DESCRIPTION

TITLE OF THE INVENTION: TRAINING METHOD FOR QUERY REFORMULATION MODEL, QUERY REFORMULATION METHOD, AND RELATED APPARATUSES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present disclosure claims the priority to the Chinese patent application with the filling No. CN2025102121935 filed with the Chinese Patent Office on February 25, 2025, and entitled "TRAINING METHOD FOR QUERY REFORMULATION MODEL, QUERY REFORMULATION METHOD, AND RELATED APPARATUSES", the contents of which are incorporated herein by reference in entirety.

TECHNICAL FIELD

[0002] The present disclosure relates to the field of machine learning, and in particular, to a training method for a query reformulation model, a query reformulation method, and related apparatuses.

BACKGROUND ART

[0003] Query reformulation is a technique widely used in the IT field, which refers to the process of modifying the way a query is phrased without changing semantic meaning of the query, so that the query is more specific and unambiguous. Query reformulation has wide applications in search scenarios and large-model conversation scenarios. For example, in the search domain, user queries may have grammatical issues, such as missing sentence components or typos, which may introduce uncertainty in query comprehension. Through query reformulation techniques, user queries may be corrected, thereby providing a text that is semantically clear and has a well-defined target for subsequent processing.

[0004] The reformulated queries may often better recall content related to the queries, which is crucial for enhancing user experience. However, it has been found in practice that the current reformulation method has poor performance for specific professional fields such as automobiles.

SUMMARY

[0005] To address at least one defect in the prior art, the present disclosure provides a training method for a query reformulation model, a query reformulation method, and related apparatuses, which enable the trained query reformulation model to comprehend queries in the automotive field and improve the accuracy of reformulated queries.

[0006] Embodiments of the present disclosure are implemented as follows.

[0007] According to a first aspect, the present disclosure provides a training method for a query reformulation model, where the method includes:

[0008] obtaining a fine-tuning sample of a pre-trained model, wherein the fine-tuning sample includes a query reformulation sample and a reformulation auxiliary sample in the automotive field, and the reformulation auxiliary sample is configured to enhance a comprehension ability and a reformulation ability of the pre-trained model to a query; and

[0009] training the pre-trained model with the fine-tuning sample to obtain a query reformulation model.

[0010] According to a second aspect, the present disclosure provides a query reformulation method, where the method includes:

[0011] receiving an original query input by a user; and

[0012] processing the original query and a preset automotive brand tree by the query reformulation model obtained by the method to obtain a reformulated query of the original query, wherein the automotive brand tree records a plurality of automobile brands and vehicle information of vehicles released by each automobile brand.

[0013] According to a third aspect, the present disclosure provides a training apparatus for a query reformulation model, wherein the apparatus includes:

[0014] a sample obtaining module, configured to obtain a fine-tuning sample of a pre-trained model, wherein the fine-tuning sample includes a query reformulation sample and a reformulation auxiliary sample in the automotive field, and the reformulation auxiliary sample is configured to enhance a comprehension ability and a reformulation ability of the pre-trained model to a query; and

[0015] a model training model, configured to train the pre-trained model with the fine-tuning sample to obtain a query reformulation model.

[0016] According to a fourth aspect, the present disclosure provides a query reformulation apparatus, wherein the apparatus includes:

[0017] a query receiving module, configured to receive an original query input by a user; and

[0018] a query reformulation module, configured to process the original query and a preset automotive brand tree by the query reformulation model obtained by the method to obtain a reformulated query of the original query, wherein the automotive brand tree records a plurality of automobile brands and vehicle information of vehicles released by each automobile brand.

[0019] According to a fifth aspect, the present disclosure provides a storage medium, wherein the storage medium stores a computer program, and the computer program, when executed by a processor, implements the training method for the query reformulation model and the query reformulation method.

[0020] According to a sixth aspect, the present disclosure provides an electronic device, wherein the electronic device includes a processor and a memory, the memory stores a computer program, and the computer program, when executed by the processor, implements the training method for the query reformulation model and the query reformulation method.

[0021] Compared with the prior art, the present disclosure has the following beneficial effects.

[0022] According to the training method for the query reformulation model, the query reformulation method, and the related apparatuses provided in the present disclosure, an electronic device obtains a fine-tuning sample of a pre-trained model, wherein the fine-tuning sample includes a query reformulation sample and a reformulation auxiliary sample in the automotive field, and the reformulation auxiliary sample is configured to enhance a comprehension ability and a reformulation ability of the pre-trained model to a query; and the electronic device trains the pre-trained model with the fine-tuning sample to obtain a query reformulation model. Therefore, when the model is trained, the query reformulation sample in the automotive field is used, and the auxiliary sample configured to enhance the comprehension ability and the reformulation ability of the pre-trained model to a query is introduced, so that the trained query reformulation model can comprehend queries in the automotive field and improve the accuracy of reformulated queries.

