

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第6898322号
(P6898322)

(45) 発行日 令和3年7月7日 (2021. 7. 7)

(24) 登録日 令和3年6月14日 (2021. 6. 14)

(51) Int. Cl.	F I
G 1 0 L 15/22 (2006. 01)	G 1 0 L 15/22 3 0 0 U
G 1 0 L 15/10 (2006. 01)	G 1 0 L 15/10 5 0 0 T
G 0 6 F 3/16 (2006. 01)	G 0 6 F 3/16 6 5 0
G 0 6 F 40/20 (2020. 01)	G 0 6 F 40/20
G 0 6 F 40/30 (2020. 01)	G 0 6 F 40/30

請求項の数 11 (全 21 頁)

(21) 出願番号 特願2018-525444 (P2018-525444)	(73) 特許権者 314015767 マイクロソフト テクノロジー ライセン シング, エルエルシー アメリカ合衆国 ワシントン州 9805 2 レッドモンド ワン マイクロソフト ウェイ
(86) (22) 出願日 平成28年11月10日 (2016. 11. 10)	(74) 代理人 100079108 弁理士 稲葉 良幸
(65) 公表番号 特表2019-502944 (P2019-502944A)	(74) 代理人 100109346 弁理士 大貫 敏史
(43) 公表日 平成31年1月31日 (2019. 1. 31)	(74) 代理人 100117189 弁理士 江口 昭彦
(86) 国際出願番号 PCT/US2016/061303	(74) 代理人 100134120 弁理士 内藤 和彦
(87) 国際公開番号 W02017/083504	
(87) 国際公開日 平成29年5月18日 (2017. 5. 18)	
審査請求日 令和1年10月10日 (2019. 10. 10)	
(31) 優先権主張番号 62/254, 438	
(32) 優先日 平成27年11月12日 (2015. 11. 12)	
(33) 優先権主張国・地域又は機関 米国 (US)	

最終頁に続く

(54) 【発明の名称】 対話補助

(57) 【特許請求の範囲】

【請求項 1】

ユーザに言語インタフェースを提供するための第1の統合コンポーネントを含む統合セクションと、

前記第1の統合コンポーネントからの言語イベントを処理するためのパーサを含むイベント処理セクションと、

前記イベント処理セクションから代替の出力を受信し、動作を選択し、前記受信した代替の出力に基づいて前記動作を実行させるように構成されたダイアログマネージャと、
を含む対話システムであって、

前記対話システムは、前記ユーザとの対話のダイアログ状態を維持し、前記イベント処理セクションからの前記代替の出力が、現在のダイアログ状態から次のダイアログ状態への代替の移行を表し、

前記対話システムは、複数のテンプレートのためのテンプレートライブラリをさらに含み、各ダイアログ状態が、前記テンプレートの1つまたは複数のインスタンスの相互関係の観点から定義されており、

前記パーサは、前記代替の出力を決定するために、現在のダイアログ状態および前記複数のテンプレートに従って言語入力を処理するように構成されており、

前記パーサは、前記現在のダイアログ状態の変更形態として前記代替の出力を決定するようにさらに構成されており、

前記現在のダイアログ状態の前記変更形態は、前記現在のダイアログ状態のテンプレー

10

20

トのインスタンスの相互関係の変化、又は、テンプレートのインスタンスの追加或いは除去を表す、対話システム。

【請求項 2】

前記テンプレートは、複数の入力フィールドおよび前記入力フィールドから演算可能な出力を定義する、請求項 1 に記載のシステム。

【請求項 3】

前記テンプレートは、複数の動作を指定する、請求項 2 に記載のシステム。

【請求項 4】

前記代替の出力の少なくともいくつかの各々は、1 つまたは複数の動作と関連付けられ、前記ダイアログマネージャは、前記代替の出力に従って動作を選択するように構成されている、請求項 1 に記載のシステム。

10

【請求項 5】

前記第 1 の統合コンポーネントは、前記ユーザから音声入力を受信し、前記パーサに対するテキストベースの言語イベントを決定するように構成された自動音声認識装置を含む、請求項 1 に記載のシステム。

【請求項 6】

前記統合セクションは、外部のアプリケーションとのインタフェースおよび前記イベント処理セクションにアプリケーション関連イベントを提供するアプリケーション統合コンポーネントをさらに含む、請求項 1 に記載のシステム。

【請求項 7】

20

前記アプリケーション関連イベントは、ユーザインタフェースイベント、コミュニケーションイベントおよびデータアクセスイベントの少なくとも 1 つを含む、請求項 6 に記載のシステム。

【請求項 8】

顕著な情報を記憶するストレージをさらに含み、前記イベント処理セクションによって受信されたイベントの処理に従って前記顕著な情報を更新するように構成され、前記パーサが、前記イベント処理セクションによって処理された以前のイベントから決定された顕著な特徴の情報をを使用して言語入力を処理するように構成されている、請求項 1 に記載のシステム。

【請求項 9】

30

前記ダイアログマネージャは、前記代替の出力に従って動作を選択する機械学習コンポーネントを含む、請求項 1 に記載のシステム。

【請求項 10】

ユーザに言語インタフェースを提供するための統合コンポーネントを含む統合セクションと、

前記統合コンポーネントからの言語イベントを処理するためのパーサを含むイベント処理セクションと、

前記イベント処理セクションから代替の出力を受信し、動作を選択し、前記受信した代替の出力に基づいて前記動作を実行させるように構成されたダイアログマネージャと、

を含む対話システムであって、

40

前記対話システムは、前記ユーザとの対話のダイアログ状態を維持し、前記イベント処理セクションからの前記代替の出力が、現在のダイアログ状態から次のダイアログ状態への代替の移行を表し、

前記対話システムは、複数のテンプレートのためのテンプレートライブラリをさらに含み、各ダイアログ状態が、テンプレートのグラフで定義され、前記テンプレートは、少なくとも 1 つの入力を有するルートテンプレート及び少なくとも 1 つの子テンプレートを含み、それぞれの前記子テンプレートは、ルートテンプレートの入力に方向付けられたリンクを介して接続されている出力を有し、

前記パーサは、前記代替の出力を決定するために、現在のダイアログ状態および複数の前記テンプレートに従って言語入力を処理するように構成されており、

50

前記パーサは、前記現在のダイアログ状態の変更形態として前記代替の出力を決定するようにさらに構成されており、

前記現在のダイアログ状態の前記変更形態は、前記現在のダイアログ状態のテンプレートのインスタンスの相互関係の変化、又は、テンプレートのインスタンスの追加或いは除去を表す、対話システム。

