

A SYSTEM FOR AND METHOD OF CONSTRUCTING A RESPONSE KNOWLEDGE BASE

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of priority to the Singapore application no. 10202401018W filed 8 April, 2024, the contents of which are hereby incorporated by reference in their entirety for all purposes.

TECHNICAL FIELD

[0002] This application relates generally to the field of conversational knowledge base, and more particularly, to a system and method of constructing a response knowledge base.

BACKGROUND

[0003] Conventional approach towards assessing conversational quality often relies on the assessment of a subject matter expert, which is subjective in nature and non-quantifiable. As an example, a sales team with better sales performance are viewed to have a better conversational quality which may not be an accurate representation. Such conventional methods are unable to define or quantify merits in responses or replies, nor is able to associate target goals with quality of responses.

SUMMARY

[0004] According to an aspect, disclosed herein a system. The system comprises: memory storing instructions; and a processor coupled to the memory and configured to process the stored instructions to implement: a module configured to perform a method of constructing a response knowledge base, the method including: providing a candidate dataset to a first machine learning

arm, the first machine learning arm trained based on a first target outcome, wherein the candidate dataset comprises a plurality of candidate prompt-response pairs; providing the candidate dataset to a second machine learning arm, the second machine learning arm trained based on a second target outcome; and based on a first output from the first machine learning arm and a second output from the second machine learning arm, determining a leading candidate prompt-response pair from the plurality of candidate prompt-response pairs; and incorporating the leading candidate prompt-response pair into the response knowledge base.

[0005] According to another aspect, disclosed herein a method to constructing a response knowledge base. The method comprises: providing a candidate dataset to a first machine learning arm, the first machine learning arm trained based on a first target outcome, wherein the candidate dataset comprises a plurality of candidate prompt-response pairs; providing the candidate dataset to a second machine learning arm, the second machine learning arm trained based on a second target outcome; based on a first output from the first machine learning arm and a second output from the second machine learning arm, determining a leading candidate prompt-response pair from the plurality of candidate prompt-response pairs; and incorporating the leading candidate prompt-response pair into the response knowledge base.

[0006] According to another aspect, disclosed herein a response knowledge base constructed based on the method as recited above. The response knowledge base comprises a plurality of reference prompts and a plurality of reference responses, wherein multiple ones of the plurality of reference responses correspond to one of the plurality of reference prompts.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Various embodiments of the present disclosure are described below with reference to the following drawings:

[0008] FIG. 1 is a schematic diagram showing a system according to embodiments of the present disclosure;

[0009] FIG. 2 is a schematic diagram showing a structure of a machine learning arm according to various embodiments;

[0010] FIG. 3 is a schematic diagram illustrating a training of the machine learning arm of FIG. 2;

[0011] FIG. 4 is a schematic diagram showing a structure of another machine learning arm according to various embodiments;

[0012] FIG. 5 is a schematic diagram illustrating a training of the machine learning arm of FIG. 4;

[0013] FIG. 6 is a schematic diagram showing a system with a common feature extractor according to embodiments of the present disclosure;

[0014] FIG. 7 is a flow chart illustrating a method of constructing a response knowledge base according to embodiments of the present disclosure; and

[0015] FIG. 8 is a schematic diagram of a processor system.

DETAILED DESCRIPTION

[0016] The following detailed description is made with reference to the accompanying drawings, showing details and embodiments of the present disclosure for the purposes of illustration. Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments, even if not explicitly described in these other embodiments. Additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0017] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0018] In the context of various embodiments, the term “about” or “approximately” as applied to a numeric value encompasses the exact value and a reasonable variance as generally understood in the relevant technical field, e.g., within 10% of the specified value.

[0019] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0020] As used herein, terms “concurrently”, “simultaneously”, “at the same time”, or the like, may refer to events or actions that coincide or overlap within a period of time, regardless of whether the events start at the same time instant, and regardless of whether the events end at the same time instant.

[0021] As used herein, the term “prompt” may be used interchangeably with the terms “question”, “remark”, “hint”, “mention”, “enquiry”, “inquiry”, etc. and may generally refer to a sentence worded or expressed by a party or a questioning party, such as a customer or user, so as to elicit information or to elicit a response from a counter party or answering party, such as a sales agent or an automated assistant.

[0022] As used herein, the term “response” may be used interchangeably with “answer”, “reply”, “observation”, “comment”, “respond”, etc. and may generally refer to a sentence worded or expressed by an answering party, such as a sales agent or an automated assistant, as a reaction or response to a prompt, a question, or an inquiry from a questioning party, such as a customer or user.

[0023] Conventional approach towards assessing conversational quality often relies on the assessment of a subject matter expert, which is subjective in nature and non-quantifiable. In the context of telesales or tele-marketing, the general consensus is that in response to specific questions from the customers/target subject, certain responses, phrases or keywords used by the

sales agent may help in achieving predetermined target outcomes, such as increasing the success rate of transaction or sales, or to increasing the call duration.

[0024] In the context of automated conversation system, such as intelligent car assistants, conventional systems often act as a receiver of instructions, such as performing a function as instructed or providing information as requested, and hence often comprises only terse, simple or predictable responses. Therefore, interactions between the car owner and the intelligent car assistant are often unable to reflect well the deeper conversational responses and insights into the car owner's experience. Examples from such car owner's experience may include a measure of satisfaction of the car owner.

[0025] The present disclosure comprises a system and a method of constructing a response knowledge base according to various embodiments. The present disclosure also includes a response knowledge base comprising a plurality of prompts and a corresponding plurality of responses. The response knowledge base may be used to provide a suitable response or answer in reply to a prompt or question.

