

【公報種別】特許法第17条の2の規定による補正の掲載
 【部門区分】第6部門第3区分
 【発行日】平成30年11月15日(2018.11.15)

【公開番号】特開2016-192020(P2016-192020A)
 【公開日】平成28年11月10日(2016.11.10)
 【年通号数】公開・登録公報2016-063
 【出願番号】特願2015-70768(P2015-70768)
 【国際特許分類】

G 0 6 F 17/30 (2006.01)

【F I】

G 0 6 F 17/30 3 1 0 Z

G 0 6 F 17/30 1 8 0 A

G 0 6 F 17/30 3 4 0 B

【手続補正書】
 【提出日】平成30年10月3日(2018.10.3)
 【手続補正1】
 【補正対象書類名】明細書
 【補正対象項目名】全文
 【補正方法】変更
 【補正の内容】
 【発明の詳細な説明】
 【発明の名称】音声対話装置、音声対話方法及びプログラム
 【技術分野】
 【0001】

本発明は、ユーザとの音声対話を通じて、情報検索等の処理を行う音声対話装置に関するものである。

【背景技術】
 【0002】

従来のパーソナルコンピュータ等における情報検索システムにおいては、検索精度を向上させるため、提示した検索結果に対して、ユーザの満足度を推定することが広く行われている。具体的には、例えば、検索結果として提示した複数の情報から、ユーザがいずれかの情報をクリック（選択）した場合には、ユーザは当該検索結果に満足したと推定することができる。また、例えば、各検索結果のリンク先におけるユーザの滞在時間が長いほど、ユーザが当該検索結果に満足したと推定することができる。

【0003】

これに対して、音声対話を通じて情報検索を行う場合には、入力クエリであるユーザの発話に対して、検索結果として1つの情報を提示するだけで処理が完結することも多い。しかも、このような音声対話システムは、車両を運転中のドライバーにより利用されることが想定されるところ、運転中のドライバーがシステムに対して返すことのできる応答は限られている。したがって、音声検索の場合には、マウスのクリック操作等、従来用いられていた手法に基づいて、検索結果に対するユーザの満足度の推定を行うことは困難である。

【0004】

なお、本発明に関連する先行技術として、以下の先行技術文献がある。

【先行技術文献】
 【非特許文献】

【0005】

【非特許文献1】Qi Guo他「Mining Touch Interaction Data on Mobile Devices to Pre

dict Web Search Result Relevance」SIGIR '13, July 28-August 1, 2013, Dublin, Ireland.

【発明の概要】

【発明が解決しようとする課題】

【0006】

本発明は、上記の問題に鑑みてなされたものであり、音声検索の結果に対するユーザの満足度を適切に推定することができる、音声対話装置等を提供することを目的とする。

【課題を解決するための手段】

【0007】

本発明の音声対話装置は、ユーザの発話内容を認識する音声認識部と、前記音声認識部にて認識されたユーザの入力クエリに応じて、情報を検索する情報検索部と、前記情報検索部にて検索された情報を音声により前記ユーザに提示するための音声応答出力部と、前記音声応答出力部から出力される音声応答に対するユーザの言語的な反応を示す言語的反応特徴量を、前記音声応答を構成する単語ごとに取得する言語的反応特徴量取得部と、前記音声応答出力部から出力される音声応答に対するユーザの非言語的な反応を示す非言語的反応特徴量を、前記音声応答を構成する単語ごとに取得する非言語的反応特徴量取得部と、前記言語的反応特徴量取得部にて取得された言語的反応特徴量及び前記非言語的反応特徴量取得部にて取得された非言語的反応特徴量に基づいて、前記音声認識部における前記入力クエリの音声認識に対する前記ユーザの満足度と、前記情報検索部における情報検索に対する前記ユーザの満足度とを推定する、満足度推定部とを備えた構成を有している。