【請求項 11】

ダイアログを介してユーザと対話する方法であって、

ユーザに言語インタフェースを提供するための統合コンポーネントを操作することと、
前記統合コンポーネントから言語イベントを処理するためのパーサを操作することと、
代替の出力を前記パーサから受信し、動作を選択し、前記受信した代替の出力に基づいて前記動作を実行させるように構成されたダイアログマネージャを操作することと、

複数のテンプレート及びユーザとの対話のためのダイアログ状態を維持することであって、前記ダイアログ状態は、テンプレートのグラフで定義され、前記テンプレートは、少なくとも 1 つの入力を有するルートテンプレート及び少なくとも 1 つの子テンプレートを含み、それぞれの前記子テンプレートは、ルートテンプレートの入力に方向付けられたリンクを介して接続されている出力を有している、

前記パーサからの代替の出力に基づいて、現在のダイアログ状態から次のダイアログ状態への代替の移行を選択することで、現在のダイアログ状態を次のダイアログ状態へ移行させることと、

を含み、

前記パーサは、前記代替の出力を決定するために、現在のダイアログ状態および複数の前記テンプレートに従って言語入力を処理するように構成されており、

前記パーサは、前記現在のダイアログ状態の変更形態として前記代替の出力を決定するようにさらに構成されており、

前記現在のダイアログ状態の前記変更形態は、前記現在のダイアログ状態のテンプレートのインスタンスの相互関係の変化、又は、テンプレートのインスタンスの追加或いは除去を表す、方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

この出願は、2015 年 11 月 12 日に出願された、「AN ATTENTIVE COMMUNICATION ASSISTANT」と称する米国仮特許出願第 62/254,438 号明細書の利益を主張する。この出願は、参照により本明細書に組み込まれる。

【0002】

この発明は、対話補助システムに関し、具体的には、音声を使用してシステムと対話するユーザのための自動補助に関する。

【背景技術】

【0003】

以前の自動ダイアログシステムは、手動で構築されるスロットフィリングアプリケーションに基づいてきた。これらは、通常は、手動で調整され、入力として英語のサブセットのみを受け入れる（これは、以前の自動ダイアログシステムの使用を難しくし、学習を困難にする傾向がある）。そのようなシステムのいくつかは、機械がユーザから会話についての追加情報を収集するモードである混合主導をサポートする。ここ最近では、部分観測マルコフ決定過程（POMDP）手法は、システムの状態を把握するために部分隠れマルコフ過程を使用しており、システムは、各時間に複数の状態を把握し、システムは、各時間に最も有力な推測に従って行動する。そのような以前のシステムでは、過程の意味論は、ハンドコーディングされてきたか、または、ダイアログが十分簡単なものである場合は、簡単な確率的過程として符号化されてきた。意味論は、単語および/または文脈の意味

10

20

30

40

50

または動作と結びつけられている。

【発明の概要】

【0004】

電話による補助の分野では、以前の電話アシスタントは、一般に、対話エージェントではなく、代わりに、発声コマンド/応答システムであった。多くのシステムでは、ユーザは、情報または動作の断片を要求することができ、システムは、音声認識装置が正確である場合や、ユーザがシステムの語彙の範囲内で要求を発した場合は、適切に応答する。しかし、一般に、システムは、脆いものであり、言い換えを理解できないものであり、セッションにわたって文脈を保持できないものであり、大抵の場合は、対話セッション内でさえ文脈を保持できないものであった。

10

【0005】

一態様では、一般に、対話アシスタントは、ユーザとの複数ターン制の対話ダイアログを実施し、その対話ダイアログでは、文脈はターン間で維持され、システムは、ダイアログを管理して、ユーザのための推論される目標を達成する。システムは、ユーザに言語インタフェースを提供するための第1の統合コンポーネントを含む統合セクションを含む。また、システムは、第1の統合コンポーネントからの言語イベントを処理するためのパーサを含むイベント処理セクションも含む。システムのダイアログマネージャは、イベント処理セクションから代替の出力を受信し、動作を選択し、受信した代替の出力に基づいて動作を実行させるように構成される。システムは、ユーザとの対話のダイアログ状態のためのストレージをさらに含み、イベント処理セクションからの代替の出力は、現在のダイアログ状態から次のダイアログ状態への代替の移行を表す。システムは、多数のテンプレートのためのストレージをさらに含み、各ダイアログ状態は、テンプレートの1つまたは複数のインスタンスの相互関係の観点から定義される。

20

【0006】

別の態様では、一般に、対話システムの多数のコンポーネントのパラメータ値を決定するための方法が使用される。システムは、イベントのシーケンスを処理するように構成され、イベントは、言語イベントおよびアプリケーション関連イベントを含み、イベントの処理は、言語イベントを構文解析することと、ダイアログ状態のシーケンスを決定することと、ダイアログ状態のシーケンスに対応するイベントのシーケンスから出力動作のシーケンスを決定することとを含む。方法は、多数のイベントのシーケンスおよび対応する出力動作のシーケンスを収集することを含む。反復は、繰り返される。各反復は、イベントシステムの現在のパラメータ値を使用してイベントのシーケンスを処理することによって、イベントのシーケンスおよび対応する出力動作のシーケンスを処理することを含む。この処理には、イベントのシーケンスからダイアログ状態のシーケンスを決定することが含まれる。出力動作のシーケンスは、ダイアログ状態のシーケンスから決定される。決定された出力動作のシーケンスと収集された出力動作のシーケンスとの比較は、システムの多数のコンポーネントのパラメータ値を更新するために使用される。反復の繰り返しは、伝導の終了に達すると完了する。システムの多数のコンポーネントのパラメータ値は、反復の結果を使用して設定される。

30

【0007】

1つまたは複数の実施形態の利点は、ダイアログ状態が定義されるテンプレートの使用が、それらの状態を明示的に指定する必要なく、可能なダイアログ状態の多数のセットの使用を可能にすることである。その上、システムの構造は、機械学習およびニューラルネットワークコンポーネントのパラメータ値(「訓練」)の効率的で効果的な決定を可能にする。

