AAAI 2024 Workshop on Reinforcement Learning and Large Language Model に採択されました

**ドメインチャットボット用のRAG最適化のための強化学習**

**Mandar Kulkarni、Praveen Tangarajan、Kyung Kim、Anusua Trivedi**

フリップカートのデータサイエンス

シアトル, ワシントン, アメリカ

|  |  |  |
| --- | --- | --- |
| arXiv:2401.06800v1 [cs.CL] 2024年1月10日 | **アブストラクト** | 文脈的な質問回答タスクで良いことになり、 |
| 大きい言語モデル(LLM)の出現を使って、conversa- | 関連するテキストをコンテキストとして指定すると、LM は生成できます。 |
| その情報を使用して質問に答えます。 リトリバル Aug- |
| ドメインユースケースでは、条件付きアシスタントが流行しています。 |
| メンテッドジェネレーション(RAG)は、使用される主要な技術の一つです |
| LLMは、文脈的な質問に答える能力を得る |
| ドメインデータに関する質問に答えるためのチャットボットの構築 |
| 広範囲にわたるトレーニング、およびRetrieval Augmented Generaによる |
| tion (RAG) は、ボットがドメイン固有の回答を可能にする | RAG は 2 つのコンポーネントで構成されます。: retrieval モデルと |
| お問い合わせ この論文では、RAGベースのアプローチについて説明しています。 | LLMをベースとした答え生成モデル。 リトリーバル |
| Fre-を使用してユーザーの問い合わせに答えるチャットボットの構築 | ユーザーのクエリに関連したコンテキストをモデル化します。 クエリ |
| よくある質問(FAQ) 社内研修 | 取得されたコンテキストは、LM に LLM に入力されます。 |
| infoNCE の損失、および exper を使用してモデルを埋め込みます。 | 回答を生成する適切なプロンプト。 お問い合わせ |
| 社内モデルが機能するimental結果が実証される |
| RAG ベースのチャットボットには、いくつかのチャルがあります。 |
| よく知られた汎用パブよりも大幅に優れています- |
| 実用的なスタンドポイントからのレンゲ。 |
| lic embedding モデル、両方の条件の検索精度と |
| アウトオブドメイン(OOD)クエリ検出。 LLM として、LLM を使用します。 | 1。 retrieval モデルが関連したコンテキストを取得できなかった場合、gen- |
| オープンAPIベースの有料チャットGPT (gpt-35-turbo-16k-0613) | erated 回答は不正確で非公式です。 ザ・オブ・ザ・ |
| モデル。 以前に取得されたコンテキストが、 | retrieval は、英語以外の問い合わせに対してより困難です。 |
| 特定のパターン/シーケンスに対する回答を生成するために使用される |
| 例:コードミックスヒングリ語 |
| 問い合わせ(例、フォローアップクエリ) したがって、スコープがあります |
| 2. 有料 API ベースの LLM (例、ChatGPT) の場合、1 件あたりのコスト |
| LLMトークン数とコストを最適化します。 想定される |
| 呼び出しは入力とアウト数に基づいて計算されます。 |
| 固定検索モデルとLMは、数を最適化します。 |
| LLMトークンは、Reinforcement Learning (RL) を使用します。 スペシフィ- | トークンを入れます。 コンテキストで渡されるトークンの数が多い |