【0008】

この構成によれば、検索結果として出力される応答音声に対するユーザの言語的な反応と、非言語的な反応に基づいて、ユーザの満足度を推定することができる。ここで、言語的な反応とは、間投詞を含む発話や舌打ち、ため息等の発声による反応であり、非言語的な反応とは、言語的な反応である発話ないし発声以外のユーザの身体的活動（振る舞い）や、心拍数、発汗量等の生体反応である。これらはいずれも、ユーザの自然な反応であって、特別な操作を要しない。したがって、この構成によれば、音声対話を通じた情報検索に対して、適切にユーザの満足度を推定することができる。

【0009】

また、この構成によれば、ユーザの言語的な反応を示す特徴量と、非言語的な反応を示す特徴量は、検索結果である音声応答を構成する単語ごとに取得し、入力クエリの音声認識に対するユーザの満足度と、情報検索に対するユーザの満足度を推定する。音声検索の場合、ユーザの入力クエリの音声認識の結果及び入力クエリから推定されるユーザの検索意図に基づいて、検索結果である音声応答が出力される。このため、ユーザが音声検索の結果に関して不満を持つとすれば、音声認識結果と情報検索結果のうちの少なくともいずれかが原因であると考えられる。そして、ユーザの不満がいずれに起因するかを推定することができれば、検索結果提示後の対話を適切に選択したり、推定結果をフィードバックして検索精度を向上させたりすることができる。したがって、この構成によれば、ユーザの満足度に基づき、音声対話を通じた情報検索を精度よく行うことができる。

【0010】

本発明の音声対話装置において、前記満足度推定部は、学習により生成された系列ラベリングモデルを用いて、各前記満足度の推定を行ってよい。

【0011】

この構成によれば、音声応答を構成する単語と、音声認識に対する満足度及び情報検索に対する満足度を、音声応答を構成する単語について、満足度を正解ラベルとして付与した系列ラベリングモデルを用いて、簡易に、かつ、精度よく推定することができる。

【0012】

本発明の音声対話装置において、前記言語的反応特徴量取得部は、さらに、前記音声応答の出力後の所定時間内の前記ユーザの言語的反応特徴量を取得し、前記非言語的反応特

微量取得部は、さらに、前記音声応答の出力後の所定時間内の前記ユーザの非言語的反応特徴量を取得してよい。

【0013】

この構成によれば、音声応答出力時、どの単語を出力している時にどのようなユーザの反応が得られたかに加えて、音声応答出力後の所定時間内に、どのようなユーザの反応が得られたかも考慮されて、ユーザの満足度が推定される。このように、出力音声応答の各単語と対応付けられる短期的なユーザの反応特徴量に加えて、音声応答全体と対応付けられる長期的なユーザの反応特徴量も取得することで、さらに精度よくユーザの満足度を推定することができる。

【0014】

本発明の音声対話装置において、前記言語的反応特徴量は、前記ユーザの発話の有無、発話内容のうちの少なくともいずれか1つを含み、前記非言語的特徴量は、前記ユーザの表情、ジェスチャー、所定のセンサから取得される身体状態データ、操作の有無のうちの少なくともいずれか1つを含んでよい。

【0015】

本発明の音声対話装置において、前記情報検索部は、前記満足度推定部において推定された、前記情報検索に対する前記ユーザの満足度が所定の閾値よりも低い場合に、情報検索をやり直してよい。

【0016】

この構成によれば、ユーザの満足度に応じて、情報検索からやり直すことで、適切な情報提示を行うことができる。

【0017】

本発明の音声対話装置において、前記音声認識部は、前記満足度推定部において推定された、前記入力クエリの認識結果に対する前記ユーザの満足度が所定の閾値よりも低い場合に、前記入力クエリの認識をやり直してよい。

【0018】

この構成によれば、ユーザの満足度に応じて、音声検索からやり直すことで、適切な情報提示を行うことができる。

【0019】

本発明の音声対話方法は、ユーザの発話内容を音声認識するステップと、認識されたユーザの入力クエリに応じて、情報を検索するステップと、検索された情報を音声により前記ユーザに提示するための音声応答を出力するステップと、出力される音声応答に対するユーザの言語的な反応を示す言語的反応特徴量を、前記音声応答を構成する単語ごとに取得するステップと、出力される音声応答に対するユーザの非言語的な反応を示す非言語的反応特徴量を、前記音声応答を構成する単語ごとに取得するステップと、取得された言語的反応特徴量及び非言語的反応特徴量に基づいて、前記入力クエリの音声認識に対する前記ユーザの満足度と、情報検索に対する前記ユーザの満足度とを含む、音声応答として提示された情報に対するユーザの満足度を推定するステップとを備える。