BRIEF DESCRIPTION OF DRAWINGS

[0023] To more clearly illustrate the technical solutions of the embodiments of the present disclosure, the drawings required in the embodiments will be briefly described below. It should be understood that the following drawings only illustrate some embodiments of the present disclosure and therefore should not be considered as limitations of the scope, and for those of ordinary skill in the art, other related drawings can be obtained according to these drawings without inventive efforts.

[0024] FIG. 1 is a schematic flowchart of a training method for a query reformulation model according to an embodiment of the present disclosure;

[0025] FIG. 2 is a schematic flowchart I of a query reformulation method according to an embodiment of the present disclosure;

[0026] FIG. 3 is a schematic flowchart II of a query reformulation method according to an embodiment of the present disclosure;

[0027] FIG. 4 is a schematic diagram of a structure of a training apparatus for a query reformulation model according to an embodiment of the present disclosure;

[0028] FIG. 5 is a schematic diagram of a structure of a query reformulation apparatus according to an embodiment of the present disclosure; and

[0029] FIG. 6 is a schematic diagram of a structure of an electronic device according to an embodiment of the present disclosure.

DETAILED DESCRIPTION OF EMBODIMENTS

[0030] To make the objectives, technical solutions, and advantages of embodiments of the present disclosure clearer, the following clearly and completely describes the technical solutions in embodiments of the present disclosure with reference to the drawings in embodiments of the present disclosure. It is clear that the described embodiments are some rather than all of embodiments of the present disclosure. The components of the embodiments of the present disclosure, as generally described and illustrated in the figures herein, can be arranged and designed in a wide variety of different configurations.

[0031] Therefore, the following detailed description of the embodiments of the present disclosure provided in the drawings is not intended to limit the scope of the present disclosure as claimed, but is merely representative of selected embodiments of the present disclosure. All other embodiments obtained by those of ordinary skill in the art based on embodiments of the present disclosure without inventive efforts shall fall within the protection scope of the present disclosure.

[0032] It should be noted that similar reference numerals and letters indicate similar items in the following drawings, and therefore, once an item is defined in one of the drawings, no further definition or explanation is required in the following drawings.

[0033] In the description of the present disclosure, it should be noted that terms "first", "second", "third", and the like are merely used to distinguish between descriptions, but cannot be understood as an indication or an implication of relative importance. In addition, terms "include", "comprise", or any other variants thereof are intended to cover a non-exclusive

inclusion, so that a process, a method, an article, or a device that includes a list of elements includes those elements, and also includes other elements which are not expressly listed, or further includes elements inherent to this process, method, article, or device. An element preceded by "includes a ..." does not, without more constraints, preclude the existence of additional identical elements in the process, method, article, or device that includes the element.

[0034] As described in the background, the current reformulation method has poor performance for specific professional fields such as automobiles.

[0035] In the research process, the implementation of query reformulation using both a rule-based method and a model-based method is provided. The following provides a detailed explanation of the principles and drawbacks of these two methods.

[0036] The rule-based method is a query reformulation method based on predefined rules. This method is used to supplement missing sentence elements or pronouns in a query by setting up a series of fixed rules. For example, if a key noun is missing from a query, the rule-based method may supplement this missing noun by a preset rule. The core of this rule-based method is to enable machines to automatically correct queries based on these rules by utilizing a fixed rule base. The rule-based method is simple to implement, and easy to program and maintain. Therefore, missing parts in the query can be quickly supplemented, and the clarity of the query is improved. However, due to the reliance on predefined rules, the rule-based method is difficult to fully cover complex or variable issues. The supplemented queries may not conform to natural language habits and may not be read smoothly enough. In addition, for queries in professional fields, the rule-based method may lack sufficient professional knowledge support, resulting in inaccurate reformulation results.

[0037] The model-based method implements query reformulation using a language model. This method trains a large-scale language model to enable this model to comprehend and rephrase queries. The model-based method can learn a large amount of language data during the training process, thereby improving the accuracy and naturalness of query reformulation. The language model can generate natural and smooth query reformulation in most cases after extensive training, and the performance of the language model on general queries is generally superior to that of the rule-based method. However, the model-based method performs poorly in specialized field, especially when there is a lack of sufficient expertise. There may be flaws in the model training process, such as ambiguity in the handling of professional terminology. In addition, in some specific tasks, the reformulation result generated by the model method may not be accurate enough, leading to problems occurring in subsequent processing links.

[0038] For example, in the automotive field, "Seal" actually refers to the name of a vehicle; however, when a language model is used for query reformulation, it is possible to identify "Seal" as an animal, thereby introducing ambiguity in the reformulated query.

[0039] Based on the discovery of the above technical problems, the inventors, through inventive work, provide the following technical solutions to solve or improve the above problems. It should be noted that the above defects in the solutions in the prior art are the results obtained by the inventors after practice and careful research. Therefore, the discovery process of the above problems and the solutions provided by the embodiments of the present disclosure for the above problems below should be the contributions made by the inventors to the present disclosure in the process of invention and creation, and should not be understood as the technical content known to those skilled in the art.

[0040] In view of the above problems, an embodiment of the present disclosure (hereinafter referred to as this embodiment) provides a training method for a query reformulation model. As shown in FIG. 1, the method includes:

[0041] S1A: Obtaining a fine-tuning sample of a pre-trained model.