| 外部からRAGにポリシーベースモデルをご提案します。 | API 呼び出しあたりのコストが高くなります。 高い容積を使って |
| 方針行動によるRAGパイプラインと相互作用する | ユーザーの問い合わせの費用は重要になることができます。 |
| コストを最適化するためのポリシーを更新します。 ポリシーモデル |
| 3。 RAGでマルチターン会話を有効にするには、コンバー- |
| 二つのアクションを実行できます。FAQ コンテキストを取得したり、再実行をスキップしたりできます。 |
| セーションの歴史は維持され、渡される必要があります |
| トライバル。 Open API ベースの GPT-4 を評価として使用しています。 |
| あらゆるクエリでLM。 より大きな入力が大きいことが知られています。 |
| ボットの答えの質を評価するモデル。 利用する |
| ケンサイズは正確さまたは幻覚の低下にとして導きます |
| 適切な報酬シェーピング、GPT-4 の評価は、 |
| 数値報酬。 それからポルを使ってポリシーモデルを訓練して下さい | LLMには、関連するインフォアを選択するための追加のタスクがあります。 |
| 複数のトレーニングチャットセッションで icy グラデーション. ポルとして- | 大きい文脈からの mation (Liu ら 2023a). |
| icyモデル、パブリックgpt-2モデル、 |
| この論文では、まずRAGベースのアプローチを記述します。 |
| 社内のBERTモデル。 提案された RL ベースの op を使って |
| Fre-を使用してユーザーの問い合わせに答えるチャットボットの構築 |
| 類似性しきい値と組み合わせて、我々はできる |
| よくある質問(FAQ) ドメインがある |
| 取得中に重要な(〜31%)コスト節約を達成する |
| 少し改善された正確さ。 結果の実演 | クレジットに関する72のFAQからなるFAQデータセット |
| FAQチャットボットでは、提案された RL アプローチはジェネリック | カード申請プロセス。 FAQ のデータセットを an- |
| 既存の RAG パイプラインで実験できます。 | 一般カード情報に関するユーザーの問い合わせ先- |
| **導入事例** | およびポスト・アプリケーション・クエリ。 社内のリトリーバルを訓練 |
| (埋め込み) info ノイズコントラストエストマを用いたモデル |
| tion (infoNCE) 損失 (van den Oord, Li, Vinyals 2019) |
| 大きい言語モデル(LLM)の出現によって、私達は従います |
| 英語とヒングリ語のパラフレーズのクエリで作成 |
| 会話アシスタントをさらに高めた使用に役立ちます |
| ChatGPT を使う 埋め込むモデルはmaxiに訓練されます- |
| ドメインユースケース。 大規模な web スケール テキストの cor で訓練 |
| query-question と query-QnA の類似性を mize。 エクスペリメン |
| 指示調整やReinなどのアプローチでpus |
| 社内モデルがSignifiを実行すると、タル結果が表示されます - - - |
| 人間のフィードバック(RLHF)、LMs と学ぶ力学 |
| 公の事前訓練された埋め込むモデル再よりcantlyよりよい |
|  | 著作権©2024、人工の進歩のための協会 | ガーディング検索精度とアウトオブドメイン(OOD)クエリ |
| インテリジェンス(www.aaai.org)。 お問い合わせ | 検出。 OODの質問は、do-に関連しない質問です |

