CLAIMS

Claim 1: A method for performing intent detection and classification, the method comprising:

receiving an input text;

reformulating the input text as one or more prompts;

passing the one or more prompts through a language model to obtain one or more states based on the input text, wherein the language model comprises a classifier head and a language modeling head configured to operate in parallel with the classifier head, and wherein the language model is jointly trained with a classification task and a language modeling task, and using at least two outlier samples, wherein one of the at least two outlier samples is generated by a first technique and another of the at least two outlier samples is generated by a second technique;

based on the one or more states, predicting a classifier output by the classifier head, and simultaneously generating a language modeling prediction by the language modeling head; and

combining the classifier output and the language modeling prediction to generate an output prediction, wherein the output prediction is out-of-distribution if the language modeling prediction is indicative that the one or more states are predicted to be outside an in-domain label space, or otherwise, the output prediction is an in-distribution intent, wherein the classifier output comprises the in-distribution intent falling within the in-domain label space,

wherein the first technique comprises performing a convex combination of a first embedding (h_x) associated with a first training input (x) and a second embedding ($h_{x'}$) associated with a second training input (x'), the first training input and the second training input being from different classes, and

wherein the second technique comprises:

computing a mean value (μ) and a standard deviation (σ) of all sentence lengths in a training set of the in-domain label space;

removing stopwords from the training set to construct a set of words (W); and

uniformly sampling from a uniform distribution of the set of words (*W)* to form the outlier sample comprising a generic sample represented by:

$$x_i \sim U(W) \text{ for } 0 < i \leq l$$

where $l \sim N(\mu, \sigma^2)$.

Claim 2: The method as claimed in claim 1, wherein

each of the one or more prompts comprises an input slot and an answer slot,

reformulating the input text as the one or more prompts comprises filling or replacing the input slot with the input text, and

the method further comprises:

mapping the output prediction to a set of tokens;

determining a token with a highest average probability or a majority voting from the set of tokens; and

filling the answer slot with the determined token.

Claim 3: The method as claimed in claim 2, wherein the one or more prompts comprise at least one of the following:

a first comment about the input text and a first question concerning the token;

a second comment about the input text and a second question concerning the token, the second comment being different from the first comment and the second question being different from the first question;

a third comment about the input text and a topic concerning the token;

a first query about the input text and a user enquiry about the token; or

a second query about the input text and an enquiry about the token, the first query being different from the second query and the user enquiry being different from the enquiry.

Claim 4: The method as claimed in any one of claims 1 to 3, wherein reformulating the input text as one or more prompts further comprises pre-processing the input text.

Claim 5: The method as claimed in claim 4, wherein pre-processing the input text comprises removing a non-alphanumeric character and/or a stopword originating from the input text.

Claim 6: The method as claimed in any one of claims 1 to 5, wherein the language model further comprises a pre-trained model in upstream communication with the classifier head and the language modeling head, and the one or more states comprise one or more hidden states generated by the pre-trained module.

Claim 7: The method as claimed in claim 6, wherein the pre-trained model comprises a Bidirectional Encoder Representation from Transformer (BERT), or a Generative pre-trained transformer (GPT), or a Pathways Language Model (PaLM), or a Large Language Model Meta AI (Llama), or Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA), or Generalized Autoregressive Pretraining for Language Understanding (XLNet), a Robustly Optimized BERT Pretraining Approach (RoBERTa), or a Decoding-enhanced BERT with Disentangled Attention (DeBERTa).

Claim 8: The method as claimed in any one of claims 1 to 7, wherein an embedding (h) of the outlier sample generated by the first technique is represented by:

$$h = \theta h_x + (1-\theta) h_{x'}$$

where a distribution of θ is a uniform distribution given by $\theta \sim U(0, 1)$.

Claim 9: The method as claimed in any one of claims 1 to 8, wherein the classification task comprises:

applying, by the classifier head, a softmax function on a plurality of hidden states to obtain classifier logits, wherein the plurality of hidden states is generated by passing training prompts reformulated from a training input text of a specific class through the language model; and

training the classifier head with cross entropy loss based on the classifier logits.

Claim 10: The method as claimed in claim 9, wherein the language modeling task comprises:

applying, by the language modeling head, a conditional softmax function on the plurality of hidden states to obtain language modeling logits; and

training the language modeling head with binary cross entropy loss based on the language modeling logits.

Claim 11: The method as claimed in any one of claims 1 to 10, wherein the at least two outlier samples are trained with binary cross entropy loss.

Claim 12: The method as claimed in claim 11, wherein the language model in entirety, upon training, has a weighted loss function comprising a summation of the cross entropy loss based on the classifier logits, a weighted binary cross entropy loss based on the language modeling logits, and a weighted binary cross entropy loss associated with the at least two outlier samples.

Claim 13: The method as claimed in any one of claims 1 to 12, wherein the language model is post-trained with an adaptive decision boundary.