[0042] The fine-tuning sample includes a query reformulation sample and a reformulation auxiliary sample in the automotive field, and the reformulation auxiliary sample is configured to enhance a comprehension ability and a reformulation ability of the pre-trained model to a query.

[0043] S2A: Training the pre-trained model with the fine-tuning sample to obtain a query reformulation model.

[0044] In this way, when the model is trained, the query reformulation sample in the automotive field is used, and the auxiliary sample configured to enhance the comprehension ability and the reformulation ability of the pre-trained model to a query is introduced, so that the trained query reformulation model can comprehend the query in the automotive field and improve the accuracy of reformulated queries.

[0045] An electronic device implementing the training method for the query reformulation model provided in this embodiment can be, but is not limited to, a mobile terminal, a tablet computer, a laptop computer, a desktop computer, a server, and the like, as long as it can provide sufficient computing power for model training. The server can be a single server or a server group. The server group can be centralized or distributed (for example, the servers can be a distributed system). In some embodiments, the server can be local or remote relative to a user terminal. In some embodiments, the server can be implemented on a cloud platform; and as an example only, the cloud platform can include a private cloud, a public cloud, a hybrid

cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, and the like, or any combination thereof. In some embodiments, the server can be implemented on an electronic device having one or more components.

[0046] To make the training method for the query reformulation model provided in this embodiment clearer, the steps shown in FIG. 1 are described in detail below using a server as an electronic device for implementing the method. It should be understood that the operations in the flowcharts can be performed out of order, and that steps without logical context can be performed in reverse order or concurrently. Furthermore, those skilled in the art, with the guidance of the present disclosure, may add one or more other operations to, or remove one or more operations from, the flowcharts. Still referring to FIG. 1, the method includes:

[0047] S1A: Obtaining a fine-tuning sample of a pre-trained model.

[0048] The fine-tuning sample includes a query reformulation sample and a reformulation auxiliary sample in the automotive field, and the reformulation auxiliary sample is configured to enhance a comprehension ability and a reformulation ability of the pre-trained model to a query. In an optional embodiment, the reformulation auxiliary sample includes a reading comprehension sample and a pronoun identification sample. The reading comprehension sample includes long text content, and the label is the corresponding summary content. The pronoun identification sample includes text content with pronouns, and the label is an entity object corresponding to the pronoun.

[0049] In this regard, it should be understood that the reformulation ability of the model is found to be closely related to the comprehension ability of the model to the query in the research process. To improve the reformulation ability of the model in the automotive field, it is necessary to ensure that the model can deeply comprehend the key information and context in the query. Therefore, a plurality of automotive field data similar to reading comprehension are collated. The main objective of this data is to help the model better comprehend complex text in the automotive field. Specifically, the reading comprehension sample includes some articles in the automotive field, which can come from automotive magazines or automotive forums, and gives summaries of these articles as labels for model training. This enables the model to better comprehend professional terms and background knowledge in the automotive field, so that this model more accurately grasps the key information of the query and reduces ambiguity in the query reformulation.

[0050] In addition, it is also found from research that users often use pronouns to refer to an entity mentioned in the context when make queries. For example, many queries contain pronouns (such as "he", "she", and "it") that may cause confusion. Therefore, this embodiment

specially sorts a pronoun identification sample, which is specifically used to training the model to identify and process pronouns. The labels of these data mainly focus on identifying pronouns in the query and finding specific entities these pronouns refer to. For example, in the query "What is the price of the Magotan, and what is its wheelbase?", the pronoun "its" refers to the entity "Magotan". The model needs to be able to identify this pronoun and comprehend that the object this pronoun refers to is the "Magotan" in the previous text. Such training enables the model to clearly indicate the object referred to by the pronoun during query reformulation process, thereby avoiding ambiguity and making the query expression clearer and more accurate. Therefore, disambiguation is a very critical part in the query reformulation.

[0051] The pronoun identification sample can come from the automotive field or the non-automotive field. It can be understood that, in the training the query reformulation model, not only should the query in the automotive field be addressed, but other types of queries in non-automotive fields that the user may pose are also considered. To make the model more generalizable when processing various queries, pronoun samples are introduced from non-automotive fields.

[0052] The server can obtain the query reformulation sample by the following embodiment:

[0053] S1A-1: Obtaining a user query in the automotive field and a context of the user query.

[0054] In this embodiment, "user query" refers to a specific query made by a user in the automotive field to a system, such as "What is the fuel consumption of the Magotan?". Such queries typically appear in automotive forums or query-answer platforms. "Context" refers to the context information or other queries in a series of queries that are related to the user query. For example, the user may have asked "What is the engine type of the Magotan?" before making the above query. Such queries and the answers to the queries, then the series of queries constitute context information of the user query.