【0020】

本発明のプログラムは、音声対話を行うためのプログラムであって、コンピュータに、ユーザの発話内容を音声認識するステップと、認識されたユーザの入力クエリに応じて、情報を検索するステップと、検索された情報を音声により前記ユーザに提示するための音声応答を出力するステップと、出力される音声応答に対するユーザの言語的な反応を示す言語的反応特徴量を、前記音声応答を構成する単語ごとに取得するステップと、出力される音声応答に対するユーザの非言語的な反応を示す非言語的反応特徴量を、前記音声応答を構成する単語ごとに取得するステップと、取得された言語的反応特徴量及び非言語的反応特徴量に基づいて、前記入力クエリの音声認識に対する前記ユーザの満足度と、情報検索に対する前記ユーザの満足度とを含む、音声応答として提示された情報に対するユーザの満足度を推定するステップとを実行させる。

【発明の効果】

【0021】

本発明によれば、検索結果として出力される応答音声に対するユーザの言語的な反応と、非言語的な反応に基づいてユーザの満足度を推定するので、ユーザに特別な操作を要することなく、音声対話を通じた情報検索に対して、適切にユーザの満足度を推定することができる。

【図面の簡単な説明】

【0022】

【図1】本発明の実施の形態における音声対話装置の構成を示すブロック図

【図2】教師データの一例を示す図

【図3】満足度推定モデルを用いた満足度の推定を説明するための図

【図4】(a)、(b) 言語的／非言語的反応特徴量の例を示す図

【図5】本発明の実施の形態における音声対話装置の動作フロー図

【発明を実施するための形態】

【0023】

以下、本発明の実施の形態の音声対話装置について、図面を参照しながら説明する。

図1は、本発明の実施の形態の音声対話装置の構成を示す図である。図1に示すように、音声対話装置1は、音声認識部10と、情報検索部11と、音声応答出力部12と、ユーザ操作受付部13と、画像認識部14と、センシングデータ取得部15と、言語的反応特徴量取得部16と、非言語的反応特徴量取得部17と、満足度推定部18とを備える。また、本実施の形態において、音声対話装置1は車両に搭載され、マイク20、スピーカ21、カメラ22、タッチパネル23、生体センサ24、満足度推定モデル記憶部25の各々に接続される。

【0024】

マイク20は、音声対話装置1の主たるユーザであるドライバーの発話である音声データを取得し、音声対話装置1に出力する。スピーカ21は、音声対話装置1の音声応答を出力する。カメラ22は、ドライバーの頭部を略正面から撮像し、取得した画像データを音声対話装置1に出力する。タッチパネル23は、ディスプレイとして情報を表示し、また、各種機器を操作するための入力インターフェースとして機能する。生体センサ24は、ユーザであるドライバーの心拍数、皮膚温、発汗量、脳波等をセンシングする。

【0025】

音声認識部10は、マイク20にて取得したユーザの発話について、音声認識を行い、認識結果を情報検索部11に出力する。情報検索部11は、音声認識部10において、ユーザの発話が情報検索の入力クエリであると認識された場合、認識結果に基づき、情報検索を行う。情報検索部11は、音声対話装置1に接続された所定のデータベース（図示しない）やネットワークを介して、入力クエリに応じた情報検索を行い、検索結果を音声応答出力部12に出力する。

【0026】

音声応答出力部12は、情報検索部11における検索結果をユーザに対して音声で提示するため、ユーザの入力クエリに対応する応答音声を生成し、スピーカ21に出力する。ユーザ操作受付部13は、タッチパネル23にて入力された、ユーザの操作を受け付ける。画像認識部14は、カメラ22から取得した画像データの認識を行う。センシングデータ取得部15は、生体センサ24からセンシングデータを取得する。