[0026] The proposed system and method of the present disclosure may find application in various fields, such as telesales or customer service scenarios. In addition, the proposed system and method may be utilized for quantitatively identifying examples of responses which may be helpful towards a predetermined target outcome. In an example, the proposed system and method may aid in identifying positive responses in the sales-agent's conversation which may increase the probability of sales order. In addition, by way of learning from historical target outcomes or intentions and comparing agent responses with sales outcomes, the proposed system and method may determine words or phrases which promotes sales thus improving the performance of the sales-agents.

[0027] In the non-limiting scenario of telesales, the response knowledge base may serve as a reference for sales-agents in assisting with providing a suitable response in view of a

customer's prompt or question. According to the nature of the customer's prompt taking into consideration the target outcome, the sales-agent may determine a suitable response. The suitable response may be determined based on a probability of achieving the target outcome, such as a highest probability of achieving the target outcome. As an example, the sales-agent may be faced with a complaint from a client with a two-fold or dual target outcome, e.g. i) to reduce dissatisfaction of the customer and ii) to improve success rate of a future sales. In the example, the sales-agent may select suitable responses with higher weights on reducing dissatisfaction in the initial stage of the conversation, and to select suitable responses with higher weights on improve success rate of a future sales in the later stage.

[0028] The invention may also be applied to a multi-modal scene of the intelligent car assistance session, such as: talking about the beautiful scenery outside the car and what you see and hear with the intelligent assistant and can also assist in solving on-site problems in the form of audio and video when a fault or accident occurs. In such scenarios, it can help identify the specific response method and content of the reply, which is more helpful to improve the satisfaction of the car owner.

[0029] In the non-limiting scenario of intelligent car assistants, the response knowledge base may serve as a reference for the onboard system in responding to the car owner's or car user's prompt during a trip. It may be appreciated that the prompt need not be limited to a question or an inquiry. In various embodiments, the prompt may be a comment or a remark from the car user, about the scenery or events happening outside the car interior. In such an instance, the intelligent car assistant may take input not only from the car user but also other inputs such as visual and/or audio inputs of the environment, such as speech from a non-car owner subject, pictures of the environment, videos of the events, etc. As such, the prompt and/or the response may be multi-modal in nature, i.e. a video and speech prompt, or a text and video response. In

other example, the intelligent car assistant may assist in solving technical difficulties / faults during a breakdown or assist in communicating with a workshop after an accident.

[0030] In various embodiments, the response knowledge base constructed based on the proposed system and method may provide a plurality of responses in reply to a prompt of question. The plurality of responses may be based on a plurality of target outcomes. The response knowledge base may provide a suitable response considering a plurality of target outcomes in combination. In some embodiments, the response knowledge base may provide a suitable response based on a single one of the plurality of target outcomes. In other words, each of the plurality of responses may be a most suitable response corresponding to a respective one or combination of the plurality of target outcomes.

[0031] In various embodiments, the plurality of responses may be suitable responses corresponding to different target outcomes, such as increase in call duration, increase in success rate of sales, increase in customer satisfaction, improvement in customer experience, reduction in dissatisfaction when faced with a complaint, increase in transaction volume, increase in transaction amount, etc.

[0032] Referring to FIG. 1, in one aspect, the system 200 according to some embodiments of the present disclosure is configurable as a tool to construct a response knowledge base 400 based on a candidate dataset 120. The system 200 may include a module configured to perform a method of constructing a response knowledge base 400 based on a candidate dataset 120.

[0033] The candidate dataset 120 may be viewed as an input dataset awaiting quality assessment. The candidate dataset 120 may include a plurality of candidate prompt-response pairs (P1&R1/P2&R2/.../PN&RN) or in other words, a plurality of candidate prompts P1/P2/.../PN and a plurality of candidate responses R1/R2/.../RN. The candidate dataset 120 may be formed or compiled based performing searches on various search engines, using various web-crawlers, etc.

[0034] In various embodiments, the candidate dataset 120 may be a single mode dataset, i.e. comprising purely text content or purely audio content. In other embodiments, the candidate dataset 120 may be a multimodal dataset, i.e. comprising a combination of text, audio, visual content.

[0035] In various embodiments as shown in FIG. 1, the system 200 may include a plurality of machine learning arms 220/240, each of the plurality of machine learning arms 220/240 connects to a respective weight $W1/W2$. In an exemplary embodiment, the system 200 may include a first machine learning arm 220 connected to a first weight $W1$; and a second machine learning arm 240 connected to a second weight $W2$.

[0036] In various embodiments, the first machine learning arm 220 may be trained based on a first target outcome and the second machine learning arm 240 may be trained based on a second target outcome. The first weight $W1$ may correspond to the weightage or importance of the first target outcome. Similarly, the second weight $W2$ may correspond to the weightage or importance of the second target outcome. The first weight $W1$ and second weight $W2$ may be determined, set or selected by a user of the system 200 depending on the goal of constructing the response knowledge base 400.

[0037] In various embodiments, the system 200 may include any other numbers of machine learning arms, such as five machine learning arms each trained based on a respective target outcome.

[0038] In various embodiments, the system 200 may construct the response knowledge base 400 using multiple ones of the first weight $W1$ and multiple ones of the second weight $W2$, or in other words, varying the first weight $W1$ and second weight $W2$ based on a same candidate dataset. This enables the construction of a wide variety of different combinations of target outcomes.