[0055] In this embodiment, queries actually made by the user can be captured from a query-answer platform or a website in the automotive field. These platforms can be specialized automotive websites or automotive modules in large comprehensive websites, as these places bring together a large number of user queries and related discussions. Next, the context information of the user query is collected, which includes the context in which the user made the query and a possible set of related queries. For example, a series of queries that a user makes before and after making a query may be considered part of the context. The collection of context information ensures that the model can better comprehend the context and relevance of the queries.

[0056] S1A-2: Processing the user query and a context of the user query using a third-party large model to generate a reformulation content of the user query.

[0057] The reformulation content includes an inference part and a query reformulation part. In this regard, in this embodiment, the user query and the context thereof can be processed using a third-party large model (such as Deepseek and ChatGPT). The third-party large model analyzes the user query and the context thereof to generate the reformulation content. The generated reformulation content contains not only the reformulation version of the user query, but also inference parts that include the related context and intent of the query.

[0058] The so-called inference part refers to the process where the third-party large model not only generates the query reformulation part but also provides the inference process for generation of the query reformulation part. Specifically, the inference part mainly includes the following aspects:

[0059] First, the third-party large model needs to determine whether the current query is related to the previous user query. For example, if a user asks "What is the fuel consumption of the Magotan?" and the user previously asked "What is the engine type of the Magotan?", third-party large model needs to comprehend a correlation between these two queries. This correlation can help the third-party large model better comprehend the context and purpose of user queries.

[0060] Secondly, the third-party large model also needs to identify the automobile brands, vehicle series, or vehicle models involved in the current query. For example, when processing the query "What is the fuel consumption of the Magotan?", the third-party large model needs to identify that "Magotan" is a specific automobile brand and model. Identifying these entities helps the third-party large model comprehend the specific objects in the query more accurately.

[0061] Finally, the third-party large model needs to decide whether the query needs to be reformulated. Some queries are very clear and definite and do not need to be reformulated. For example, the query "What is the fuel consumption of the 2020 Toyota Corolla 1.8L (Automatic Transmission model)?" already includes elements such as brand, year, and model, making it sufficiently clear. However, for those ambiguous queries, the third-party large model needs to generate a clearer reformulation. In this way, the third-party large model generates complete information including not only query reformulation but also background inference.

[0062] For example, assume that the user asks "What is the fuel consumption of the Magotan?", the third-party large model may make the following inference for this query:

[0063] "The user asks 'What is the fuel consumption of the Magotan'. This query is related to the previous historical query 'What is the engine type of the Magotan', because the engine type

may affect fuel consumption. The user may want to know the fuel consumption performance of the Magotan (specific model) under different driving conditions, especially considering the engine type (such as turbocharged or naturally aspirated) of the Magotan. Therefore, the query needs to be reformulated to include more detail and context, such as requesting detailed fuel consumption data and explaining how different engine types affect fuel consumption."

[0064] Finally, the third-party large model reformulates the query as follows based on the inference content:

[0065] "The user wants to know the fuel consumption performance of the Magotan (specific model) under different driving conditions, especially considering the engine type (such as turbocharged or naturally aspirated) of the Magotan. Please provide detailed fuel consumption data and explain how different engine types affect fuel consumption."

[0066] Based on the description of the user query and content reformulation, the step S1 also includes:

[0067] S1A-3: Hiding part of the reformulation content, and taking the hidden content and the user query as a query reformulation sample.

[0068] In this embodiment, the server can hide the query reformulation part and the inference part, so that the model predicts the hidden part. It can be understood that a part of the reformulation content or some details in the inference process can be selectively hidden, so that the hidden part is predicted according to the remaining unhidden part during model training. In this way, a plurality of such query reformulation samples may not only train the query reformulation ability of the model, but also train the inference ability of the model. For example, for "user query", the third-party large model may obtain an "inference part" and a "query reformulation part". Assume that the "inference part" is hidden, and the remaining "user query" and "query reformulation part" are input into the model as query reformulation samples, and the model is trained to generate the "inference part", so that the model has a certain inference ability.

[0069] S2A: Training the pre-trained model with the fine-tuning sample to obtain a query reformulation model.

[0070] The pre-trained model refers to a large language model (LLM) that has been pre-trained on a large-scale dataset. This embodiment does not specifically limit which large language model to use, for example, the base models of the Qwen series, DeepSeek-V3, and Baichuan-Base. This model learns rich feature representation and generalization capabilities by training on a large amount of data. For example, the pre-trained model can be a pre-trained language model. Since the pre-trained model is a model pre-trained on large-scale unlabeled data and

has a certain basic language understanding ability, this pre-trained model can be fine-tuned in a supervised manner through specific task samples to better solve specific tasks. In this process, this embodiment uses three types of samples: a query reformulation sample, a reading comprehension sample, and a pronoun identification sample. In an optional embodiment, the step S2A may include:

[0071] S2A-1: Training the pre-trained model with the query reformulation sample, the reading comprehension sample and the pronoun identification sample to obtain an initial model.