40

【0008】

本発明の他の特徴および利点は、以下の説明および請求項から明らかである。

【図面の簡単な説明】

【0009】

【図1】対話補助システムのブロック図である。

50

【図 2】カードライブラリを示す図である。

【図 3】ダイアロググラフを示す図である。

【発明を実施するための形態】

【0010】

図 1 は、対話補助システム 100 のブロック図であり、対話補助システム 100 は、例えば、ユーザによって開始されたタスクを遂行するために、複数ターン制のダイアログでユーザと対話するように構成される。そのようなタスクの一例は、システムの音声対話を通じてピザを注文することであり得る。しかし、システムによって、実質的により広いタスクタイプのセットを処理できることを理解すべきである。

【0011】

システムは、外部の統合セクション 110 を含む。外部の統合セクション 110 は、一般に、ユーザや、ユーザと関連付けられた（例えば、ユーザによって制御されるかまたはユーザと対話する）外部のアプリケーションまたはシステムとの直接通信インタフェースを提供する。ユーザとの直接通信の形態の 1 つ（以下で詳細に論じる）は、音声統合コンポーネント 111 を活用する。音声統合コンポーネント 111 は、自動音声認識装置および音声合成器を含む。自動音声認識装置および音声合成器によって、システムは、例えば、電話もしくは他のリモート接続上で、または、ローカルマイクロフォンおよびスピーカ上で、音声によってユーザと直接通信できる。直接通信の他の形態（一般に、各々が、図 1 には示されていない外部の統合セクション 110 の別個のコンポーネントを備える）は、例えば、直接的なグラフィカルユーザインタフェース（GUI）対話であるユーザとのテキストベースの直接通信を提供することができる。外部のインタフェースセクションは、任意選択により、アプリケーション統合コンポーネント 112 を含む。コンポーネント 112 が統合を提供する外部のアプリケーションの例は、Eメールアプリケーションまたはウェブブラウザアプリケーションを含む。Eメールアプリケーションまたはウェブブラウザアプリケーションによって、例えば、システムは、Eメールメッセージを送受信したり、情報の取得または提供を行うためにウェブサイトと対話したりすることができる。

【0012】

外部の統合セクションのコンポーネント 111、112 は、イベントを生成する。イベントは、システム 100 を通じて処理される。例えば、音声統合コンポーネント 111 は、ユーザが出した発声の自動演算転写を提供することができる。システムは、イベントインタープリタセクション 120 を含む。イベントインタープリタセクション 120 は、個々のイベントインタープリタ 121、122 を含む。その結果、例えば、必須ではないが、イベントインタープリタセクション 120 の各コンポーネントが外部の統合セクション 110 の異なるコンポーネントに対するイベントを処理するという形で、外部の統合セクション 110 の各コンポーネント 111、112 は、イベントインタープリタセクションの対応するコンポーネント 121、122 にそのイベントを送る。ユーザからの音声ベースの入力の場合、音声統合コンポーネント 111 は、自動音声認識の結果（例えば、言葉の転写、N 最良転写、単語ラティスなど）をイベントインタープリタの意味パーサコンポーネント 121 に送る。

【0013】

イベントインタープリタセクション 120 のコンポーネントによる処理に役立つシステム 100 の一態様は、本明細書では顕著な特徴の情報 115 と呼ばれる情報の維持である。例えば、この情報は、システムが、ユーザの発声を解釈する際、ユーザの入力における暗黙参照を解決する際およびイベントの処理の間の論理条件を解決する際に使用する文脈上の情報を提供する。顕著な特徴の情報の一部は、比較的静的なもの（例えば、ユーザの住所および電話番号を表す）であり得る。そのような情報は、目標を達成するための住所情報を提供するために使用することができ、ユーザから明示的な承認を引き出せる可能性がある。顕著な特徴の情報の一部は、例えば、同じダイアログの以前の入力によって決定されているかまたはダイアログの合成テキスト出力に従って決定されているなど、はるかに一過性のものであり得る。非常に簡単な例では、そのような情報は、「それ」などのユ

10

20

30

40

50

ーザの入力における言葉を解決するために使用することができる。顕著な特徴の情報セクション 115 には、外部の統合セクション 110 のコンポーネント、イベントインタプリタセクションのコンポーネントおよびダイアログコントローラ 130 から直接もたらされる場合を含めて、複数の情報源が存在しうる。それらについては以下でさらに詳細に説明する。また、顕著な特徴の情報セクション 115 の情報は、例えば、外部の統合セクションが使用することもできる。例えば、顕著な特徴の情報は、ユーザへのテキストベースまたは音声ベースの出力のためのテキスト生成に役立ち得る。

【0014】

イベントインタプリタ 120 のコンポーネントが動作する方法にとって重要なシステム 100 の別の態様は、カードテンプレートライブラリ 125 で定義される「カード」の使用、および、これらのカードの観点から指定されるダイアログ状態 135 の維持に関連する。以下でさらに詳細に論じられるように、外部の統合セクション 110 からイベントインタプリタセクション 120 によって受信された各イベント 118、119 に対し、イベントインタプリタセクションは、1組の提案「計画」129を提供する。一般に、計画は、あるダイアログ状態から別のダイアログ状態へ（一般に、現在のダイアログ状態 135 からその現在の状態の変更形態である別の状態へ）の移行を指定する。そのような変更形態は、カードをグラフに追加すること、カードを別のカードと交換することまたはグラフにおいて既存のカードを再配列することを含み得る。それに加えて、計画は、移行と関連付けられた1つまたは複数の動作を指定し、一般に、計画の新しい状態は、それらの動作が実行されていることを指定する。例えば、動作は、電話番号をダイヤルすること、Eメールを送信すること、何らかのテキストをユーザに読み聞かせることまたはユーザに質問すること（「Eメールの送信先は？」など）をシステムに行わせることができる。一般に、イベントインタプリタセクション 120 は、次の状態に関する最終的な決定を試みない。その決定は、ダイアログコントローラ 130 に委ねられ、ダイアログコントローラ 130 は、提案計画 129 を受信し、計画を選択し、実行すべき1つまたは複数の関連動作を開始する。