Claim 14: A system for performing intent detection and classification using a language model comprising a classifier head and a language modeling head configured to operate in parallel with the classifier head, the system comprising:

a memory; and

at least one processor unit in communication with the memory, the at least one processor unit configured to:

receive an input text,

reformulate the input text as one or more prompts,

pass the one or more prompts through the language model to obtain one or more states based on the input text, wherein the language model is configured to be jointly trained with a classification task and a language modeling task, and using at least two outlier samples, wherein one of the at least two outlier samples is to be

generated by a first technique and another of the at least two outlier samples is to be generated by a second technique, and

combine a classifier output and a language modeling prediction to generate an output prediction,

wherein based on the one or more states, the classifier head is configured to predict the classifier output, and simultaneously, the language modeling head is configured to generate the language modeling prediction; and

wherein the output prediction is out-of-distribution if the language modeling prediction is indicative that the one or more states are predicted to be outside an in-domain label space, or otherwise, the output prediction is an in-distribution intent, wherein the classifier output comprises the in-distribution intent falling within the in-domain label space;

wherein the first technique is configured to perform a convex combination of a first embedding (h_x) associated with a first training input (x) and a second embedding ($h_{x'}$) associated with a second training input (x'), the first training input and the second training input being from different classes, and

wherein the second technique is configured to:

compute a mean value (μ) and a standard deviation (σ) of all sentence lengths in a training set of the in-domain label space;

remove stopwords from the training set to construct a set of words (W); and

uniformly sample from a uniform distribution of the set of words (W) to form the outlier sample comprising a generic sample represented by:

$$x_i \sim U(W) \text{ for } 0 < i \leq l$$

where $l \sim N(\mu, \sigma^2)$.

Claim 15: The system as claimed in claim 14,

wherein each of the one or more prompts comprises:

an input slot for filling the input text; and

an answer slot,

wherein the system further comprises a verbalizer configured to map the output prediction to a set of tokens, and

wherein the at least one processor unit is further configured to determine a token with a highest average probability or a majority voting from the set of tokens and fill the answer slot with the determined token.

Claim 16: The system as claimed in claim 14 or 15, wherein the at least one processor unit is further configured to remove a non-alphanumeric character and/or a stopword originating from the input text.

Claim 17: The system as claimed in any one of claims 14 to 16, wherein the language model further comprises a pre-trained model in upstream communication with the classifier head and the language modeling head, the pre-trained model being configured to generate the one or more states.

Claim 18: The system as claimed in claim 17, wherein the pre-trained model comprises a Bidirectional Encoder Representation from Transformer (BERT), or a Generative pre-trained transformer (GPT), or a Pathways Language Model (PaLM), or a Large Language Model Meta AI (Llama), or Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA), or Generalized Autoregressive Pretraining for Language Understanding (XLNet), a Robustly Optimized BERT Pretraining Approach (RoBERTa), or a Decoding-enhanced BERT with Disentangled Attention (DeBERTa).

Claim 19: The system as claimed in any one of claims 14 to 18, wherein during training with the classification task, the classifier head is configured to apply a softmax function on a plurality of hidden states to obtain classifier logits, and configured to be trained with cross entropy loss based on the classifier logits, wherein the plurality of hidden states is generated by passing training prompts reformulated from a training input text of a specific class through the language model.

Claim 20: The system as claimed in claim 19, wherein during training with the language modeling task, the language modeling head is configured to apply a conditional softmax function on the plurality of hidden states to obtain language modeling logits, and

configured to be trained with binary cross entropy loss based on the language modeling logits.

Claim 21: The system as claimed in any one of claims 14 to 20, wherein the at least two outlier samples are configured to be trained with binary cross entropy loss.

Claim 22: The system as claimed in claim 21, wherein the language model in entirety, upon training, has a weighted loss function comprising a summation of the cross entropy loss based on the classifier logits, a weighted binary cross entropy loss based on the language modeling logits, and a weighted binary cross entropy loss associated with the at least two outlier samples.

Claim 23: The system as claimed in any one of claims 14 to 22, wherein the language model is configured to be post-trained with an adaptive decision boundary.

Claim 24: A method for training a language model for intent detection and classification, the method comprising:

jointly training the language model with a classification task and a language modeling task; and

using at least two outlier samples, wherein one of the at least two outlier samples is generated by a first technique and another of the at least two outlier samples is generated by a second technique different from the first technique,

wherein the first technique comprises performing a convex combination of a first embedding (h_x) associated with a first training input (x) and a second embedding ($h_{x'}$) associated with a second training input (x'), the first training input and the second training input being from different classes, and

wherein the second technique comprises:

computing a mean value (μ) and a standard deviation (σ) of all sentence lengths in a training set of an in-domain label space;

removing stopwords from the training set to construct a set of words (W); and

uniformly sampling from a uniform distribution of the set of words (*W)* to form the outlier sample comprising a generic sample represented by:

$$x_i \sim U(W) \text{ for } 0 < i \leq l$$

where $l \sim N(\mu, \sigma^2)$.

Claim 25: The method as claimed in claim 24, wherein an embedding (*h*) of the outlier sample generated by the first technique is represented by:

$$h = \theta\, h_x + (1\text{-}\theta)\, h_{x'}$$

where a distribution of θ is a uniform distribution given by $\theta \sim U(0, 1)$.

Claim 26: The method as claimed in claim 24 or 25, wherein the classification task comprises:

applying, by a classifier head of the language model, a softmax function on a plurality of hidden states to obtain classifier logits, wherein the plurality of hidden states is generated by passing training prompts reformulated from a training input text of a specific class through the language model; and

training the classifier head with cross entropy loss based on the classifier logits.

Claim 27: The method as claimed in claim 26, wherein the language modeling task comprises:

applying, by a language modeling head of the language model, a conditional softmax function on the plurality of hidden states to obtain language modeling logits; and

training the language modeling head with binary cross entropy loss based on the language modeling logits, the language modeling head being configured to operate in parallel with the classifier head.

Claim 28: The method as claimed in any one of claims 24 to 27, wherein using the at least two outlier samples comprising training the at least two outlier samples with binary cross entropy loss.

Claim 29: A non-transitory computer-readable storage medium comprising computer readable instructions operable when executed by a computer to perform intent detection and classification, the computer readable instructions configured to perform a method as claimed in any one of claims 1 to 13 and/or a method as claimed in any one of claims 24 to 28.