[0072] In this regard, the server may fine-tune the pre-trained model using these three types of samples, where each type of sample type corresponds to a specific task. The query reformulation task requires the model to learn to reformulate the query, the reading comprehension task requires the model to comprehend and answer the query, and the pronoun identification task requires the model to accurately identify the pronouns and the referents of the pronouns in the text. Through this step-by-step training, the model gradually masters the basic skills on these tasks and forms an initial model.

[0073] S2A-2: Performing hybrid training on the initial model with the query reformulation sample, the reading comprehension sample and the pronoun identification sample to obtain the query reformulation model.

[0074] Next, to further improve the performance of the model on specific tasks, the initial model is continuously subjected to hybrid training. The hybrid training refers to that three different types of task samples, namely a query reformulation sample, a reading comprehension sample and a pronoun identification sample, are used simultaneously in the same training process, and the training losses of the three different types of task samples are weighted and summed to obtain the loss of a comprehensive model, so that the parameter of the initial model is adjusted. In this way, the model not only continues to consolidate previously learned skills, but can also establish connections between these different tasks, thus better comprehending the complexity of the language. Finally, the model after hybrid training performs better when processing query repetitive tasks.

[0075] Based on the query reformulation model obtained in the above embodiment, this embodiment further provides a query reformulation method. As shown in FIG. 2, the method includes:

[0076] S1B: Receiving an original query input by a user.

[0077] S2B: Processing the original query and a preset automotive brand tree by the query reformulation model to obtain a reformulated query of the original query.

[0078] The automotive brand tree records a plurality of automobile brands and vehicle information of vehicles released by each automobile brand.

[0079] In this way, when the query reformulation model optimizes the original query, taking the automotive brand tree as embedded prior knowledge can help the model better understand the correlation and classification structure between automobile brands, which provides important support for the query reformulation task, and enables the model to generate more targeted and diverse reformulated queries.

[0080] An electronic device implementing the query reformulation method provided in this embodiment can be, but is not limited to, a mobile terminal, a tablet computer, a laptop computer, a desktop computer, a server, and the like, as long as it can provide sufficient computing power for model training. The server can be a single server or a server group. The server group can be centralized or distributed (for example, the servers may be a distributed system). In some embodiments, the server can be local or remote relative to a user terminal. In some embodiments, the server can be implemented on a cloud platform; and as an example only, the cloud platform can include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, and the like, or any combination thereof. In some embodiments, the server can be implemented on an electronic device having one or more components.

[0081] To make the query reformulation method provided in this embodiment clearer, the steps shown in FIG. 2 are described in detail below using a server as an electronic device for implementing the method. It should be understood that the operations in the flowcharts can be performed out of order, and that steps without logical context can be performed in reverse order or concurrently. Furthermore, those skilled in the art, with the guidance of the present disclosure, can add one or more other operations to, or remove one or more operations from, the flowcharts. Still referring to FIG. 2, the method includes:

[0082] S1B: Receiving an original query input by a user.

[0083] For example, assume that a user is using an online query-answer platform specific to the automotive field. A user asked: "I want to buy a Corolla as a commuter car, and what is the approximate fuel consumption of the Corolla?". In this case, the user entered this sentence into the search box of the platform, which constitutes an original query. After receiving the input original query, the server optimizes the query with the trained query reformulation model so as to ensure that the query is more clearly and accurately expressed.

[0084] S2B: Processing the original query and a preset automotive brand tree by the query reformulation model to obtain a reformulated query of the original query.

[0085] The automotive brand tree records a plurality of automobile brands and vehicle information of vehicles released by each automobile brand. In other words, an "automotive brand tree" is a data structure used to record and organize detailed information of various automobile brands and vehicles released by each automobile brand. Similar to a tree diagram, the root node is the automobile brand, and the branches are various vehicle series and models under this brand, thus providing prior knowledge in the automotive field for the query reformulation model.

[0086] In the research process, it is also found that the current query reformulation method generally submits the reformulated query after the original problem is optimized to other large model models for processing, or feeds the reformulated query back to the user for confirmation. However, these two approaches have distinct limitations. The former cannot effectively guarantee the accuracy and reliability of reformulated queries due to the lack of a strict verification mechanism for the reformulated queries. This may lead to deviations or misunderstandings in subsequent processing steps, affecting the quality of the final results. The latter relies too much on the user's ability to make queries and the comprehension depth of the queries. If the user does not have enough professional knowledge or expression ability, the query may not be reformulated accurately enough, and the efficiency and the effect of the entire query solving process are affected.

[0087] In the research process, it is also found that when a vehicle is inquired, if accurate information is desired, it is necessary to include the complete elements required for the vehicle query, including "vehicle brand, vehicle model, production year, engine type, specific needs" and the like. Therefore, this embodiment can verify the reformulated query by using the complete elements provided in advance. Based on the inventive concept, as shown in FIG. 3, based on FIG. 2, the query reformulation method provided in this embodiment further includes:

[0088] S3B: Obtaining existing elements in the reformulated query.

[0089] For example, the server can continue to call the query reformulation model to process the reformulated query and extract existing elements in the reformulated query.

[0090] S4B: If the existing elements are compared with the preset complete elements and there are missing elements in the reformulated query, taking information corresponding to the missing elements in the automotive brand tree as missing information of the reformulated query.