【0015】

イベント 118、119 を処理することは、イベントインタプリタセクション 120 が1組の提案計画をダイアログコントローラ 130 に提供することを続行する。また、本質的には、自律的に生成された可能な移行は、現在の入力イベントを考慮することなく、ダイアログ状態のみに基づくため、ダイアログコントローラは、ダイアログ状態 135 からさらなる提案計画を受信することもできる。ダイアログコントローラは、次のダイアログ状態 135 を選択する。選択した次のダイアログ状態と関連付けて、ダイアログコントローラは、次のダイアログ状態をもたらす計画と関連付けられた動作 139（または、より一般的には、1つまたは複数の動作）も決定し、その計画を達成するように動作コントローラ 140 に指示する。例えば、動作は、タスクの完了に必要なある情報を要求するプロンプトを発行することによってユーザからの入力を求めることであり得る。動作コントローラ 140 は、動作を達成するために、外部の統合セクション 110 のコンポーネントと対話する。例えば、音声統合コンポーネント 111 は、動作コントローラ 140 のコマンドで、テキストを生成し、音声によってユーザに提示するためにそのテキストを合成することができる。一般に、このイベント処理のサイクルは、タスクが完了するまで、ユーザとの複数の対話「ターン」）に対して続く。

【0016】

システムの態様は、ダイアログのためのすべての可能な構造を事前に決定しておかなければならないわけではないことである。むしろ、ダイアログの構造は、カードテンプレートライブラリ 125 で定義される1組のカードテンプレート（一般に、手動で作成されている）に基づいて、ランタイムの間に決定される。以下でより完全に論じられるように、一般に、カードは、事前に定義されたタイプの出力および多くのフィールド（各々は、事前に定義されたタイプのもの）を有する。これらのフィールドの値は、入力として考えることができ、カードの出力を定義する。

10

20

30

40

50

【0017】

図2を参照すると、カードライブラリ125の例は、多くのカードテンプレート225を含む。図2に示されるように、4つのカードが示されており、それらは、ピザ注文ドメインに関連する。「ピザの注文」と称する1つのカードは、「PizzaOrder」という事前に定義されたタイプの出力を有する。このカードは、4つのフィールドを有し、4つのフィールドは共に、ピザ注文を特徴付けるのに十分な情報を提供する。「Address」という事前に定義されたタイプの「住所」フィールドは、注文の配達住所を提供し、「CCInfo」（すなわち、クレジットカード情報）という事前に定義されたタイプの「支払い」フィールドは、注文のクレジットカード支払いの詳細を提供し、「Pizza」という事前に定義されたタイプのアイテムのセット（またはシーケンス）である「ピザ」フィールドは、注文自体の詳細を提供し、最後に、「Phone」（すなわち、電話番号）という事前に定義されたタイプの「電話」フィールドは、注文が来た番号を提供する。

10

【0018】

一般に、各データタイプは、そのデータタイプの出力値を提供することができる1つまたは複数のカードを有する。図2に示されるように、「友人の場所」というカードは、「Location」というタイプの出力を提供する。例えば、ユーザが、彼の友人のJohnの家に配達予定のピザを注文する場合は、「友人の場所」というカードを使用して、ユーザが注文しているピザをどこに配達するかを決定することができる。

20

【0019】

図3を参照すると、ダイアログ状態135は、対話の状態を表すダイアロググラフ335で表すことができる。グラフのノードは、カードライブラリ125からのカードのインスタンス、および、各々があるカードの出力から他のカードの1つまたは複数の入力まで方向付けられたリンクに対応する。例示的なダイアロググラフ335として、グラフのルートノードは、「ピザの注文」というカードのインスタンスである。図3に示される状態では、「ピザの注文」というカードの「住所」の入力は、「場所からの住所」というカードにリンクされ、そのカードの「場所」の入力は、上記で紹介された「友人の場所」というカードのインスタンスにリンクされる。最後に、「友人の場所」というカードの「誰か」の入力は、「人物」というカードにリンクされる。ユーザとシステムとの間の対話の多くの異なる発声または複数のターンはこのダイアロググラフにつながる可能性があるが、発声の1つは、以下の通りであり得る。

30

ユーザ：「Johnの家にピザの配達をお願いしたいのですが」

または

ユーザ：「ピザの注文をお願いします」

システム：「どこへ配達しますか？」

ユーザ：「Johnの所に」

システム：「Johnの家ですか、それとも、彼のオフィスですか？」

ユーザ：「Johnの家に」

【0020】

いずれの時点においても、ダイアロググラフ335は、例えば、複数のカードが1つのカードへの入力値を潜在的に提供するまたは複数のルートカードがあるなど、曖昧なものであり得ることに留意されたい。ダイアログ状態135は、カードインスタンスのネスティングと同等に表すことができ、例えば、「ピザの注文」というカードインスタンスの「住所」の入力はネスティングされた「場所からの住所」というカードインスタンスであるなど、1つのカードインスタンスの入力の値は、ネストカードインスタンスとして表される。しかし、1組のカードインスタンスを使用してダイアログ状態を表す特定の方法は決定的なものではないことを理解すべきである。

40

【0021】

上記で紹介されるように、音声入力が一例である（テキスト入力に加えて）、システムによる言語入力の処理に移ると、外部の統合セクション110は、音声統合コンポーネン

50

ト 1 1 1 を含み、音声統合コンポーネント 1 1 1 は、イベントインタープリタセクション 1 2 0 の意味パーサ 1 2 1 に自動処理音声入力 1 1 8 (すなわち、言語入力)を提供する。音声認識装置の出力は、テキスト(または言葉のシーケンスの他の表現)の形態であり、N 最良リストまたは 1 組の可能な言葉のシーケンスを表すラティスの形態の音声入力の代替の解釈を表し得る。

【 0 0 2 2 】

意味パーサ 1 2 1 は、音声認識装置 1 1 0 の出力を処理する際に、顕著な特徴の情報 1 1 5 およびカードテンプレート 1 2 5 を活用する。一般に、ユーザからの各入力発声に対し、パーサは、特定のカードのフィールド値および新しいカードの参照が入力で表されているかどうかを判断する。意味パーサの出力は、ダイアログコントローラ 1 3 0 に送られ、ダイアログコントローラ 1 3 0 は、その時点で知っている情報(顕著な特徴の情報 1 1 5 を含む)に基づいて、ユーザとの対話をどのように導くかを決定する。全体的な目標および従属サブダイアログまたは下位目標を表すダイアロググラフ 3 3 5 を維持するのは、ダイアログコントローラ 1 3 0 である。