主要なデータ、例えば、今日の天気は? すべてのユーザーのために クエリは、コンとしてトップkのFAQ(質問+回答)を取得しています。テキストを LLM に入力して回答を生成します。 お問い合わせ 前の2つのクエリ、回答、FAQのcon-を維持複数ターンの会話を有効にするための歴史としてのテキスト。 お問い合わせ 有料APIベースチャットGPT(gpt-35-turbo-16k-0613) LLM として実験されるすべての実験のため。 GPT-4 は、 ボットはエンドツーエンドのRAGのための質(よい/悪い)に答えます パイプライン。

RLベースのアプローチを提案し、num-num-を最適化LLMに渡されたトークンのバー。 弊社では、cer-についてクエリのパターン/シーケンス、私たちは良い答えを得ることができます ボットからFAQのコンテキストをフェッチすることなく。 例-このようなシナリオのアンプルは、次のクエリのために1.することができます。 既に存在していた場合、FAQ コンテキストは取得できません。 前回のクエリで取得; 2. クエリのシーケンス 同じFAQを参照すると、コンテキストはのみ取得できます 開始時に3. OODのクエリでは、LMはそれをプロンプト -ボットを誘導して回答を生成できます。 このインを使用して-視力では、ポリシーのグラデーションに基づくアプローチをoptiに提案しています。LLMトークンの数をマイズし、したがって、コスト。 ザ・オブ・ザ・ ポリシーモデルへの入力はステートです。 以前のクエリ、以前のポリシーアクション、および現在の お問い合わせ ポリシーモデルは、 [FETCH] または [NO\_FETCH] は、 通常のRAGパイプラインが実行される ポリシーネットワークが [FETCH] アクションを取るとき。 いつか ポリシーネットワークは [NO\_FETCH] アクション、FAQ con を選択します。テキストは取得されません。 クエリとコンテキスト(ケースで空) LLM に [NO\_FETCH] アクションが入力されます。 お問い合わせ GPT-4 は報酬モデルとして、品質評価を変換します (Good/Bad) 適切な報酬を使用して数値報酬に シェーピング。 LLMが良い答えを生成する場合(評価されるように) GPT-4) コンテキストをフェッチすることなくも、高い 将来のそのような行動を促進するための肯定的な報酬。 なし FAQ 文脈を fetching すると、間違った答えにつながります。ネガティブな報酬(例えば、このポリシーがポリシーの場合に起こりうる) ドメインクエリの [NO\_FETCH] アクションを選択します。 過去の関連コンテキストなし。 トレーニングチャットの場合 各州のセッションでは、ポリシーモデルがアクションをサンプル ac-上の現在の確率分布によるとお問い合わせ それから生成します(状態、行為、報酬)軌跡 現在のポリシーから複数回サンプリングすることで ポルicy モデルが cu- でポリシーグラデーションを使用して更新されます。mulative 報酬。 方針モデルとして、実験した 社内のBERTとパブリックgpt-2モデル。 図1は、 提案された政策に基づくアプローチのアーキテクチャ。 エクスカーターimentalの結果は、ポリシーモデルが提供していることを示しています FAQ コンテキストを返したときにのみ取得することでトークン保存お問い合わせ 単純な類似性しきい値と組み合わせた場合ベース最適化、トークンの節約を実現 テストチャットセッションで 91 件のクエリで achiev-少し改善された正確さ(手動による評価される ラベリング)通常のRAGパイプラインよりも。 これは、 提案されたアプローチの有効性。

**関連作品**   
ジェネレーションAIとLMの最近の進歩により、 リトリバル拡張生成(RAG)(Lewis et al. 2021) アプローチは、contexの好ましい戦略として登場しました-

2

**社内の組み込みモデルのトレーニング**

FAQチャットボットでRAGを使用するには、まず最初に取得する必要があります 与えられたクエリに答えるために最も関連性の高いFAQをトップk。 フェッチへ top-k FAQ は、組み込みモデルがクエリをエンコードするのに使われます よくある質問 質問と回答をまとめました。 よくある質問と使い方は、FAQのベクトルを取得することです。 コサインSIM-ilarity 測定を使用して、FAQ をランク付けします。 関連する。 一般的な汎用性で実験 訓練済みのモデルとモデルを事前に訓練 ドメイン内データ。 社内の組み込みモデルを訓練 infoNCEの損失(Oord、Li、Vinyals 2019)で私たち-よく知られている埋め込みモデル(e5ベースv2)の重量を 初期化。 InfoNCEは、コンの効果的な損失を証明しました自己指示された設定の下でtrastive学習、特に (ゴヤールら、2021年) 公に利用できる大使館を微調整-dingモデル *e5ベースv2* (Wang et al. 2022) インドメインを使用して infoNCE 損失のデータ。

組み込みのトレーニングのためのデータセットを作成します。 ChatGPT および小さい手動タギング。 複数の生成 質問のパラフレーズとしてFAQごとの英語の質問。 お問い合わせ インドでは英語圏の人口が大きいため、 プラットフォーム上の多くのヒングリ語のクエリ。 また、生成する いくつかの ヒンジ語 パラフレーズ 質問 per FAQ to support code-Hinglish クエリをミックスします。 次に、手動で小さなセットを追加します。 答えることができるクエリを追加するコルパス よくある質問 回答の内容をベースとします。 ～3.5k のお問い合わせ トレーニング、検証のための1kクエリ、および1014クエリ テスト。