【0027】

言語的反応特徴量取得部16は、応答音声出力部12にて生成され、スピーカ21から出力される応答音声に対するユーザの言語的反応を、音声認識部10における認識結果に基づいて取得し、満足度推定部17に出力する。言語的反応特徴量とは、検索結果である応答音声に対してなされたユーザの発話ないし発声である。例えば、ユーザが全く意図しない検索結果が出力された場合、ユーザは、「あれ?」、「えー」等の間投詞を発したり、舌打ちしたり、ため息をついたりする。このように、音声応答出力時のユーザの発話ないし舌打ち音やため息音を含む発声には、音声応答に対するユーザの満足度が表れる。言

語的反応特徴量取得部 16 は、音声応答の出力から所定の時間内に、音声認識部 10 にて、このようなユーザの言語的反応が検出された場合に、音声応答を構成する各単語と、各発話ないし発声に対応付けて満足度推定部 18 に出力する。

【0028】

非言語的反応特徴量取得部 17 は、応答音声出力部 12 にて生成され、スピーカ 21 から出力される応答音声に対するユーザの非言語的反応を、画像認識部 14 における認識結果やユーザ操作受付部 13 にて受け付けたユーザ操作の内容、センシングデータ取得部 15 にて取得されたセンシングデータに基づいて取得し、満足度推定部 17 に出力する。非言語的反応特徴量とは、検索結果である応答音声に対してなされた、言語的反応である発話ないし発声以外のユーザの身体的活動（振る舞い）及び心拍数や発汗量等の生体反応である。

【0029】

例えば、ユーザが全く意図しない検索結果が出力された場合、ユーザは、首を傾げたり、音声応答を無視して、入力クエリとは無関係の機器操作を行ったりする。また、例えば、それまでリラックスしていたユーザが、音声応答の出力により、心拍数や発汗量が増大するなど、緊張した状態になった場合、音声応答に対して不満を持ったことが推測される。このように、音声応答出力時のユーザの振る舞いや生体反応には、音声応答に対するユーザの満足度が表れる。非言語的反応特徴量取得部 17 は、音声応答の出力から所定の時間内に、画像認識部 14 やユーザ操作受付部 13、センシングデータ取得部 15 にて、このようなユーザの非言語的反応が検出された場合に、音声応答を構成する各単語と、ユーザの各振る舞いや生体反応に対応付けて満足度推定部 18 に出力する。

【0030】

満足度推定部 18 は、満足度推定モデル記憶部 25 に記憶されるモデルを参照して、言語的反応特徴量取得部 16 及び非言語的反応特徴量取得部 17 にて取得された各特徴量に基づいて、ユーザの満足度を推定する。

【0031】

ところで、音声対話装置が情報検索を行う場合、ユーザが入力するクエリにおいて、ユーザの検索の意図が完全に明確であるとは限らない。例えば、ユーザが、アイティーラボという会社の場所情報を検索する意図を有する場合、ユーザが「アイティーラボはどこですか」というクエリを入力（発話）した場合には、ユーザの検索意図は入力クエリにおいて明確であるといえる。これに対し、ユーザが単に「アイティーラボ」というクエリを入力した場合には、ユーザの検索意図は多義的に解釈しうる。このような場合、音声対話装置は、ユーザの入力クエリである発話を音声認識した上で、推定されるユーザの検索意図に即した情報検索を行い、その結果を提示することになる。すなわち、音声対話装置が出力する検索結果には、音声認識の結果と、推定されるユーザの検索意図に基づく検索結果とが含まれているといえることができる。

【0032】

したがって、音声対話装置による情報検索に対して、ユーザは、入力クエリの音声認識結果または推定されたユーザの検索意図に基づく検索結果の少なくともいずれかについて不満を持つ可能性がある。例えば、ユーザが「アイティーラボ」という会社の場所を検索する意図を持って、入力クエリとして「アイティーラボ」と発話したのに対して、「ラボ」を「ラブ」と誤認識し、「アイティーラブはパソコンショップです」との音声応答を出力した場合、ユーザは、音声認識及びそれに基づく検索結果の双方に不満を持つことになる。この場合、音声対話装置は、音声認識をやり直したり、ユーザの不満度が大きい場合には、ユーザに対してクエリの再入力を促したりすることが考えられる。