【 0 0 2 3 】

システムの 2 つの重要な態様は、例えば、明示的な値としてのまたはその出力が値を提供するカードのインスタンス化によるグラフの変数の決定に必要な情報を引き出すために、ユーザとの対話の間にダイアロググラフが構築される方法や、ユーザとの対話を導くためにダイアロググラフがどのように使用されるかである。

【 0 0 2 4 】

ダイアロググラフの構築は、特定のカードにマッピングされるユーザ入力のインスタンスを検出することができる意味パーサのオペレーションにかなり依存する。このプロセスを支援するため、一般に、各カードは、ユーザの入力を解釈する際に意味パーサによって使用される分類子またはトリガ定義(例えば、トリガフレーズのリスト)の指定と関連付けられている。以下でさらに論じられるように、そのような分類子またはトリガフレーズは、例えば、カードライブラリにカードを追加するシステム設計者が、手動で作成することができる。好ましくは、システムは、ユーザ入力と有効な解釈とをペアにする訓練データを活用するプロセスを通じて、分類子を学習する。

【 0 0 2 5 】

一般に、意味パーサ 1 2 1 は、意図の表現を生成するためにユーザ発声を解釈し、これらの意図の表現をダイアログコントローラ 1 3 0 に送る。ダイアログコントローラ 1 3 0 は、上記で紹介されたダイアロググラフ 1 3 5 を含むダイアログ状態を維持するために、これらの意図の表現を処理する。一般に、意味パーサは、例えば、エンティティ(例えば、「それ」、「Eメール」、「Dan の Eメール」、「Dan Klein の Eメール」、「TTS についての Eメール」など)の参照を解決するために、顕著な特徴の情報 1 1 5 を活用してテキスト入力 1 1 8 を処理する。意味パーサは、問題のエンティティの現在の顕著な特徴および他の同様の顕著であるエンティティを見て、正しい参照表現を決定する。上記で紹介されるように、パーサは、入力における参照であるカードテンプレートの識別や、入力からおよびそれらの識別されたカードに対する顕著な特徴からの情報の記入も行う。また、パーサは、タスク状態の一部を表す現在のダイアログ状態 1 3 5 へのアクセスも有し、ダイアログ状態で既にインスタンス化されたカードに記入する必要がある値を提供する。また、パーサは、その情報が新しいカードまたはダイアログ状態で既にインスタンス化されたカードに関係しているかどうかにかかわらず、発声において見られるエンティティで顕著な特徴の情報 1 1 5 の更新も行う。

【 0 0 2 6 】

この実施形態における意味パーサ 1 2 1 は、ハイパーグラフ重み付け演繹システム(hypergraph weighted deduction system)である(例えば、Pauls, A. D., "Optimal Search Algorithms for Structured Problems in Natural Language Processing," Ph.D. Thesis, EECS Depar

10

20

30

40

50

tment, University of California, Berkeley, 2012を参照)。そのようなパーサは、「公理」と呼ばれる「アイテム」の初期のセットから始まり、1つまたは複数の目標アイテム（発声の「完全な」解釈を表す）を生成するために、「演繹ルール」を使用して、アイテムを組み合わせるより大きなアイテムにする。意味パーサ121は、2種類のアイテムを有する。第1のアイテムは、カードとダイアログマネージャのカードグラフにおけるその場所の両方を説明する「ホーム」アイテムである。別のアイテムは、グラフに場所がないカードである「オーファン」アイテムである。すべてのホームアイテムは、目標アイテムと考えられる。

【0027】

公理に対し、このパーサは、自動音声認識装置110の出力、ダイアロググラフ135を含むタスク状態および顕著な特徴115からの情報を使用する。現在のタスクグラフのすべてのカード（ホーム）は、公理として生成される。それに加えて、パーサは、ASR転写物（またはラティス）を使用して、ある特定のキーワード、キーワードの組合せ、または、他のシステムもしくは辞書により導入される追加の情報（顕著な特徴状態など）によってもたらされるカード（オーファンまたはホーム）を生成する。例えば、「同じメッセージをJohn Smithに送信します」という発声は、「Eメール送信」という公理、「テキストメッセージ送信」という公理、「連絡先（John Smith）解決」という公理および最も顕著なメッセージを表す公理をもたらす。

【0028】

演繹ルールに対し、パーサは、次いで、カードを互いに接続することによってアイテムを組み合わせるより大きなアイテムにする方法を考慮する。例えば、パーサは、最も顕著なメッセージを表すカードを取り、その「受信者」というサブカードを「連絡先（John Smith）解決」と交換することができる。最も顕著なメッセージがEメールである場合は、「Eメール送信」とリアドレスメッセージを表すアイテムを組み合わせることができ、そうでなければ、「テキストメッセージ送信」とメッセージを組み合わせることができる。

【0029】

一般に、すべての公理が使用されるという制約も、転写物またはラティスの時間帯におけるすべての言葉が構文解析で使用される公理によって「カバーされる」という制約もないことに留意されたい。例えば、完全に有効な（ただし間違っただけ）構文解析は、上記のように受信者フィールドを交換せずに、単に、「Eメール送信（メッセージ＝最も顕著なメッセージ）」であり得る。実際に、パーサによって提案されたすべてのホーム構文解析アイテムは、「完全な」構文解析と見なされる可能性がある。それらの多くは、間違いにすぎない。

【0030】

良い構文解析と悪い構文解析のどちらかを選ぶため、パーサは、事前に指定された機能を使用して、スコアをアイテムに割り当てる。この文書でさらに論じられるように、この機能は、手動で設計することも、線形分類子またはニューラルネットワークを使用して、機械学習で推定することもできる。

【0031】

意味パーサは、構文解析を生成するため、ユーザの発声の前に決定された現在のダイアログ状態からの情報を発声自体のテキストと組み合わせる。この特徴は、文脈上の解釈を可能にする。パーサは、「メッセージ」という言葉はある文脈では「テキストメッセージ」を意味し、別の文脈では「Eメールメッセージ」を意味し得ると正しく推論することができる。文脈を知っていなければ、意味パーサは、どの分析がより正しいかを知るすべがない可能性がある。一度に1つの入力発声のみを通常は処理する以前の意味パーサとは異なり、このシステムの意味パーサは、ユーザの趣旨のいくつかの決定プロセスを知らせるため、顕著な特徴の情報および現在のダイアログ状態を考慮する。現在のシステムでは、意味パーサは、実行しているカードの顕著な特徴、タスクによって示されているそれらのカードおよび他のカードのアイデンティティ、ならびに、ユーザの音声をどのように解釈

