference on Recommender Systems (Boston, Massachusetts, USA) (RecSys '16). Association for Computing Machinery, New York, NY, USA, 191–198. https://doi.org/10.1145/2959100.2959190

[9] Jeffrey Dalton, Laura Dietz, and James Allan. 2014. Entity Query Feature Expansion Using Knowledge Base Links. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (Gold Coast, Queensland, Australia) (SIGIR '14). Association for Computing Machinery, New York, NY, USA, 365–374. https://doi.org/10.1145/2600428.2609628

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (Minneapolis, Minnesota). Association for Computational Linguistics, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[11] Siyu Duan, Wei Li, Jing Cai, Yancheng He, and Yunfang Wu. 2021. Query-Variant Advertisement Text Generation with Association Knowledge. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (Virtual Event, Queensland, Australia) (CIKM '21). Association for Computing Machinery, New York, NY, USA, 412–421. https://doi.org/10.1145/3459637.3482290

[12] Yong-Yi Fanjiang and Yu-Yu Wang. 2017. Context-aware services delivery framework for interactive mobile advertisement. Computer Communications 99 (2017), 63–76. https://doi.org/10.1016/j.comcom.2016.08.002

[13] Carlos A Gomez-Uribe and Neil Hunt. 2015. The netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems (TMIS) 6, 4 (2015), 1–19.

[14] Maria Guinea, Isabel Litton, Rigel Smiroldo, Irma Nitsche, and Eric Sax. 2020. A Proactive Context-Aware Recommender System for In-Vehicle Use. In Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing (Bangkok, Thailand) (ICVISP 2020). Association for Computing Machinery, New York, NY, USA, Article 32, 8 pages. https://doi.org/10.1145/3448823.3448852

[15] Nor Hazlina Hashim, Normalini, and Norhazlina Sajali. 2018. The Influence Factors Towards Mobile Advertising Message Content on Consumer Purchase Intention. Global Business Review 19, 5 (2018), 1187–1206. https://doi.org/10.1177/0972150918788746

[16] Elaine M. Huang, Anna Koster, and Jan Borchers. 2009. Overcoming Assumptions and Uncovering Practices: When Does the Public Really Look at Public Displays?. In Proceedings of the 6th International Conference on Pervasive Computing (Sydney, Australia) (Pervasive '08). Springer-Verlag, Berlin, Heidelberg, 228–243. https://doi.org/10.1007/978-3-540-79576-6_14

[17] Dongyeop Kang, Anusha Balakrishnan, Pararth Shah, Paul Crook, Y-Lan Boureau, and Jason Weston. 2019. Recommendation as a Communication Game: Self-Supervised Bot-Play for Goal-oriented Dialogue. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (Hong Kong, China). Association for Computational Linguistics, 1951–1961. https://doi.org/10.18653/v1/D19-1203

[18] Dimitrios Kastrinakis and Yannis Tzitzikas. 2010. Advancing Search Query Autocompletion Services with More and Better Suggestions. In Proceedings of the 10th International Conference on Web Engineering (Vienna, Austria) (ICWE'10). Springer-Verlag, Berlin, Heidelberg, 35–49.

[19] Matěj Kocián, Jakub Náplava, Daniel Štancl, and Vladimír Kadlec. 2022. Siamese BERT-Based Model for Web Search Relevance Ranking Evaluated on a New Czech Dataset. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36. 12369–12377.

[20] Kazunori Komatani and Shogo Okada. 2019. Osaka University Multimodal Dialogue Corpus (Hazumi). https://doi.org/10.32130/rdata.4.1

[21] Kai Li and Timon C Du. 2012. Building a targeted mobile advertising system for location-based services. Decision Support Systems 54, 1 (2012), 1–8.

[22] Raymond Li, Samira Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards Deep Conversational Recommendations. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (Montréal, Canada) (NIPS'18). Curran Associates

Inc., Red Hook, NY, USA, 9748–9758.