[0091] In this regard, to reduce the quantity of the missing information, the server can obtain the existing information corresponding to the existing elements from the reformulated query according to the existing elements, and screen the automotive brand tree by using the existing information to match a target vehicle with the existing information; and obtain missing

information corresponding to the missing elements from the vehicle information of the target vehicle as missing information of the reformulated query.

[0092] S5B: Supplementing the missing information into the reformulated query to obtain a secondary reformulated query of the original query.

[0093] For example, assume that the reformulated query is "What is the fuel consumption of the Toyota Corolla 1.8L Hybrid?". After analysis, it is found by the server that the existing elements include "vehicle brand, vehicle model, engine type, and specific requirements". Compared with the above complete elements, the missing element is "production year". In this case, the server screens the automotive brand tree by using "Corolla" and "1.8L Hybrid", and finds out target vehicles meeting both "Corolla" and "1.8L Hybrid". It is assumed here that there are 3 target vehicles in total, and the production years of these target vehicles are 2022, 2021, and 2019 respectively. The server supplements the year information as missing information to the query "What is the fuel consumption of the Toyota Corolla 1.8L Hybrid?" and obtains the secondary reformulated query "What is the fuel consumption of the Toyota Corolla 1.8L Hybrid? The production years are: 2022, 2021, and 2019.". Finally, the server submits the secondary reformulated query to other pre-trained large models (e.g., an automotive query-answer model) to get an answer to the query.

[0094] Based on the same inventive concept as the training method for the query reformulation model provided in this embodiment, this embodiment further provides a training apparatus for a query reformulation model, which includes at least one software function module that can be stored in a memory in the form of software or fixed in an electronic device. The processor in the electronic device is configured to execute executable modules stored in the memory, for example, the software function modules and computer programs included in the apparatus. Referring to FIG. 4, the apparatus may include the following functional components:

[0095] a sample obtaining module 11A, configured to obtain a fine-tuning sample of a pre-trained model, wherein the fine-tuning sample includes a query reformulation sample and a reformulation auxiliary sample in the automotive field, and the reformulation auxiliary sample is configured to enhance a comprehension ability and a reformulation ability of the pre-trained model to a query; and

[0096] a model training model 12A, configured to train the pre-trained model with the fine-tuning sample to obtain a query reformulation model.

[0097] In this embodiment, the sample obtaining module 11A is configured to implement the step S1A in FIG. 1, and the model training module 12A is configured to implement the step S2A in FIG. 1. Therefore, for the detailed description of the above modules, reference may be

made to specific embodiments of corresponding steps, which are not described in detail in this embodiment.

[0098] In addition, since having the same inventive concept as the training method for the query reformulation model, the training apparatus for the query reformulation model can also implement other steps or substeps of the method by the modules.

[0099] Optionally, the model training module 12A is further specifically configured to:

[00100] train a pre-trained model with a query reformulation sample, a reading comprehension sample and a pronoun identification sample to obtain an initial model; and

[00101] perform hybrid training on the initial model with the query reformulation sample, the reading comprehension sample and the pronoun identification sample to obtain a query reformulation model.

[00102] Optionally, the sample obtaining module 11A is further specifically configured to:

[00103] obtain a user query in the automotive field and a context of the user query;

[00104] process the user query and the context of the user query using a third-party large model to generate a reformulation content of the user query, where the reformulation content includes an inference part and a query reformulation part; and

[00105] hide part of the reformulation content, and take the hidden content and the user query as a query reformulation sample.

[00106] In addition, this embodiment further provides a query reformulation apparatus. As shown in FIG. 5, the apparatus includes:

[00107] a query receiving module 11B, configured to receive an original query input by a user; and

[00108] a query reformulation module 12B, configured to process the original query and a preset automotive brand tree by the query reformulation model to obtain a reformulated query of the original query, wherein the automotive brand tree records a plurality of automobile brands and vehicle information of vehicles released by each automobile brand.

[00109] In this embodiment, the query receiving module 11B is configured to implement the step S1B in FIG. 2, and the query reformulation module 12B is configured to implement the step S2B in FIG. 2. Therefore, for the detailed description of the above modules, reference can be made to specific embodiments of corresponding steps, which are not described in detail in this embodiment.

[00110] Since having the same inventive concept as the query reformulation method, the query reformulation apparatus may also implement other steps or substeps of the method by the modules.

[00111] Optionally, the query reformulation module 12B is further configured to:

[00112] obtain existing elements in the reformulated query;

[00113] take information corresponding to the missing elements in the automotive brand tree as missing information of the reformulated query if the existing elements are compared with the preset complete elements and there are missing elements in the reformulated query; and

[00114] supplement the missing information into the reformulated query to obtain a secondary reformulated query of the original query.

[00115] In addition, the functional modules in the embodiments of the present disclosure can be integrated together to form an independent part, or each module can exist independently, or two or more modules can be integrated to form an independent part.