【0033】

他方で、入力クエリを正しく認識できても、ユーザの検索意図を誤推定し、「アイティーラボはデンソーの関連企業です」との音声応答を出力した場合、ユーザは音声認識には不満はないが、検索結果に不満を持つことになる。この場合には、音声対話装置は、検索をやり直したり、ユーザの不満度が大きい場合には、ユーザに対して検索意図を推定する

ための質問を出力したりすることが考えられる。すなわち、ユーザが応答音声に対してどのような不満をどの程度持つかに応じて、音声対話装置におけるその後の処理が変わり得る。

【0034】

本発明の音声対話装置1は、この点に着目し、満足度推定部18は、取得した各反応特徴量に基づいて、音声認識部10における入力クエリの認識結果に対するユーザの満足度と、情報検索部11における情報検索に対するユーザの満足度とを推定する。

【0035】

なお、図1に示す音声対話装置1は、CPU、RAM、ROM、HDD等を備えたコンピュータにより実現される。CPUがROMに記憶されたプログラムを読み出して実行することにより、音声対話装置1の各部の機能が実現される。このような音声対話装置を実現するためのプログラムも本発明の範囲に含まれる。

【0036】

ここで、満足度推定部18が参照する、満足度推定モデルについて、図2及び図3を参照して説明する。本実施の形態において、満足度推定モデルは、例えば、CRF (Conditional Random Fields) や、SSVM (Structured Support Vector Machine) 等の、教師あり機械学習手法によって生成される。図2は、満足度推定モデルを生成するための教師データの一例を示す図である。図2の例では、ユーザの入力クエリは「アイティーラボ」であり、それに対する音声応答は「アイティーラブは東京都渋谷区・・・」である。この教師データにおいては、スピーカが、音声応答のうち「アイティーラブ」と出力した時、ユーザは、自己の入力クエリが誤認識されたことに気づき、「えっ？」という言葉を発し、また、目を大きく見開き、驚きの表情に変化した。また、心拍数、発汗量も増大している。この時の正解データ、すなわち、ユーザの音声認識に対する満足度及び検索結果に対する満足度は、それぞれ、25%、30%である。こうして、ユーザの言語的／非言語的反応特徴量と満足度とが対応付けられる。なお、音声応答は、音声周波数の変化に基づいて、単語分割を行う。

【0037】

また、図2の例においては、応答音声の出力後、所定時間Tまでの間に、ユーザは、タッチパネル上で、ラジオをオンにする操作を行った。ラジオの操作は、図2の例における入力クエリとも、また、応答音声とも無関係である。検索により提示された情報に不満を持つ場合、ユーザはこのような無関係な操作を行うことも多い。したがって、本実施の形態では、単語単位だけではなく、より長期的にユーザの反応を観察して得られる特徴量をも考慮して、ユーザの満足度を推定している。

【0038】

また、図2に示すような教師データに基づいて、各単語の音声認識に対する満足度と当該音声応答全体の音声認識に対する満足度との関係、及び、各単語についての情報検索に対する満足度と当該音声応答全体の情報検索に対する満足度との関係、並びに、当該音声応答全体の音声認識に対する満足度及び当該音声応答全体の検索に対する満足度と、音声応答として提示した情報に対する満足度との関係も学習される。これにより、音声認識・情報検索のいずれに対してどの程度満足しているか、また、より具体的に音声応答のどの部分に対して満足または不満であるかを推定することができる。さらに、提示した情報全体に対する満足度も推定することができる。図3は、このようにして生成された満足度推定モデルを用いた満足度の推定を説明するための図である。図2の例では、長期的反応特徴量は、提示した情報に対する満足度を示す特徴量として説明したが、本実施の形態では、それ以外にも、図3に示すように、音声応答全体の音声認識に対する満足度を示す特徴量や、音声応答全体の情報検索に対する満足度を示す特徴量も取得される。具体的には、例えば、ユーザの発話内容がネガティブであり、その発声音量が大きい場合には、全体の音声認識または全体の情報検索に対する不満度が大きいことを示しているとみることができる。