[0039] In various embodiments, the response knowledge base 400 may include a plurality of reference prompt-response pairs (PR1&RR1/PR2&RR2/.../PRN&RRN). The plurality of reference prompt-response pairs (PR1&RR1/PR2&RR2/.../PRN&RRN) may include a plurality of reference prompts PR1/PR2/.../PRN and a plurality of reference responses RR1/RR2/.../RRN. In various embodiments, each of the plurality of reference responses RR1/RR2/.../RRN may correspond to one of the plurality of reference prompts PR1/PR2/.../PRN. In various embodiments, multiple ones of the plurality of reference responses (e.g. RR1/RR2) may correspond to one of the plurality of reference prompts (e.g. PR1).

[0040] In various embodiments, the system 200 may further include a comparison model 320 in signal communication with the response knowledge base 400. The comparison model 320 may compute or determine a similarity score based on a selected reference prompt PR_i from the response knowledge base 400 and each of the plurality of candidate prompt-response pairs (P1&R1/P2&R2/.../PN&RN). In various embodiments, the comparison model 320 may compute the similarity score based on a selected reference prompt PR_i and each of the candidate prompt P1/P2/.../PN of the candidate dataset 120.

[0041] Still referring to FIG. 1, operation of the system 200 may be described as follows. In various embodiments, a method of constructing a response knowledge base 400 begins with providing an input 125 from the candidate dataset 120 or a portion of the candidate dataset 120 to a first machine learning arm 220 of the system 200. Concurrently or sequentially, the candidate dataset 120 or the portion of the candidate dataset 120 is independently and parallelly provided to the second machine learning arm 240 as well as the comparison model 320 of the system 200.

[0042] Responsive to inputting the candidate dataset 120 to the first machine learning arm 220, the first machine learning arm 220 provides a first output 225. The first output 225 may

comprise a plurality of first probabilities. Each of the plurality of first probabilities may correspond to a probability of achieving the first target outcome for the corresponding candidate prompt-response pair (P1&R1/P2&R2/.../PN&RN).

[0043] Similarly, responsive to inputting the candidate dataset 120 to the second machine learning arm 240, the second machine learning arm 240 provides a second output 245. The second output 245 may comprise a plurality of second probabilities. Each of the plurality of second probabilities may correspond to a probability of achieving the second target outcome for the corresponding candidate prompt-response pair (P1&R1/P2&R2/.../PN&RN).

[0044] In addition, responsive to inputting the candidate dataset to the comparison model 320, the system 200 may compute a similarity output of the comparison model 320. The similarity output may include a respective similarity score 325 between each of the plurality of candidate prompt-response pairs (P1&R1/P2&R2/.../PN&RN) and a selected reference prompt PR_i from the response knowledge base 400. In various embodiments, a similarity output of the comparison model 320 corresponds to a plurality of similarity scores 325.

[0045] Each of the similarity score 325 may correspond to a probability of similarity between a selected reference prompt PR_i from the response knowledge base 400 and the corresponding candidate prompt-response pairs (P1&R1/P2&R2/.../PN&RN). In various embodiments, each of the plurality of similarity probabilities may correspond to a probability of similarity between a selected reference prompt PR_i and the respective candidate prompt P1/P2/.../PN. In an example, a probability of similarity may range between “0” and “1”, wherein “1” corresponds to the pair of selected reference prompt PR_i and the respective candidate prompt (e.g. P1) being identical.

[0046] Further, the system 200 may take a weighted sum of the first output 225, the second output 245, and the similarity score 325. In various embodiments, the system 200 may determine a first weighted output 227 based on the first output 225 and the first weight W1.

Similarly, the system 200 may determine a second weighted output 247 based on the second output 245 and the second weight W2. Further, the system 200 may also determine a weighted similarity score 327 based on the similarity score 325 and a similarity weight WS. In various embodiments, a sum of the first weight W1, the second weight W2 and the similarity weight WS is equal to 1.

[0047] In various embodiments, the system 200 may determine a sum of the first weighted output 227, the second weighted output 247, and the weighted similarity score 327 to determine a weighted outcome 420. In various embodiments, the weighted outcome 420 may include a plurality of weighted probabilities corresponding to a combination of the first weighted output 227 and the second weighted output 247. In various embodiments, the weighted outcome 420 may include a plurality of weighted probabilities corresponding to a combination of the first weighted output 227, the second weighted output 247, and the weighted similarity score 327.

[0048] The weighted outcome 420 may include a plurality of weighted probabilities corresponding to a combination of the first target outcome and the second target outcome. By adjusting the first weight W1 and the second weight W2, the system 200 may achieve different combinations of target outcomes. The weighted outcome 420 may also take into consideration the weightage of the similarity score between each candidate prompt-response pairs (P1&R1/P2&R2/.../PN&RN) of the candidate dataset 120 and the selected reference prompt PRi from the response knowledge base 400. For example, the system 200 may adjust the similarity weight WS to give more or less weightage to similarity scores in determining the weighted outcome 420.

[0049] Based on the weighted outcome 420 and the plurality of weighted probabilities, the system 200 may then determine a leading candidate prompt-response pair 422 from the plurality of candidate prompt-response pairs (P1&R1/P2&R2/.../PN&RN). In various embodiments, the leading candidate prompt-response pair 422 may be determined based on one with the highest

weighted outcome 420 or weighted probability, i.e 0.7 as shown in FIG. 1. The leading candidate prompt-response pair 422 may then be incorporated into the response knowledge base 400.

[0050] In various embodiments, the leading candidate prompt-response pair 422 (e.g. P1&R1) may be incorporated into the response knowledge base 400 by replacing one of the plurality of reference responses (e.g. PR2/RR2) with the candidate response R1 of the leading candidate prompt-response pair 422 (e.g. P1/R1). In some embodiments, the corresponding selected reference response RR_i of the selected reference prompt PR_i may be replaced by the leading candidate prompt-response pair 422.