10

20

30

40

50

するかを決定する際のエージェントの出力動作および出力発声を考慮する。

【 0 0 3 2 】

意味パーサ 1 2 1 の機能をレビューするため、意味パーサ 1 2 1 は、音声統合コンポーネント 1 1 1 からのテキスト入力 1 1 8 の解釈に基づいて顕著な特徴の情報を使用および更新する。また、意味パーサ 1 2 1 は、現在のダイアログ状態 1 3 5 を使用し、ダイアログコントローラ 1 3 0 に送る情報を生成する。また、その入力を処理する際、意味パーサは、テンプレートライブラリの特定のカードテンプレートと関連付けられたルールまたは他の情報を用いるように構成され、それにより、カードのインスタンスに対応するテキストのインスタンスの検出が可能になり、また、カードによって参照される（すなわち、カードの入力および出力として）既知のデータタイプのインスタンスを構文解析または検出するようにも構成される。従って、カードテンプレートライブラリは、意味パーサまたはシステムの他のコンポーネントの変更を必要とすることなく、システムのドメインを拡張するための方法を提供する。カードの指定の分離および意味パーサの構成により、意味パーサ自体を変更する必要なく、新しいカードを作成することによる新しいタスクへのシステムの拡張が可能になることに留意されたい。

10

【 0 0 3 3 】

ここでダイアログコントローラ 1 3 0 のオペレーションに移ると、ダイアログコントローラの機能は、各イベント入力後、本質的に次に何をするかを決定することである（すなわち、それは「決定者」である）。1組の候補計画（意味パーサ 1 2 1 によって提案されたおよび/またはダイアログ状態のカードから自律的に）を考慮すると、ダイアログマネージャは、それらの計画のうちのどれを選択してその動作を実行すべきかを選ぶ。しかし、ダイアログマネージャは、何もしないこと、特定の計画を実行すべきかどうかを尋ねること、数個の計画のうちのどれを実行すべきかをユーザに尋ねることを理解すべきである。これらの代替案は、「何もしない」というカード、「承認」というカード、「選ぶ」というカードまたは「助けを求める」というカードを使用することによって、計画として策定することもできる。ダイアログマネージャは、これらの代替の計画を候補計画のプールに追加し、他の提案計画と並行してそれらをスコア付けする。

20

【 0 0 3 4 】

従って、ダイアログマネージャ 1 3 0 は、3つの責任を有する。最初に、ダイアログマネージャ 1 3 0 は、スコア関数に従って提案計画をランク付けする。次いで、ダイアログマネージャ 1 3 0 は、それらの計画のスコアと併せて新しい候補計画を追加する。最後に、ダイアログマネージャ 1 3 0 は、最良のスコア計画を選ぶ。計画のスコア関数は、手で構築されるものでも、機械学習を使用するものでもよい。機械学習アルゴリズムは、構造化サポートベクタマシンなど、いかなる構造化分類子でも、ランキングアルゴリズムでもよい。

30

【 0 0 3 5 】

ダイアログマネージャ 1 3 0 の役割の別の態様は、ダイアログ状態 1 3 5 の「焦点」の維持である。上記で論じられるように、ダイアログ状態 1 3 5 のダイアロググラフ 3 3 5 は、カードのインスタンスの相互接続（または同等のネスティング）によって形成される。ダイアログマネージャは、これらのカードのうちの1つをダイアログの焦点として識別する。この焦点カードは、動作を呼び起こすカードであり得る。焦点の識別は、例えば、意味パーサが動作から生じ得る入力を解釈する際に役立つ。

40

【 0 0 3 6 】

引き続き言語の入力および出力の事例を考慮すると、ダイアログマネージャによって選択されたある特定の動作は、テキストの生成と、音声出力の場合は、そのテキストの音響バージョンの合成とを伴う。音声統合コンポーネント 1 1 1 は、テキスト生成器と、テキスト生成器の出力を受信してオーディオ提示のための出力を提供する音声合成器とを含む。

【 0 0 3 7 】

カード設計者が出力のための言語情報を指定する際、カード設計者は、ストリングを直

50

接記載するよりむしろ、メッセージを説明する簡単な論理形式を使用して言語情報を指定する。次いで、この論理形式は、テキスト生成器によってストリングとしてレンダリングされる。生成メカニズムの中央集権化は、2つの主要な利益を提供する。第1に、カードインプリメンタは、文法のルール、正しい代名詞の使用などについて知る必要はない。第2に、言語の変形例を自動的に導入し、ユーザとシステムとの間の対話をより自然なものにすること(「JohnにEメールを送信しますか? (「would you like to send John an email」、「would you like to send an email to John?」)」)が簡単である。

【0038】

論理形式は、論じられているエンティティの特性(例えば、Eメールの送信者または件名)またはダイアログのタスク(例えば、システムが、現在のタスクを終了する前に、新しく受信したEメールを読むべきかどうかを尋ねる)を説明することができる。可能な一実施形態では、論理形式は、トップレベルのダイアログ作用、イベント、エンティティまたは属性としてすべてのノードが指定され、追加のキーが追加の精緻化を提供する(イベントの参加者、エンティティの名前など)、簡単なネオデイヴィッドソン意味論に対応するキー値ペアとして実装される。最後に、現在のダイアログ状態の断片に対応するこの論理形式の一部分(カードまたはエンティティ)は、適切な識別子でタグ付けされる。

【0039】

例えば、ダイアログエージェントは、以下のように、「パワーラインレポート」という件名でEメールを送信するというその趣旨を表現することができる。

```
{
  type: InformIntention
  body: {
    type: event
    name: send
    agent: DialogueAgent
    id: card###
    theme: {
      type: entity
      name: email
      subject: power line report
      id: email###
    }
  }
}
```

【0040】

この論理形式は、多くの方法でレンダリングすることができる。

「パワーラインレポート」という件名でEメールを送信する

パワーラインレポートについてのEメールを送信する

パワーラインレポートEメールを送信する

それを送信する

など。

【0041】

意味パーサ121と同様に、音声統合コンポーネント111のテキスト生成器は、顕著な特徴の情報115を使用して、論理形式からテキストをどのように生成するかを決定する。上記の例で見られるように、システムは、他のEメールと比べて問題のEメールがどれほど顕著であるかに応じて、様々な程度の特異性で問題のEメールを説明することができる。