モデルを訓練するために、次の損失関数を使用します。

|  |  |  |
| --- | --- | --- |
|  | 拡張(exp)シム( )*ログインお問い合わせ, zログイン*)*ツイート*) | (1) (1) (1) |
| *ログインお問い合わせ* ================================================================================================================================================================================================================================================================ | *ログイン*=11ツイート*ログイン* ================================================================================================================================================================================================================================================================*お問い合わせ*. . 拡張(exp)シム( )*ログインお問い合わせ, zログイン*)*ツイート*) |  |

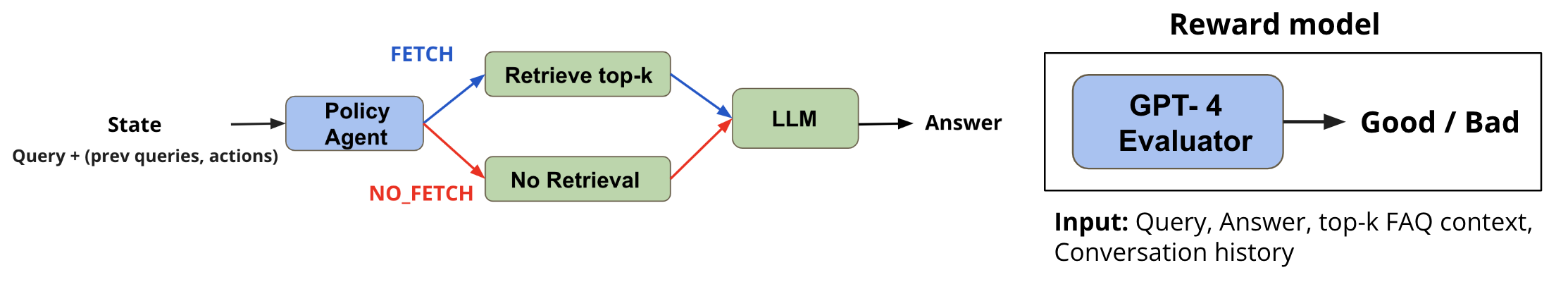
詳しくはこちら *ログインお問い合わせ* そして、 *ログインログイン* 肯定的なペアを示します(すなわち、質問のパラ-同じFAQのフレーズ, *ツイート* バッチ サイズを示します、 *ログイン*温度を示します。 すべての実験のために、私達は置きました *ネクタイ* お問い合わせ 8と8 *ログイン* 0.1 にコサインの類似性を sim に使用する。

Eq.1から見たように、情報NCEの損失は具体的にスーツです-クエリFAQでマッピングされたデータをトレーニングできるので、 肯定的な組を必要とし、残りのサンプルをとして扱います インバッチネガティブ。 埋め込むモデルを微調整 2つの目的:query-QnAの類似性および照会を最大限に活用して下さい-質問の類似性。

表1はトップ1およびトップ3の正確さの比較を示します 一般的な英語とヒンジ語のクエリの結果-目的の公共モデル(e5ベースv2)、微調整されたモデル トリプルロスと情報NCE損失で。 ランキング結果は query-QnA cosine の類似性に基づいて計算される。 それはある場合もあります 社内モデルがinfoNCEで訓練されたことを見る 未訓練モデルよりも大幅に優れています。

次にパブの検出性能を比較します。ドメイン内とドメイン外でのライセンスと社内モデル (OOD) サンプル。 ドメイン内の30個のデータセットを作成しました。 OOD はそれぞれに問い合わせます。 インドメインの問い合わせには英語と英語が含まれています。 OOD クエリには、挨拶、ac- が含まれますが、ヒングリ語のクエリ知識・一般非ドメインに関する質問 肯定的かつ負のクエリのスコアを計算しました 公正で微調整されたモデルで。 表2は示します