[0051] In other embodiments, the leading candidate prompt-response pair 422 may be incorporated into the response knowledge base 400 without replacing any of the plurality of reference responses RR₁/RR₂/.../RR_N. In yet other embodiments, the replacement or addition of the leading candidate prompt-response pair 422 may be determined based on the weighted probability of the leading candidate prompt-response pair 422 and a threshold. In yet other embodiments, the replacement or addition of the leading candidate prompt-response pair 422 may be determined based on the weighted probability of the leading candidate prompt-response pair 422 and the corresponding weighted probability as previously computed for the selected reference prompt-response pair PR_i&RR_i.

[0052] It may be appreciated that during the initial phases or stages of constructing the response knowledge base 400, the leading candidate prompt-response pair 422 may typically be added to the response knowledge base 400 instead of replacing any of the selected reference prompt-response pair PR_i&RR_i. The above-described method of constructing the response knowledge base 400 may be iteratively performed based on different candidate dataset 120 as input.

[0053] In an exemplary response knowledge base constructed by the system 200 as shown in Table 1, the response knowledge base may include a plurality of reference prompts PR1/PR2/.../PRN and a plurality of reference responses RR1/RR2/.../RRN. In various embodiments, multiple ones of the plurality of reference responses (e.g. RR1/RR2) may correspond to one of the plurality of reference prompts (e.g. PR1). In other words, the response knowledge base 400 may include multiple corresponding reference responses (e.g. RR1/RR2) to a common reference prompt (e.g. PR1 being identical to PR2). Referring to Table 1, the response knowledge base 400 may include multiple reference responses (RR1, RR2) directed towards different target outcomes for a common reference prompt (PR1).

Table 1. Exemplary response knowledge base

	Reference Prompt		Reference Response
PR1	The scenery is beautiful	RR1	You may also consider visiting the nearest national park located 15 minutes away
PR2	The scenery is beautiful	RR2	You may consider arriving earlier the next time coinciding with the sunrise.
...	...	...	...
PRN	...	RRN	...

[0054] In various embodiments, the candidate dataset 120 may include multiple candidate prompt-response pairs (P1&R1/P2&R2/.../PN&RN) with identical prompts but differing responses, i.e. P1/P2/.../PN are the same but R1/R2/.../RN are different. This illustrates a scenario where the same prompt is used to determine a most suitable response. In other embodiments, the dataset 120 may include multiple candidate prompt-response pairs (P1&R1/P2&R2/.../PN&RN) with differing prompts and corresponding differing responses. This illustrates a scenario where an unstructured candidate dataset is used for constructing the response knowledge base.

[0055] FIG. 2 illustrates a first machine learning arm 220 according to various embodiments. The first machine learning arm 220 may comprise a first feature extractor 222 in serial

connection with a first machine learning model 224. In various embodiments, the first feature extractor 222 may be a text-based feature extractor such as sentence transformer models for text encoding. In an example, the first feature extractor 222 may use a SBERT (Sentence-Bidirectional Encoder Representations from Transformers) encoding method based on Word2Vec, which is a trained sentence transformer model for vectorized coding. In other examples, methods such as One-Hot encoding, TF-IDF (Term frequency-inverse document frequency) method, etc. may also be used. The sentence transformer model may include capabilities of semantic understanding of text inputs, which helps in obtaining the semantic features of the text inputs and aids in later machine learning model training.

[0056] In various embodiments, the first machine learning model 224 may be a deep learning model such as a multi-layer perceptron (MLP) model, a convolutional neural network (CNN) model or a Generative pre-trained transformer (GPT) model. The MLP may include a first input layer, a plurality of first hidden layers, and a first output layer. In an example, the plurality of first hidden layers may include 10 to 20 layers. In another example, the plurality of first hidden layers may include 15 layers. The first machine learning model 224 may be trained based on a first target outcome, such as success rate in completing sales. It may be appreciated that other deep learning models may also be utilized as the first machine learning model 224.

[0057] FIG. 2 may also be used to illustrate a second machine learning arm 240 according to various embodiments. Similarly, the second machine learning arm 240 may comprise a second feature extractor 242 in serial connection with a second machine learning model 244. In various embodiments, similar to the first feature extractor 222, the second feature extractor 242 may also be a text-based feature extractor such as sentence transformer models for text encoding.

[0058] In various embodiments, the second machine learning model 244 may be a deep learning model such as a multi-layer perceptron (MLP) model, a convolutional neural network

(CNN) model or a Generative pre-trained transformer (GPT) model. The MLP may include a second input layer, a plurality of second hidden layers, and a second output layer. In an example, the plurality of second hidden layers may include 10 to 20 layers. In another example, the plurality of second hidden layers may include 12 layers. The second machine learning model 244 may be trained based on a second target outcome, such as customer satisfaction.

[0059] In various embodiments, the system 200 may provide an input 125 from the candidate dataset 120 to the first feature extractor 222 to obtain a first feature set 223, and this may be followed by providing the first feature set 223 to the first machine learning model 224. The first feature set 223 may correspond to embeddings from the input 125. Thereafter, the first machine learning model 224 may provide the first output 225 corresponding to the first target outcome. Similarly, the system 200 may also provide the input 125 from the candidate dataset 120 to the second feature extractor 242 to obtain a second feature set 243, and this may be followed by providing the second feature set 243 to the second machine learning model 244. The second feature set 243 may correspond to embeddings from the input 125. Thereafter, the second machine learning model 244 may provide the second output 245 corresponding to the second target outcome. In various embodiments, the first feature extractor 222 may be configured differently from the second feature extractor 242. As such, the first feature set 223 may be different from the second feature set 243.