【0042】

各論理形式に対し、テキスト生成器は、1組の候補発声を生成する(明示的にまたは構

10

20

30

40

50

文森として実現される)。これらの候補は、再帰的に生成される。論理形式(またはその断片)を考慮すると、テキスト生成器は、ストリングとして論理形式全体をレンダリングするためのルールまたはテンプレートを使用してその一部を実現するためのルールを有することができ、次いで、各サブパートを独立して実現することによって完了する。

【0043】

これらのルールは、手動で記載することも、構造化言語資源と非構造化言語資源の両方から自動的に取り出すこともできる。例として、所定のいかなるイベントに対しても、テキスト生成器は、OntoNotesコーパスからのデータならびに論理形式の自然言語表現を選ぶためのオープンウェブおよび人間の注釈者からのデータ集合体を使用する。OntoNotesコーパスは、各動詞をその論証とリンクし、システムの実施形態は、それらの論証とペアになった動詞のレンダリングを提供するために情報を使用することができる。

10

【0044】

1組の候補発声を考慮すると、テキスト生成器は、スコア関数を使用して各発声を重み付けし、各発声のスコアに比例して各発声からサンプリングする。意味パーサのように、この機能は、手動で設計することも、線形分類子またはニューラルネットワークを使用する機械学習によって推定することもできる。

【0045】

自動音声認識および音声合成は必要ではないことを除き、音声入力処理および音声出力の提供の説明は、テキストベースの入力および出力に適用可能である。その上、対応するイベントインタープリタ122によるアプリケーション統合コンポーネント112からのイベントの処理は、ダイアログ状態135の既存のカードに値を記入することによってまたはアプリケーションイベントと関連付けられたカードのインスタンスを導入する状態変化を伴う計画を提案することによって、提案されたダイアログ状態の変化を直接供給することができる。

20

【0046】

上記で説明される様々なコンポーネントによって実装される手順は、完全に手動で構成しなければならないわけではない。むしろ、多くのコンポーネントは、トリガ段階とカードテンプレートとの関連付けまたはユーザへの出力の論理形式など、手動で構成された態様を有し得るが、システムのオペレーションの多くは、代替案の比較的限られたセットからのランク付けおよび選択を伴う。これらの代替案のランク付けおよび選択は、ユーザとシステムとの間の対話の代表的なデータを使用する自動技法を使用するように構成されていてもよい。

30

【0047】

一般に、機械学習(ML)または人工ニューラルネットワーク(ANN)の使用の様々な公知の技法は、システムの様々なコンポーネント(例えば、パーセプトロン、フィードフォワードシステム、畳み込みニューラルネットワーク、長期・短期メモリシステム、注意システム、サポートベクタマシンまたは他の訓練アルゴリズムなどの構造分類子またはランキングアルゴリズム、交差エントロピー、尤度、確率、誤差率または他の尺度などの最適化関数)によって使用される。非常に一般的には、訓練システムは、システムのすべての入力、出力および中間状態を取り、音声統合コンポーネント111の音声認識装置およびテキスト生成器の性能、意味パーサ121の性能、ならびに、ダイアログマネージャ130の性能を最適化する。最適化は、交差エントロピー、尤度、確率、誤差率または他の尺度の関数であり得る。

40

【0048】

MLおよびANN技法の使用の一態様は、所望の成果を表す「訓練データ」の必要性である。例えば、ダイアログマネージャの事例では、この訓練データは、意味パーサによって提案された計画に基づく選択すべき望ましい計画の表示を含み得る。システムがそのような訓練データを決定する方法の1つは、「ウィザードオブオズ」(Woz)モードである。このモードでは、ユーザは、システムに何でも自由に尋ねることができ、ヒューマン

50

エージェントは、彼の力を最大限まで出して、システムへの返答を形成する（またはいくつかのサービスが利用可能ではないと応答する）。このモードは、データ収集や、プロトコルに組み込まれた自動または手動補助による後の最適化への支援を意図する。

【 0 0 4 9 】

「ヒューマンオペレータ」モードでは、人間は、システムへの返答を形成せず、むしろ、システムによって識別されたオプションから選択を行う。例えば、ヒューマンオペレータには、音声統合コンポーネントによって決定されたトップの自動転写物を提示することができ、ヒューマンオペレータは、「最良」のものを選択することができる。同様に、意味パーサによって提案された計画の中から、ヒューマンオペレータは、入力および現在のダイアログ状態を考慮して、最も適切なものを選択することができる。ヒューマンオペレータの選択は、記録され、かつ、ヒューマンオペレータなしの完全自動モードでの自動オペレーションのための自動選択またはスコア特徴を訓練するために使用される。

10

【 0 0 5 0 】

ヒューマンオペレータモードと完全自動モードとの中間として、混合自動/ヒューマンモードは、選択およびスコア付けまたは代替案を実行し、いくつかの事例では、人の介入なしで進行する。しかし、自動手順に不確実性または曖昧さがあることを自動選択またはスコア付けが示す場合（例えば、複数のオプションが同様のスコアを有するため）は、ヒューマンオペレータは、ヒューマンオペレータモードと同様に、決定を行うことが求められる。これらの決定は、MLコンポーネントおよびANNコンポーネントの今後の訓練のための訓練データを増大するために使用される。

20

【 0 0 5 1 】

訓練データの別の供給源は、システムにおける中間データの人間の注釈からもたらされる。例えば、ヒューマンエキスパートは、手動で意味構文解析の正しい構文解析に注釈を付け、同じ構文解析を生成するようにパーサを訓練することができる（恐らくは、部分的な信用を可能にするために、構文解析における間違っただサブカードの数を最小化するなどの損失関数を使用して）。