3



プロフィール 1:ドメインチャットボットのRAGを最適化するためのポリシーエージェントベースのアーキテクチャを提起しました。

現在のクエリ、現在のFAQのコンテキスト、応答gener-ボットから隔離され、会話履歴(を含む) 前のクエリ、回答、FAQ文)。 プロではない理想または地上の真理応答でGPT-4をバイドします。 代わりに、 生成された an の品質を評価するために GPT-4 をプロンプトします。現在のFAQで利用可能な情報に基づいてスワア コンテキストと会話履歴。 GPT-4 は eval に指示されます。出力を正確に一致するかどうかに基づいてuateFAQ文脈で提供される形成。 追加またはde-入手可能な情報からの逸脱は、事実として宣言されています 失敗は評価され、悪いです。 GPT-4は、 格付けの背後にある良いか悪いと理由として最終的な評価。 OOD のクエリでは、ボットは "do" を提供する そうでない」答えは、GPT-4 は、このような解答を評価するように指示されています。よいようにスワース。 表3および4は照会、ボット応答を示し、 GPT-4評価例

|  |  |
| --- | --- |
| **お問い合わせ** | カードの年間料金があります |
| **ボット応答** | はい、年会費はあります クレジットカード500 ただし、この料金が免除されます 年間支出が上回っている場合オフ INR 3.5ラック。 |
| **gpt4 楕円** | お問い合わせ |
| **gpt4 理由** | 回答は直接アドレスです。クエリと提供 関連する情報 よくある質問 答えは 情報と一致する よくある質問文脈、その旨を示す INRの年間料金があります クレジット カードの500、 すると、その人は、 年間費はINR 3.5を超える ラックス。 |

表3: ドメイン内クエリに対するGPT-4評価

チャットセッションのシーケンスでボット出力を評価する 25件の問い合わせ 英語とヒングリ語で構成されたクワリー ドメインクエリ、OD クエリ(例、天気)、 ご挨拶・確認・お問い合わせ等 GPT-4 エバーサ―これらの回答に対して100%の精度を提供しました。 マニュアル 再確認のために評価の検証を行いましたお問い合わせ よくある質問ボットができることを示す評価 ユーザーの質問に正確に応答します。

3

最初のクエリのみのFAQコンテキスト、およびすべてのサブス-quent クエリは、このコンテキストを利用することができます。 塀、フェッチ すべてのクエリに関するFAQのコンテキストは、そのような下に不要な シナリオ。

・ **OODのクエリ**: : :

LLM は OOD の質問に「do」と答えるために促されます お問い合わせ したがって、FAQのコンテキストを取得しても OOD のクエリでは、LM は無視するように指示されます。 フェンス, コンテキストの取得は OOD クエリでは不要 お問い合わせ

上記のリストから見ることができるように、多くのケースがあります FAQ のコンテキストを取得する必要がない場合 通常の RAG パイプラインと同じ応答、そこにfore は、渡されたトークンを最適化するためのスコープです。 LLMは、精度を取引せずにコストを削減します。 強化学習(RL)をベースにしたap-をご提案いたします。LLM に渡されたトークンの数を最適化するアプローチ 図1は、政策に基づく提案されたアプローチを示しています RAGを最適化するためのエージェント。 具体的には、方針を訓練します GPT-4 評価者から報酬を最大化するためのネットワーク。 ザ・オブ・ザ・ RAGパイプラインの外に存在するポリシーネットワークと、 ポリシーをチューニングすることで最適化します。 わたしたちがそうでないと仮定します。 レトリバルモデルやLMLのグラデーションにアクセスし、 これらのコンポーネントを固定として扱います。

ポリシーネットワークへの入力は、状態です。 以前のクエリ、対応するアクション(to-にエンコード)kens と現在のクエリ。 RAGパイプラインは、 歴史としての最後の2つの会話、ポリシーモデルの 同様に、以前の2つのクエリをコンテキストとして保持します。 ザ・オブ・ザ・ ポリシーネットワークは、現在のクエリに対して2つのアクションを取ることができます。 [FETCH]と[NO\_FETCH]。 名前が示すように、いつ ポリシーネットワークは、[FETCH]アクション、通常のRAGを選択します パイプラインを実行します。 ポリシーネットワークが選択される場合 [NO\_FETCH] アクション、FAQ コンテキストは取得されません。 現在のクエリはLMMに直接入力します。 エルム 回答を生成します。 現在のクエリ、回答、 会話履歴は GPT-4 に入力され、これにより、 ボットは良いか悪いと答えます。