[23] Yanen Li, Anlei Dong, Hongning Wang, Hongbo Deng, Yi Chang, and ChengXiang Zhai. 2014. A Two-Dimensional Click Model for Query Auto-Completion. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (Gold Coast, Queensland, Australia) (SIGIR '14). Association for Computing Machinery, New York, NY, USA, 455–464. https://doi.org/10.1145/2600428.2609571

[24] Yahui Liu, Wei Bi, Jun Gao, Xiaojiang Liu, Jian Yao, and Shuming Shi. 2018. Towards Less Generic Responses in Neural Conversation Models: A Statistical Re-weighting Method. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (Brussels, Belgium). Association for Computational Linguistics, 2769–2774. https://doi.org/10.18653/v1/D18-1297

[25] Zhibin Liu, Zheng-Yu Niu, Hua Wu, and Haifeng Wang. 2019. Knowledge Aware Conversation Generation with Explainable Reasoning over Augmented Graphs. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (Hong Kong, China). Association for Computational Linguistics, 1782–1792. https://doi.org/10.18653/v1/D19-1187

[26] Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. Towards Conversational Recommendation over Multi-Type Dialogs. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (Online). Association for Computational Linguistics, 1036–1049. https://doi.org/10.18653/v1/2020.acl-main.98

[27] Nurul Lubis, Sakriani Sakti, Koichiro Yoshino, and Satoshi Nakamura. 2019. Positive Emotion Elicitation in Chat-Based Dialogue Systems. IEEE/ACM Transactions on Audio, Speech, and Language Processing 27, 4 (2019), 866–877. https://doi.org/10.1109/TASLP.2019.2900910

[28] Wenchang Ma, Ryuichi Takanobu, and Minlie Huang. 2021. CR-Walker: Tree-Structured Graph Reasoning and Dialog Acts for Conversational Recommendation. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (Online and Punta Cana, Dominican Republic). Association for Computational Linguistics, 1839–1851. https://doi.org/10.18653/v1/2021.emnlp-main.139

[29] Sean MacAvaney, Andrew Yates, Arman Cohan, and Nazli Goharian. 2019. CEDR: Contextualized Embeddings for Document Ranking. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (Paris, France) (SIGIR'19). Association for Computing Machinery, New York, NY, USA, 1101–1104. https://doi.org/10.1145/3331184.3331317

[30] Rishi Madhok, Shashank Mujumdar, Nitin Gupta, and Sameep Mehta. 2018. Semantic understanding for contextual in-video advertising. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32.

[31] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and Design Space for Interactive Public Displays. In Proceedings of the 18th ACM International Conference on Multimedia (Firenze, Italy) (MM '10). Association for Computing Machinery, New York, NY, USA, 1285–1294. https://doi.org/10.1145/1873951.1874203

[32] Rodrigo Nogueira and Kyunghyun Cho. 2019. Passage Re-ranking with BERT. arXiv preprint arXiv:1901.04085 (2019).

[33] Ramith Padaki, Zhuyun Dai, and Jamie Callan. 2020. Rethinking query expansion for BERT reranking. In Advances in Information Retrieval: 42nd European Conference on IR Research (Lisbon, Portugal). Springer, 297–304.

[34] Terry R. Payne, Esther David, Nicholas R. Jennings, and Matthew Sharifi. 2006. Auction Mechanisms for Efficient Advertisement Selection on Public Displays. In Proceedings of the 4th European Workshop on Multi-Agent Systems.

[35] Ajax Persaud and Irfan Azhar. 2012. Innovative mobile marketing via smartphones. Marketing Intelligence & Planning 30, 4 (2012), 418–443.

[36] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv:1908.10084 [cs.CL]

[37] Wataru Sakata, Tomohide Shibata, Ribeka Tanaka, and Sadao Kurohashi. 2019. FAQ Retrieval Using Query-Question Similarity and BERT-Based Query-Answer Relevance. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (Paris, France) (SIGIR'19). Association for Computing Machinery, New York, NY, USA, 1113–1116. https://doi.org/10.1145/3331184.3331326

[38] Ryoichi Shibata, Shoya Matsumori, Yosuke Fukuchi, Tomoyuki Maekawa, Mitsuhiko Kimoto, and Michita Imai. 2022. Utilizing Core-Query for Context-Sensitive Ad Generation Based on Dialogue. In Proceedings of the 27th International Conference on Intelligent User Interfaces (Helsinki, Finland) (IUI '22). Association for Computing Machinery, New York, NY, USA, 734–745. https://doi.org/10.1145/3490099.3511116

[39] Masahiro Shiomi, Kazuhiko Shinozawa, Yoshifumi Nakagawa, Takahiro Miyashita, Toshio Sakamoto, Toshimitsu Terakubo, Hiroshi Ishiguro, and Norihiro Hagita. 2013. Recommendation Effects of a Social Robot for Advertisement-Use Context in a Shopping Mall. International Journal of Social Robotics 5, 2 (2013), 251–262. https://doi.org/10.1007/s12369-013-0180-4