[0060] FIG. 3 illustrate a method of training the first machine learning arm 220 and/or a second machine learning arm 240 according to various embodiments. The method of training may include training the first machine learning arm 220 based on a first training dataset 130. The first training dataset 130 may include a plurality of training prompt-response pairs T1/T2/.../TN each labelled with a respective first target outcome 134. In various embodiments, each of the first target outcome 134 may be a value corresponding to a first desired outcome.

[0061] During the training process of the first machine learning arm 220, the plurality of training prompt-response pairs $T1/T2/.../TN$ of the first training dataset 130 may be provided as input 135 to the first feature extractor 222 to obtain a first feature set 223, and the first feature set 223 may be provided to the first machine learning model 224. The first machine learning model 224 may be trained based on the first target outcome 134.

[0062] The method of training may also include training the second machine learning arm 240 based on a second training dataset 132. The second training dataset 132 may also include a plurality of training prompt-response pairs $T1/T2/.../TN$ each labelled with a respective second target outcome 136. In various embodiments, each of the second target outcome 136 may be a value corresponding to a second desired outcome.

[0063] Similarly, during the training process of the second machine learning arm 240, the plurality of training prompt-response pairs $T1/T2/.../TN$ of the second training dataset 132 may also be provided as input 135 to the second feature extractor 242 to obtain a second feature set 243, and the second feature set 243 may be provided to the second machine learning model 244. The second machine learning model 244 may be trained based on the second target outcome 136.

[0064] In various embodiments, the first machine learning arm 220 and the second machine learning arm 240 may be trained independently from one another. In various embodiments, each of the first target outcome 134 or second target outcome 136 may be quantifiable by a binary value, such as 0 or 1, corresponding to a success or a failure. As an example, for the first training dataset 130, each of the plurality of training prompt-response pairs $T1/T2/.../TN$ may be labelled with “1” corresponding to “successful sales” or “0” corresponding to “unsuccessful sales”. As another example, for the second training dataset 132, each of the plurality of training prompt-response pairs $T1/T2/.../TN$ may be labelled with “1” corresponding to “increase in

call duration by 30 seconds or more” or “0” corresponding to “increase in call duration below 30 seconds”.

[0065] In various embodiments, the first training dataset 130 and the second training dataset 132 may each include 500 to 1000 training prompt-response pairs T1/T2/.../T500 or T1/T2/.../T1000. In various embodiments, with reference to the first target outcome 134, half of the first training dataset 130 may be labelled with “1” or “successful” and the other half of the first training dataset 130 may be labelled with “0” or “unsuccessful”. Similarly, with reference to the second target outcome 136, half of the second training dataset 132 may be labelled with “1” or “successful” and the other half of the second training dataset 132 may be labelled with “0” or “unsuccessful”. The ratio between the number of “1” labels and the number of “0” labels may be determined by a user.

[0066] In various embodiments, the training prompt-response pairs T1/T2/.../TN may be deterministic in nature. In other words, the training prompt may include a clear indication / goal, and the respective prompt response pair may include a clear outcome, i.e. may be labelled as “1” or as “0” distinctively.

[0067] In some embodiments, each of the first machine learning arm 220 and the second machine learning arm 240 may be trained with different training dataset 130/132 or different training prompt-response pairs T1/T2/.../TN. However, the different training datasets 130/132 may be labelled with a same classification of target outcomes. As such, the first target outcome 134 of the first training dataset 130 and the second target outcome 136 of the second training dataset 132 may correspond to each other. For example, both the first training dataset 130 and the second target outcome 136 may be labelled for a common target outcome of “improving customer experience”.

[0068] FIG. 4 illustrate a scenario where the system 200 is provided with a candidate dataset which is a multi-modal candidate dataset 120. FIG. 4 illustrates the first machine learning arm

220 receiving the multi-modal candidate dataset 120. In various embodiments, the multi-modal candidate dataset 120 may include a combination of text content, audio content, visual content, 3D contents, etc.

[0069] In various embodiments, the multi-modal candidate dataset 120 may include a plurality of multi-modal candidate prompt-response pairs. Each of the plurality of multi-modal candidate prompt-response pairs may include a multi-modal candidate prompts and a multi-modal candidate responses. In various embodiments, each of the multi-modal candidate prompt and the multi-modal candidate response may include a combination of text content, audio content, visual content, 3D contents, etc.

[0070] In an exemplary embodiment, a multi-modal candidate prompt of the multi-modal candidate dataset 120 may include a combination of a picture of an environment and an audio exclamation from a car user. The corresponding multi-modal candidate response in the multi-modal candidate dataset 120 may include a list of potential stopover locations with similar scenery and the relevant navigation information.

[0071] Referring to FIG. 4, in various embodiments, a first feature extractor 222 of the first machine learning arm 220 may include a plurality of first multi-modal feature extractors 2222/2224/2226. Each of the plurality of first multi-modal feature extractors 2222/2224/2226 may be configured to extract features from different modes of input. In an exemplary embodiment, a text input 1252 from the candidate dataset 120 may be provided to the first multi-modal feature extractor 2222 of the first machine learning arm 220, an audio input 1254 may be provided to the first multi-modal feature extractor 2224 of the first machine learning arm 220, and a visual input 1256 may be provided to the first multi-modal feature extractor 2226 of the first machine learning arm 220. A combined first feature set 223 may be collected or compiled with the respective outputs from the first feature extractor 2222, the second feature

extractor 2224, and the third feature extractor 2226. The first feature set 223 may be input to the first machine learning model 224 to obtain a first output 225.