[00116] It should also be understood that if the above embodiments are implemented in a form of a software functional module and sold or used as an independent product, the functions can be stored in a computer-readable storage medium. Based on such understanding, the technical solutions of the present disclosure essentially can be, or part of the technical solutions contributing to the prior art can be, or part of the technical solutions can be embodied in the form of a software product. The computer software product is stored in a storage medium and includes several instructions for enabling a computer device (which can be a personal computer, a server, a network device or the like) to implement all or part of the steps of the method described in the embodiments of the present disclosure.

[00117] Therefore, this embodiment further provides a storage medium, where the storage medium stores a computer program, and the computer program, when executed by a processor, implements the query reformulation method provided in this embodiment. The storage medium may be any medium that can store program code, such as a USB flash drive, a removable hard disk, a read-only memory (ROM), a random access memory (RAM), a magnetic disk, or an optical disc.

[00118] This embodiment provides an electronic device for implementing the training method for the query reformulation model and the query reformulation method. As shown in FIG. 6, the electronic device can include a processor 22 and a memory 21. The memory 21 stores a computer program, and the processor reads and executes the computer program corresponding to the above embodiment in the memory 21, thereby implementing the training method for the query reformulation model and the query reformulation method provided in this embodiment.

[00119] Still referring to FIG. 6, the electronic device further includes a communication unit 23. The memory 21, the processor 22, and the communication unit 23 are electrically connected

to each other directly or indirectly via a system bus 24 to achieve data transmission or interaction.

[00120] The memory 21 can be an information recording device based on any electronic, magnetic, optical or other physical principles, configured to record execution instructions, data and the like. In some embodiments, the memory 21 can be, but is not limited to, a volatile memory, a non-volatile memory, a storage drive, and the like.

[00121] In some embodiments, the volatile memory can be a random access memory (RAM). In some embodiments, the non-volatile memory can be a read-only memory (ROM), a programmable read-only memory (PROM), an erasable programmable read-only memory (EPROM), an electric erasable programmable read-only memory (EEPROM), a flash memory, and the like. In some embodiments, the storage drive can be a disk drive, a solid state drive, any type of storage disk (such as a CD and a DVD), or a similar storage medium, or a combination thereof, or the like.

[00122] The communication unit 23 is configured to transmit and receive data via a network. In some embodiments, the network can include a wired network, a wireless network, a fiber optic network, a telecommunications network, an intranet, the Internet, a local area network (LAN), a wide area network (WAN), a wireless local area network (WLAN), a metropolitan area network (MAN), a wide area network (WAN), a public switched telephone network (PSTN), a Bluetooth network, a ZigBee network, or a near field communication (NFC) network, or any combination thereof. In some embodiments, the network can include one or more network access points. For example, the network may include wired or wireless network access points, such as base stations and/or network switching nodes, through which one or more components of the service request processing system can connect to the network to exchange data and/or information.

[00123] The processor 22 can be an integrated circuit chip having a signal processing capability, and the processor can include one or more processing cores (e.g., a single-core processor or a multi-core processor). By way of example only, the processors can include a central processing unit (CPU), an application specific integrated circuit (ASIC), an application specific instruction-set processor (ASIP), a graphics processing unit (GPU), a physical processing unit (PPU), a digital signal processor (DSP), a field programmable gate array (FPGA), a programmable logic device (PLD), a controller, a microcontroller unit, a reduced instruction set computer (RISC), or a microprocessor, or any combination thereof.

[00124] It should be understood that the structure shown in FIG. 6 is for illustration only. The electronic device can also have more or fewer components than those shown in FIG. 6, or can

have a configuration different from that shown in FIG. 6. The components shown in FIG. 6 can be implemented in hardware, software, or a combination thereof.

[00125] It should be understood that the apparatus and methods disclosed in the above embodiments can also be implemented in other ways. The apparatus embodiments described above are merely illustrative. For example, flowcharts and block diagrams in the drawings show systematic architectures, functions, and operations of the apparatus, the method and the computer program product possibly implemented according to a plurality of embodiments of the present disclosure. In this regard, each block in the flowcharts or the block diagrams can represent a portion of a module, a program segment or codes, where the portion of the module, the program segment or the codes includes one or more executable instructions for implementing specified logical functions. It should also be noted that, in some alternative implementations, the functions shown in the blocks can occur in an order different from the order shown in the drawings. For example, two consecutive blocks can, in fact, be executed substantially in parallel, and the two blocks can sometimes be executed in a reverse order, depending upon the functions involved. It should also be noted that each block of the block diagrams and/or the flowcharts and a combination of the blocks in the block diagrams and/or the flowcharts can be implemented through a dedicated hardware-based system that executes a specified function or operation, or can be implemented through a combination of a dedicated hardware and a computer instruction.

[00126] The above description is merely the various embodiments of the present disclosure. However, the protection scope of the present disclosure is not limited thereto, and any modifications and substitutions that can be easily conceived by those skilled in the art within the technical scope disclosed by the present disclosure shall fall within the protection scope of the present disclosure. Therefore, the protection scope of the present disclosure shall be subject to the protection scope of the claims.