[40] Milad Shokouhi. 2013. Learning to Personalize Query Auto-Completion. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (Dublin, Ireland) (SIGIR '13). Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/2484028.2484076

[41] Yueming Sun and Yi Zhang. 2018. Conversational Recommender System. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (Ann Arbor, MI, USA) (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 235–244. https://doi.org/10.1145/3209978.3210002

[42] Shiree Treleaven-Hassard, Joshua Gold, Steven Bellman, Anika Schweda, Joseph Ciorciari, Christine Critchley, and Duane Varan. 2010. Using the P3a to gauge automatic attention to interactive television advertising. Journal of Economic Psychology 31, 5 (2010), 777–784. https://doi.org/10.1016/j.joep.2010.03.007

[43] Paraskevi Vennou, Evangelia Mantzari, and George Lekakos. 2011. Evaluating Program-Embedded Advertisement Format in Interactive Digital TV. In Proceedings of the 9th European Conference on Interactive TV and Video (Lisbon, Portugal) (EuroITV '11). Association for Computing Machinery, New York, NY, USA, 145–154. https://doi.org/10.1145/2000119.2000149

[44] Oriol Vinyals and Quoc Le. 2015. A neural conversational model. arXiv preprint arXiv:1506.05869 (2015).

[45] Tung Vuong, Salvatore Andolina, Giulio Jacucci, and Tuukka Ruotsalo. 2021. Spoken Conversational Context Improves Query Auto-Completion in Web Search. ACM Transactions on Information Systems (TOIS) 39, 3 (2021). https://doi.org/10.1145/3447875

[46] Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. 2010. Idea Expander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli. In Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (Savannah, Georgia, USA) (CSCW '10). Association for Computing Machinery, New York, NY, USA, 103–106. https://doi.org/10.1145/1718918.1718938

[47] Po-Wei Wang, J. Zico Kolter, Vijai Mohan, and Inderjit S. Dhillon. 2018. Realtime query completion via deep language models. In The 2018 SIGIR Workshop On eCommerce. https://openreview.net/forum?id=By3VrbbAb

[48] Kirsty Williamson. 2018. Research concepts (second edition ed.). Chandos Publishing. 3–25 pages. https://doi.org/10.1016/B978-0-08-102220-7.00001-7

[49] Chenyan Xiong and Jamie Callan. 2015. Query Expansion with Freebase. In Proceedings of the 2015 International Conference on The Theory of Information Retrieval (Northampton, Massachusetts, USA) (ICTIR '15). Association for Computing Machinery, New York, NY, USA, 111–120. https://doi.org/10.1145/2808194.2809446

[50] Jheng-Hong Yang, Sheng-Chieh Lin, Chuan-Ju Wang, Jimmy Lin, and Ming-Feng Tsai. 2019. Query and Answer Expansion from Conversation History. In Proceedings of the Twenty-Eighth Text REtrieval Conference (Gaithersburg, Maryland, USA) (NIST Special Publication, Vol. 1250). National Institute of Standards and Technology (NIST). https://trec.nist.gov/pubs/trec28/papers/CFDA_CLIP.C.pdf

[51] Wei Yang, Yuqing Xie, Aileen Lin, Xingyu Li, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin. 2019. End-to-End Open-Domain Question Answering with BERTserini. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations) (Minneapolis, Minnesota). Association for Computational Linguistics, 72–77. https://doi.org/10.18653/v1/N19-4013

[52] Wei Yang, Haotian Zhang, and Jimmy Lin. 2019. Simple applications of BERT for ad hoc document retrieval. arXiv preprint arXiv:1903.10972 (2019).

[53] Xiaoyao Zheng, Yonglong Luo, Liping Sun, Ji Zhang, and Fulong Chen. 2018. A tourism destination recommender system using users' sentiment and temporal dynamics. Journal of Intelligent Information Systems 51, 3 (2018), 557–578.

[54] Kun Zhou, Yuanhang Zhou, Wayne Xin Zhao, Xiaoke Wang, and Ji-Rong Wen. 2020. Towards Topic-Guided Conversational Recommender System. In Proceedings of the 28th International Conference on Computational Linguistics (Barcelona, Spain (Online)). International Committee on Computational Linguistics, 4128–4139. https://doi.org/10.18653/v1/2020.coling-main.365