【0039】

図4(a)は、本実施の形態の音声対話装置1が取得する、言語的反応特徴量の例を示す図、図4(b)は、非言語的反応特徴量の例を示す図である。図4(a)、(b)に示すように、例えば、音声応答出力後所定時間内のユーザの発話や操作の遷移、表情の変化等は、長期的な特徴量となる。

【0040】

図5は、本発明の第1の実施の形態の音声対話装置1の動作フロー図である。まず、音声対話装置1は、ユーザの入力クエリを受け付け(ステップS1)、入力クエリの音声認識を行う(ステップS2)。続いて、音声認識結果に基づいて情報検索を行い(ステップS3)、検索結果に基づいて応答音声を出力する(ステップS4)。そして、言語的／非言語的反応特徴量を取得して(ステップS5)、ユーザの満足度推定を行う(ステップS6)。なお、前述のように、言語的／非言語的反応特徴量は、音声応答を構成する各単語と対応付けて取得されるほか(短期的特徴量)、長期的な特徴量としても取得される。したがって、反応特徴量の取得は、音声応答出力を終了する前(あるいは、音声応答出力開始前)から、音声応答出力終了後所定時間経過前まで行われる。

【0041】

そして、ステップS6にて推定されたユーザの満足度に基づき、音声認識に対する満足度が所定値未満の場合には(ステップS7にてNo)、入力クエリの受付からやり直す。その際、ユーザに対して、再度のクエリ入力を促す音声も出力してもよい。また、ユーザの満足度に応じて、ステップS2の音声認識から処理をやり直してもよい。これに対して、音声認識に対する満足度が所定値以上であるが(ステップS7にてYes)、検索結果に対する満足度が所定値未満の場合(ステップS8にてNo)、情報検索からやり直す。この場合も、満足度に応じて、ユーザに検索意図を質問するなどの処理を追加してもよい。

【0042】

以上、説明したように、本実施の形態の音声対話装置は、ユーザの入力クエリに対して情報検索を行い、検索結果を音声応答で出力する。その際、ユーザの言語的／非言語的反応特徴量を、音声応答を構成する各単語と対応付けて取得する。また、音声応答出力後、所定時間内のユーザの言語的／非言語的反応特徴量を、長期的な特徴量として取得する。そして、取得した反応特徴量に基づいて、推定モデルを参照して、音声認識に対するユーザの満足度、情報検索に対する満足度、提示した情報に対する満足度を推定する。したがって、ユーザの特別な操作を要することなく、音声対話を通じた情報検索に対して、適切にユーザの満足度を推定することができる。

【0043】

なお、上記の実施の形態では、非言語的反応特徴量として、ユーザの表情、操作の有無、生体反応を取得する場合について説明したが、本発明の範囲はこれに限られない。これらの非言語的反応特徴量のうち、いずれか1つまたは2つのみを用いてもよいし、上記以外の反応特徴量を取得してもよい。具体的には、例えば、長期的特徴量として、心拍の低周波領域から取得可能な自律神経系の情報を取得してもよい。

【0044】

また、上記の実施の形態では、ユーザの満足度に応じて、音声認識や情報検索をやり直す場合について説明したが、本発明の範囲はこれに限られない。例えば、推定された満足度を、ユーザの嗜好データベースや、どのようなコンテキストの下で検索が行われたかを示す状況データベースに蓄積し、検索結果のランキングに用いてもよい。

【産業上の利用可能性】

【0045】

本発明は、音声検索の結果に対するユーザの満足度を適切に推定することができるという効果を有し、音声対話装置等として有用である。

【符号の説明】

【0046】

1 音声対話装置

- 1 0 音声認識部
- 1 1 情報検索部
- 1 2 音声応答出力部
- 1 3 ユーザ操作受付部
- 1 4 画像認識部
- 1 5 センシングデータ取得部
- 1 6 言語的反応特徴量取得部
- 1 7 非言語的反応特徴量取得部
- 1 8 満足度推定部