[0072] Similar description may apply to the second machine learning arm 240, wherein a second feature extractor 242 of the second machine learning arm 240 may include a plurality of second multi-modal feature extractors 2422/2424/2426. Each of the plurality of second multi-modal feature extractors 2422/2424/2426 may be configured to extract features from different modes of input. In an exemplary embodiment, the text input 1252 may be provided to the second multi-modal feature extractor 2422 of the second machine learning arm 240, the audio input 1254 may be provided to the second multi-modal feature extractor 2424 of the second machine learning arm 240, and the visual input 1256 may be provided to the second multi-modal feature extractor 2426 of the second machine learning arm 240. A combined second feature set 243 may be collected or compiled with the respective outputs from the first feature extractor 2422, the second feature extractor 2424, and the third feature extractor 2426. The first feature set 243 may be input to the second machine learning model 244 to obtain a second output 245.

[0073] As non-limiting examples of the various feature extractors, a Sbert-type sentence transformer model may be used for a text input; a AudioSet model may be used for an audio input; a CLIP model (Contrastive Language-Image Pre-Training model) which is a text image comparison pre-trained model, may be used for a visual or pictorial input; a VideoCLIP model (Video Contrastive Language-Image Pre-Training model) which is a text vs. video comparison pre-trained model, may be used for video input; a CLIP-FO3D model (Contrastive Language-Image Pre-Training model for free and open-world 3D scene representations) which fuses text, images, videos and 3D features, may be used for 3D model/representation input. The features or embeddings extracted by the various feature extractors may be combined as a combined feature set for input to the machine learning model or deep learning model.

[0074] FIG. 5 illustrate a method of training the first machine learning arm 220 and the second machine learning arm 240 based on multi-modal datasets. Departing from the above single mode training dataset, the first training dataset 130 may include a plurality of multi-modal training prompt-response pairs each labelled with a respective first target outcome 134. The second training dataset 132 may also include a plurality of multi-modal training prompt-response pairs each labelled with a respective second target outcome 136. The first training dataset 130 and the respective first target outcome 134 may be used to train the first machine learning model 224. Similarly, the second training dataset 132 and the respective second target outcome 136 may be used to train the second machine learning model 244.

[0075] In various embodiments, to prevent the model from overfitting, the default L2 regularization intensity coefficient may be used during training. In exemplary embodiments, the deep learning model achieved good results in accuracy, recall, F1 and other indicators based on the actual data verification.

[0076] In various embodiments, each of the plurality of multi-modal training prompt-response pairs may include a multi-modal training prompt and a multi-modal training response. In various embodiments, each of the multi-modal training prompt and the multi-modal training response may include a combination of text content, audio content, visual content, 3D contents, etc.

[0077] FIG. 6 illustrates a system 200 according to various embodiments of the disclosure. In contrast to earlier embodiments, the system 200 may include a common feature extractor 232 connected to a first machine learning model 224; a second machine learning model 244; and a comparison model 320. In various embodiments, a candidate dataset 120 may be provided as input 125 to the common feature extractor 232. The common feature extractor 232 may extract the different features or embeddings from the candidate dataset 120 to obtain a common feature set 235. The system 200 may provide the common feature set 235 to each of the first machine

learning model 224, and the second machine learning model 242. Similar to the above embodiments, the first machine learning model 224 may provide a first output 225 based on the common feature set 235. Similarly, the second machine learning model 244 may provide a second output 245 based on the common feature set 235. The comparison model 320 may compute or determine a similarity score 325 based on a selected reference prompt PR_i from the response knowledge base 400 and the candidate dataset 120. The first output 225, the second output 245 and the similarity score 325 may be provided as a weighted sum to determine a weighted outcome 420. Based on the weighted outcome, a leading candidate prompt-response pair 422 may then be incorporated into the response knowledge base 400.

[0078] Still referring to FIG. 6, in various embodiments, the common feature extractor 232 may be a common multi-modal feature extractor 232. As such, a multi-modal candidate dataset 120 may act as an input 125 to the common multi-modal feature extractor 232 for feature extraction.

[0079] FIG. 7 illustrate a method 700 of constructing a response knowledge base according to various embodiments of the present disclosure. The method 700 may comprise: in frame 710, providing a candidate dataset to a first machine learning arm, the first machine learning arm trained based on a first target outcome, wherein the candidate dataset comprises a plurality of candidate prompt-response pairs; in frame 720, providing the candidate dataset to a second machine learning arm, the second machine learning arm trained based on a second target outcome; in frame 730, based on a first output from the first machine learning arm and a second output from the second machine learning arm, determining a leading candidate prompt-response pair from the plurality of candidate prompt-response pairs; and in frame 740, incorporating the leading candidate prompt-response pair into the response knowledge base.

[0080] In various embodiments, the method 700 may further comprise: in frame 750, computing a respective similarity score between each of the plurality of candidate prompt-

response pairs and a selected reference prompt from the response knowledge base; and in frame 760, based on the respective similarity score, determining the leading candidate prompt-response pair from the plurality of candidate prompt-response pairs.

[0081] In various embodiments, the method 700 may further comprise: in frame 770, determining the leading candidate prompt-response pair based on a weighted sum of the first output, the second output and the similarity score.

[0082] Exemplary embodiment

[0083] In an exemplary embodiment of the system 200 which comprises a first machine learning model and a second machine learning model, the first target outcome may be sales completion, and hence, the first output corresponds to a probability of sales completion. The first machine learning model was trained using a first training dataset with “1” corresponding to “successful sales” or “0” corresponding to “unsuccessful sales”.