直感的に、ポリシーモデルが間違ったアクションを選択した場合、 その場合、GPT-4 は、例えば、 pol- の場合、icy モデルは、インドメインの [NO\_FETCH] アクションを選択します。 以前のコンテキストなしでクエリ。 内にあることを観察します。この場合、テーブル5はこのような例を示します。 GPT-4評価は数値に変換されます 報酬 (*ログイン*) 表6に示す値。 後ろの直感 このリワード関数は、ポリシーモデルが実行したときに a [NO\_FETCH] アクションと GPT-4 がボット応答を率いる グッドとして、入力クエリがフェッチされていないことを示します コンテキストはまだ良い出力につながります。 アイテムリストに記載されている理由。 したがって、プロにそのような行動を遠隔にし、高い肯定的な報酬を与えます。 もし、 ポリシーモデルが[FETCH]アクションを再生し、小さなposを付与します。正式な報酬 GPT-4 がボット応答を ei の悪い場合-アクションのサイダーは、リモートを解除するための負の報酬を提供します そのような行動は、例えば、ポリシーモデルは負の報酬を得る 必要に応じてコンテキストをフェッチしないとき。 最小限に GPT-4評価の費用は、事例のみを評価します。 ポリシーモデルが[NO\_FETCH]アクションを実行する場所。 お問い合わせ

5月5日

どこまでも *ログイン* です。 *割引率* そして、 *ネクタイ* 数値を示す チャットセッションでの問い合わせ 値を設定 *ログイン* 小さなものへ 値(0.1)は、即時の報酬を世話しているので、例えば、 ob-クエリごとに良い出力を達成します。

6つのチャットセッションでポリシーモデルを訓練します。の 168 クエリ. 追加チャットセッションを作成 チャットセッション内のクエリをランダムにシャッフルすることで。 セッション内の各クエリに対して、 [FETCH] または [NO\_FETCH] 初期ポリシーに基づくアクション ザ・サムpledアクションが実行され、ボットレスポンスが生成されます。 GPT-4 は、ボット応答を評価し、評価を提供します。 数値を再変換するGood または Bad として、ワード。 チャットセッションごとに複数の時間を繰り返す 1733(州、行動、評価)タプルのデータセットを作成します。 方針モデルとして、社内のプリトトレーニング済みBERTを利用しています。 モデル。 社内のBERTモデルは同じarchitecを持っています-12 層とエンコーディングで bert-base-uncased としての ture 次元 768。モデルはマスクされた言語を使用して訓練されます モデリング(MLM)と次の文予測(NSP) 商品説明、ユーザーレビュー、 など。 Softmax のアクティベーションでリニアレイヤーを追加します。 最後のレイヤーの[CLS]トークンの埋め込みで、ステートをマップ 2次元アクション空間に表現。 追加する ポリシーアクションを表す vocab に 2 つの新しいトークン, [FETCH] と [NO\_FETCH] をランダムに初期化 埋め込む。 ポリシーネットワークはポリシーで訓練されています Eqに示す勾配の損失および不利な規則化の損失。 3。 Entropyの正規化は調査および調査を助けるために示されます より良い最適化につながる (Ahmed et al. 2019).

*ログインツイート* ================================================================================================================================================================================================================================================================ *ツイート* ログイン *パイツイート*( )*は、ツイートお問い合わせツイートツイート*)*ツイートツイート ツイート 存储温度*( )*パイツイート*( )*は、ツイートお問い合わせツイートツイート*) ) (3)

どこまでも *フリガナ* entropy を示す *ツイート* entropy を示す 損失の重量。 セット内容 *ツイート* 実験中の実験を0.1に。 *ツイートツイート* インディ状態を cates *ツイートツイート*時間ステップおよび *は、ツイート* 取られた行動を示す お問い合わせ *ツイートツイート*時間ステップ。