INDUSTRIAL PRACTICALITY

[00127] In summary, the present disclosure provides a training method for a query reformulation model, a query reformulation method, and related apparatuses, in which not only a query reformulation sample in the automotive field is used during model training, but also an auxiliary sample configured to enhance the comprehension ability and reformulation ability of a pre-trained model to a query is introduced, so that the trained query reformulation model can comprehend queries in the automotive field and improve the accuracy of reformulated queries.

CLAIMS

What is claimed is:

Claim 1: A training method for a query reformulation model, comprising:

obtaining a fine-tuning sample of a pre-trained model, wherein the fine-tuning sample comprises a query reformulation sample and a reformulation auxiliary sample in the automotive field, and the reformulation auxiliary sample is configured to enhance a comprehension ability and a reformulation ability of the pre-trained model to a query; and

training the pre-trained model with the fine-tuning sample to obtain a query reformulation model.

Claim 2: The training method for the query reformulation model according to claim 1, wherein the reformulation auxiliary sample comprises a reading comprehension sample and a pronoun identification sample.

Claim 3: The training method for the query reformulation model according to claim 2, wherein the step of training the pre-trained model with the fine-tuning sample to obtain a query reformulation model comprises:

training the pre-trained model with the query reformulation sample, the reading comprehension sample and the pronoun identification sample to obtain an initial model; and

performing hybrid training on the initial model with the query reformulation sample, the reading comprehension sample and the pronoun identification sample to obtain the query reformulation model.

Claim 4: The training method for the query reformulation model according to claim 3, wherein the hybrid training refers to that the query reformulation sample, the

reading comprehension sample and the pronoun identification sample are used simultaneously in a same training process, and training losses of the query reformulation sample, the reading comprehension sample and the pronoun identification sample are weighted and summed to obtain a comprehensive model loss, so that parameters of the initial model are adjusted.

Claim 5: The training method for the query reformulation model according to any one of claims 1 to 4, wherein the step of obtaining a query reformulation sample of the pre-trained model comprises:

obtaining a user query in the automotive field and a context of the user query;

processing the user query and the context of the user query using a third-party large model to generate a reformulation content of the user query, wherein the reformulation content comprises an inference part and a query reformulation part; and

hiding part of the reformulation content, and taking the hidden content and the user query as the query reformulation sample.

Claim 6: The training method for the query reformulation model according to claim 5, wherein the step of hiding part of the reformulation content and taking the hidden content and the user query as the query reformulation sample comprises:

hiding an inference part in the reformulation content, and taking the query reformulation part and the user query as the query reformulation sample.

Claim 7: A query reformulation method, comprising:

receiving an original query input by a user; and

processing the original query and a preset automotive brand tree by the query reformulation model obtained by the method according to any one of claims 1

to 6 to obtain a reformulated query of the original query, wherein the automotive brand tree records a plurality of automobile brands and vehicle information of vehicles released by each automobile brand.

Claim 8: The query reformulation method according to claim 7, further comprising:

obtaining existing elements in the reformulated query;

taking information corresponding to the missing elements in the automotive brand tree as missing information of the reformulated query when the existing elements are compared with preset complete elements and there are missing elements in the reformulated query; and

supplementing the missing information into the reformulated query to obtain a secondary reformulated query of the original query.

Claim 9: The query reformulation method according to claim 8, wherein the step of taking information corresponding to the missing elements in the automotive brand tree as missing information of the reformulated query comprises:

obtaining existing information corresponding to the existing elements from the reformulated query, and screening the automotive brand tree by using the existing information so as to match a target vehicle with the existing information; and

obtaining information corresponding to the missing element from vehicle information of the target vehicle as the missing information of the reformulated query.

Claim 10: A training apparatus for a query reformulation model, comprising:

a sample obtaining module, configured to obtain a fine-tuning sample of a pre-trained model, wherein the fine-tuning sample comprises a query reformulation

sample and a reformulation auxiliary sample in the automotive field, and the reformulation auxiliary sample is configured to enhance a comprehension ability and a reformulation ability of the pre-trained model to a query; and

a model training model, configured to train the pre-trained model with the fine-tuning sample to obtain a query reformulation model.

Claim 11: A query reformulation apparatus, comprising:

a query receiving module, configured to receive an original query input by a user; and

a query reformulation module, configured to process the original query and a preset automotive brand tree by the query reformulation model obtained by the method according to any one of claims 1 to 5 to obtain a reformulated query of the original query, wherein the automotive brand tree records a plurality of automobile brands and vehicle information of vehicles released by each automobile brand.

Claim 12: A storage medium, wherein the storage medium stores a computer program, and the computer program, when executed by a processor, implements the training method for the query reformulation model according to any one of claims 1 to 6 and the query reformulation method according to any one of claims 7 to 9.

Claim 13: An electronic device, wherein the electronic device comprises a processor and a memory, the memory stores a computer program, and the computer program, when executed by the processor, implements the training method for the query reformulation model according to any one of claims 1 to 6 and the query reformulation method according to any one of claims 7 to 9.