[0084] In addition, the second target outcome may be increase in call duration, and hence, the second output corresponds to a probability of an increase in call duration. The second machine learning model was trained using a second training dataset with “1” corresponding to “increase in call duration by 30 seconds or more” or “0” corresponding to “increase in call duration below 30 seconds”. Examples of the candidate prompts, the candidate responses, the first output, and the second output are shown in Table 2.

[0085] In the process of constructing the response knowledge base, it was found that responses which are concise and directed at the customer's concern(s) have a higher probability success, i.e. success of closing a sale.

Table 2. Exemplary candidate dataset with computed first output and second output

Candidate Prompt	Candidate Response	First output	Second output

The interest rate is too high	I understand your concerns. Interest is indeed an important factor in choosing financial products. However, our interest setting is based on risk assessment and market standards. Although the interest rate is relatively high, our products provide greater flexibility and more convenient services. If you have other needs, we can recommend more suitable products for you.	0.73	0.65
The interest rate is too high	Thank you for your feedback. Interest is part of our product pricing, which reflects market interest rates and risk costs. However, we can strive for more favorable interest rates for you based on your credit status and cooperation history. Would you like to provide some additional information? So that we can better evaluate.	0.81	0.82
The interest rate is too high	I understand your concerns about interest. Our products do have certain interest costs, but compared with other products on the market, our services are more comprehensive, including fast approval, flexible repayment methods, etc. If you feel that the interest rate is too high, we can introduce you to some low-interest but slightly more stringent alternatives.	0.65	0.72
The interest rate is too high	I fully understand your concerns about interest. Our interest rates are set based on market conditions and risk assessment, but we also hope that customers can get benefits. At present, we have some preferential activities, such as early repayment interest reduction, etc. You can consider these offers, or tell me your specific needs, and I will do my best to help you.	0.85	0.85
The interest rate is too high	Thank you for your feedback. Interest is part of our product pricing, which ensures that we can provide high-quality services and guarantees. However, we are also constantly optimizing our products to meet the needs of our customers. If you have other suggestions or needs, please feel free to tell us, and we will do our best to improve.	0.64	0.71

[0086] The proposed response knowledge base and the proposed system for constructing the response knowledge base may be implemented by a processor system 900 as illustrated in the schematic block diagram of FIG. 8. Components of the processing system 900 may be provided within one or more computing device to carry out the functions of the modules or any other modules. One skilled in the art will recognize that the exact configuration or arrangement illustrated in FIG. 8 is provided by way of example only, e.g., each processing system provided may be different and the exact configuration of processing system 900 may vary.

[0087] In embodiments of the present disclosure, the processing system 900 may include a controller 901 and user interface 902. User interface 902 is configured to enable manual interactions between a user and the computing module as required. For this purpose, the processing system 900 includes the input/output components required for the user to enter instructions to provide updates to each of the modules. A person skilled in the art will recognize that components of user interface 902 may vary from embodiment to embodiment but may typically include one or more input devices 935 such as but not limited to a touchscreen, a keyboard, a joystick, a mouse, a microphone, etc. The user interface 902 also includes EEG sensors 933 that can be attached to the user's head to sense the user's brain activity. The user interface 902 can also include a media player 940, which can be in the form of one or more playback devices, including but not limited to a display, a speaker, earphones, headsets, etc.

[0088] The controller 901 is configured to be in data communication with the user interface 902 via bus 915. The controller 901 includes memory 920 and processor 905 mounted on a circuit board to process instructions and data, e.g., to perform the method of the present disclosure. The controller 901 includes an operating system 906, an input/output (I/O) interface 930 for communicating with user interface 902, and a communications interface, e.g., a network card 950. The network card 950 may, for example, be configured to send data from the

controller 901 via a wired or wireless network to other processing devices or to receive data via the wired or wireless network. Wireless networks that may be utilized by the network card 950 include, but are not limited to, Wireless-Fidelity (Wi-Fi), Bluetooth, Near Field Communication (NFC), cellular networks, satellite networks, telecommunication networks, Wide Area Networks (WAN), and etc.

[0089] Memory 920 and operating system 906 are in data communication with central processing unit (CPU) 905 via bus 910. The memory 920 may include both volatile and non-volatile memory. The memory 920 may include more than one of each type of memory, e.g., Random Access Memory (RAM) 923, Read Only Memory (ROM) 925, and a mass storage device 945. The mass storage device 945 may include one or more solid-state drives (SSDs). One skilled in the art will recognize that the memory described above includes non-transitory computer-readable media and shall be taken to include all computer-readable media except for a transitory, propagating signal. Typically, instructions are stored as program code in the memory but can also be hardwired. Memory 920 may include a kernel and/or programming modules such as a software application that may be stored in either volatile or non-volatile memory.

[0090] Herein, the term “processor” is used to refer generically to any device or component that can process computer-readable instructions, including for example, a microprocessor, microcontroller, programmable logic device, or other computational device. That is, processor 905 may be provided by any suitable logic circuitry for receiving inputs, processing them in accordance with instructions stored in memory, and generating outputs (for example to the memory components or media player 940). In the present disclosure, processor 905 may be a single core or multi-core processor with memory addressable space. In one example, processor 905 may be multi-core, comprising—for example—an 8 core CPU. In another example, it could be a cluster of CPU cores operating in parallel to accelerate computations.