広範なテストチャットセッションでポリシーモデルをテストします の 91 件のクエリ, の in-ドメインの問い合わせ(カードの申請方法)、挨拶/理解-edgment の質問(こんにちは、ok、涼しく、ありがとうu)および OOD の照会。 取得したトークン貯蓄の数を判断するため 提案された最適化は、トークンの数を比較します 設定に通常のRAGパイプラインでLMに渡される RAGの最適化についてご紹介します。 数字 LLM 呼び出しのトークンは tiktoken を使って計算された ライブラリ。 通常の RAG パイプラインでは、FAQ コンテキストを取得します。 クエリごとに

ポリシーに基づくアプローチにより、ポリシーモデルが提供 各クエリに対するアクションに対する確率予測。 お問い合わせ モンテカルロドロップアウト方式で不確実性を推定 平均10の予測上の確率。 A [フェッチ] または [NO\_FETCH] アクションは、平均値に基づいて選択されます 確率。

**トップ1スコア(SimThr)の類似点**

先ほど説明したように、社内の組み込みモデルが訓練された infoNCE の損失はのための正確な類似点のスコアを提供します よくある質問と差別化スコア w.r.t. OOD クエリ. エルム top-1の類似点スコアが大きい場合、コールを避けることができます。

6月6日

|  |  |
| --- | --- |
| **お問い合わせ** | i のキャッシュバックが得られるもの スワギー |
| **ポリシーアクション** | [フェッチ] |
| **ボット回答** | 4% 無制限のキャッシュバック クレジットカードでスイギーに。 |
| **人間の楕円** | お問い合わせ |
| **お問い合わせ** | pvrについて |
| **ポリシーアクション** | [NO\_FETCH] は、 |
| **ボット回答** | 4% 無制限のキャッシュバック クレジットカードでPVRに。 |
| **人間の楕円** | お問い合わせ |
| **お問い合わせ** | 治療効果のいくら |
| **ポリシーアクション** | [NO\_FETCH] は、 |
| **ボット回答** | 4% 無制限のキャッシュバック クレジットカードでのお支払い |
| **人間の楕円** | お問い合わせ |

表9:シーケンスの政策モデル予測 同じFAQを参照するクエリ

gpt-2 は、最後のレイヤーの [SEP] から embedding を使用します。ken とそれの上に線形層を追加します。 すべて 残り 実験的な設定は、社内BERTと同じです。 実験。 表10は比較結果を表示します。 使用方法 方針としてgpt-2モデルはトークンsavの~25%を提供します同じ精度で つまり、社内で BERTモデルは、限られたトレーニングセットでより良いを実行します ドメイン事前トレーニングのため。

|  |  |  |  |
| --- | --- | --- | --- |
| **セットアップ** | **# トークン** | **プロモーション** | **Acc。** |
| ログイン Thr + ポリシー (gpt-2) | 124924 | ～25% | 1。 |
| ログイン Thr + ポリシー (BERT) | 115259 | ～31% | 1。 |

表10:オープンでトークン保存の比較 ポリシーモデルとしてのgpt-2

**コンテンツ**

本稿では、AG ベースのチャットボットを an-ドメインを使用してクレジットカード関連の問い合わせをスワッピング FAQ データセット。 問題のパラファラーゼデータを生成しました。 パブリックLLMモデルをモデル化し、社内の埋め込みを訓練 infoNCE 損失を使用して検索するためのモデル。 インハウス モデルは、より大幅に向上するために示されていました ランキング精度の観点から一般公開モデル OOD のクエリ検出。 更に、私達はspeのためにそれを気づくクエリのcificパターン/シーケンス、取得する必要はありません よくある質問 回答を得るためのコンテキスト。 固定を想定 リトリーバルモデルとLLMは、to-数を最適化します。Reinforcement Learning を使って LLM に渡された kens。 お問い合わせ RAGに外部を取り戻すポリシーモデルを訓練 コンテキストを取得するかどうかを決定します。 GPT-4 を使用した 報酬モデルとして。 RLベースの最適化を組み合わせる 類似性しきい値で重要なトークン節約につながります 精度を少し高めながら。 提案する方針-ベースアプローチはジェネリックであり、既存の用途で使用できます。

7月7日