[0091] Further, one skilled in the art will recognize that certain functional units in this description have been labelled as modules throughout the specification. The person skilled in the art will also recognize that a module may be implemented as circuits, logic chips or any sort of discrete component. Still further, one skilled in the art will also recognize that a module may be implemented in software which may then be executed by a variety of processor architectures. In embodiments of the disclosure, a module may also comprise computer instructions or executable code that may instruct a computer processor to carry out a sequence of events based on instructions received. In further embodiments, the module may comprise a combination of different types of modules or sub-modules. The choice of the implementation of the modules may be determined by a person skilled in the art and does not limit the scope of the claimed subject matter in any way.

[0092] All examples described herein, whether of apparatus, methods, materials, or products, are presented for the purpose of illustration and to aid understanding, and are not intended to be limiting or exhaustive. Modifications may be made by one of ordinary skill in the art without departing from the scope of the invention as claimed.

CLAIMS

1. A system, comprising:
memory storing instructions; and
a processor coupled to the memory and configured to process the stored instructions to implement:
a module configured to perform a method of constructing a response knowledge base, the method including:
providing a candidate dataset to a first machine learning arm, the first machine learning arm trained based on a first target outcome, wherein the candidate dataset comprises a plurality of candidate prompt-response pairs;
providing the candidate dataset to a second machine learning arm, the second machine learning arm trained based on a second target outcome; and
based on a first output from the first machine learning arm and a second output from the second machine learning arm, determining a leading candidate prompt-response pair from the plurality of candidate prompt-response pairs; and
incorporating the leading candidate prompt-response pair into the response knowledge base.
2. The system as recited in claim 1, wherein the first output comprises a plurality of first probabilities, each of the plurality of first probabilities is a probability of achieving the first target outcome for a respective one of the plurality of candidate prompt-response pairs, and wherein the second output comprises to a plurality of second probabilities, each of the plurality of second probabilities is a probability of achieving the second target outcome for a respective one of the plurality of candidate prompt-response pairs.

3. The system as recited in claim 1, wherein the method further comprising:
computing a respective similarity score between each of the plurality of candidate prompt-response pairs and a selected reference prompt from the response knowledge base; and
based on the respective similarity score, determining the leading candidate prompt-response pair from the plurality of candidate prompt-response pairs.
4. The system as recited in claim 3, wherein the method further comprising: determining the leading candidate prompt-response pair based on a weighted sum of the first output, the second output and the similarity score.
5. The system as recited in claim 3, wherein computing the respective similarity score comprises: providing each of the plurality of candidate prompt-response pairs and the selected reference prompt to a comparison model, wherein a similarity output of the comparison model corresponds to the respective similarity score.
6. The system as recited in claim 1, wherein
providing the candidate dataset to the first machine learning arm comprises: providing the candidate dataset to a first feature extractor to obtain a first feature set, and providing the first feature set to a first machine learning model; and
providing the candidate dataset to the second machine learning arm comprises: providing the candidate dataset to a second feature extractor to obtain a second feature set, and providing the second feature set to a second machine learning model.

7. The system as recited in claim 6, wherein the first feature extractor comprises a plurality of first multi-modal feature extractors, and the second feature extractor comprises a plurality of second multi-modal feature extractors

8. The system as recited in claim 6, wherein each of the first machine learning model and the second machine learning model is a respective deep learning model.

9. The system as recited in claim 1, wherein the response knowledge base comprises a plurality of reference prompts and a plurality of reference responses, each of the plurality of reference responses corresponding to one of the plurality of reference prompts.

10. The system as recited in claim 9, wherein multiple ones of the plurality of reference responses correspond to one of the plurality of reference prompts.

11. The system as recited in claim 9, wherein incorporating the leading candidate prompt-response pair into the response knowledge base comprises: replacing one of the plurality of reference responses with the a respective candidate response of the leading candidate prompt-response pair.

12. The system as recited in claim 1, wherein the candidate dataset is a multi-modal dataset.

13. The system as recited in claim 12, wherein the multi-modal dataset comprises a plurality of multi-modal prompts and a plurality of multi-modal responses.

14. The system as recited in claim 1, further comprising: providing the candidate dataset to a common feature extractor to obtain a common feature set, and providing the common feature set to the first machine learning arm and the second machine learning arm.

15. The system as recited in claim 1, further comprising: training the first machine learning arm based on a first training dataset, the first training dataset labelled with a respective first target outcome; and training the second machine learning arm based on a second training dataset, the second training dataset labelled with a respective second target outcome.

16. The system as recited in claim 1, wherein each of the first target outcome and the second target outcome is quantifiable by a binary value.

17. A method to constructing a response knowledge base, comprising:

providing a candidate dataset to a first machine learning arm, the first machine learning arm trained based on a first target outcome, wherein the candidate dataset comprises a plurality of candidate prompt-response pairs;

providing the candidate dataset to a second machine learning arm, the second machine learning arm trained based on a second target outcome;

based on a first output from the first machine learning arm and a second output from the second machine learning arm, determining a leading candidate prompt-response pair from the plurality of candidate prompt-response pairs; and

incorporating the leading candidate prompt-response pair into the response knowledge base.

18. The method as recited in claim 17, further comprising:

computing a respective similarity score between each of the plurality of candidate prompt-response pairs and a selected reference prompt from the response knowledge base; and based on the respective similarity score, determining the leading candidate prompt-response pair from the plurality of candidate prompt-response pairs.

19. The method as recited in claim 18, further comprising: determining the leading candidate prompt-response pair based on a weighted sum of the first output, the second output and the similarity score.

20. A response knowledge base constructed based on the method as recited in claim 1, the response knowledge base comprising a plurality of reference prompts and a plurality of reference responses, wherein multiple ones of the plurality of reference responses corresponds to one of the plurality of reference prompts.