【 0 0 5 2 】

一般に、中間データの注釈は必要ではない。例えば、意味パーサによる構文解析出力は、実行された場合に（例えばダイアログマネージャによって選ばれた計画として）ヒューマンエージェントの挙動と整合するであろう構文解析を見つけることによって決定しなければならない「潜在的変数」として扱われる。例えば、そのような手法のための訓練データは、ユーザの発声およびイベント（例えば発声または新しいEメールの到着）のシーケンスを含みうる。出力は、それらのイベントに対応する動作のシーケンスである。EMアルゴリズムなどの教師なし訓練アルゴリズム（例えば、A . P . Dempster , N . M . Laird , and D . B . Rubin . " Maximum Likelihood from Incomplete Data Via the EM Algorithm , " Journal of the Royal Statistical Society : Series B , 39 (1) : 1 - 38 , November 1977を参照）は、どの構文解析（ならびにダイアログマネージャおよび動作コントローラ出力）が正しい出力のシーケンスを生成することができるかを推論するために使用することができる。一般に、別個のパラメータは、音声統合コンポーネント111、意味パーサ121およびダイアログマネージャ130などの各コンポーネントに対してこのような方法によって訓練される。

30

40

【 0 0 5 3 】

そのような訓練アルゴリズムの重要なコンポーネントは、「説明のつかない」ヒューマンオペレータ生成出力および動作（すなわち、システムが利用可能な構文解析または動作によって生成することができない出力）を処理する能力である。このような出力は、特別な構文解析および動作を提供することによって説明することができる。特別な構文解析および動作は、可能ないかなる出力も説明することができるが、訓練の間のその使用には通常の構文解析および動作が好まれるように重いペナルティが科される。

50

【0054】

音声統合コンポーネントのテキスト生成部分を訓練するための一手法は、ヒューマンエキスパートを利用する。シューマンエキスパートは、システムによって生成されたサンプル論理形式に対応するテキストを提供する。テキスト生成器は、入力として論理形式を考慮し、例えば生成テキストのBLEU (<http://aclweb.org/anthology/P/P02/P02-1040.pdf>) スコアを最大化するなどの損失関数を使用して、それらの出力を生成するように訓練される。

【0055】

意味パーサは、出力計画をスコア付けまたは選択するためのアイテムの多くの特徴を使用する。これらは、以下を含む。

1. 作成された新しいカードのタイプおよび何枚作成されたか。
2. テキスト入力に応じたトリガ言葉の確率（顕著な特徴を説明するために再度スコア付けされた後である可能性がある）。
3. カードおよびエンティティの顕著な特徴。
4. 既存のカードに対して変更は何回行われたか。
5. 「実は」（ユーザが何かを変更することを希望することを示す）または「また～も」（ユーザが何かを追加することを希望することを示す）のような、談話における有益な言葉の存在。

6. 入力発声のいくつの言葉が使用されるか。
7. いくつの言葉が複数回使用されるか。
8. 組み合わせられた2つのアイテムに対して、それらのトリガ言葉（もしあれば）が入力発声においてどれだけ離れているか。
9. 以前の動作において調整された可能性がある、ヒーローエージェント対話（本明細書の他の場所で提示されるデータ収集についての議論を指す）から収集されたデータにおいて構文解析がどれだけ頻繁に起こるかによって構文解析をスコア付けする動作モデル。説明される実施形態では、動作モデルは、言葉ではなく動作に対するnグラム言語モデルである。

【0056】

システムは、直接または間接監視を使用して訓練することができる。直接監視では、システムは、候補計画（正しい計画を含む）およびダイアログ文脈のリストという形態で訓練データを受信し、正しい計画を選択するように訓練される。

【0057】

このデータを収集するための方法の1つは、ヒューマンオペレータモードを用いることである。ヒューマンオペレータモードでは、システムは、計画のランク付けされたリストを継続的に提案し、人間は、候補のうちの1つを選択したりすべての候補を拒否したりすることができる。同様に、混合モードでは、システムは、信頼度閾値（数個の計画が同様のスコアを有する際など）または別のメカニズムに基づいて助けを求めるためにヒューマンエージェントに従うことを選択することができる。人間の選択は、訓練データとして記憶することができる。

【0058】

また、ダイアログマネージャは、意味パーサの訓練に対して説明されるものと同じ潜在変数手法を使用する間接監視を考慮して、訓練され得る。別には、さらに一層間接的な形態の監視は、ヒューマンオペレータの直接の介入なしで、エンドユーザ対話から学習することである。このモードでは、システムは、強化学習エージェントとして動作し、正しい動作または間違った動作を実行したキュー（「報酬」）を探す。このとき、システムは、再び間違いを犯す可能性を低くするためにスコア関数の重みを更新する。これらのキューは、ユーザによって明示的に提供されたり（例えば、事前に指定されたキーフレーズを言うこと、電話を振ることまたはボタンを押すことによって）、あるいは、ユーザの音声の音響特性によってユーザのフラストレーションレベルを測定するなどの方法によって暗黙的に提供されたりする。学習アルゴリズムは、Q学習、深層Qネットワークまたは時間

10

20

30

40

50

差学習など、いかなる強化学習アルゴリズムでもあり得る。

【0059】

上記で説明される訓練のための手法に加えて、ユーザとの対話から収集されたデータ（トレースデータと呼ばれる）からのオフラインの統合方法またはダイアログが進行中の間のオンラインの統合方法で、システムのコンポーネントを訓練することが可能である。トレースデータから訓練する際、システムは、入力としていかなる注釈も含む完全なまたは進行中の通話記録を取り、その予測を通話記録によって提示されたものと同じものにするためにその重みを更新する。この項目の目的のため、記録は、以下を含み得る。

1．利用可能な場合はラティスまたはk最良リストを含む、ユーザの音声の音声認識転写物。

10

2．ヒューマンエージェントが音声出力を提供する場合は、エージェントに対する音声認識転写物。

3．アプリケーションイベント（例えば、グラフィカルインタフェース、DOM（文書オブジェクトモデル）イベント、HTTP要求などにおけるアイテムの選択）ならびにアプリケーションおよびデバイスからのメタデータ捕捉（例えば、時間、場所、車両速度など）を含む、アプリケーションイベント。

5．事実の後にエージェントまたは他の注釈者によって追加された文字通りの注釈または構造化された注釈。

【0060】

このデータはすべて時間的に整列され、各イベントは、その開始時刻および終了時刻を伴う。訓練は、例えば、教師なし学習技法または強化学習技法を使用して、トレースデータを予測するようにシステムを構成する。一つの手法は、生成的確率モデルに基づく。

20

【0061】

図1に戻ると、訓練は、外部の統合セクション110のコンポーネントによって生成されたイベント118、119のシーケンス、および、それらのイベントにตอบสนองしてダイアログコントローラ130によって生成することができる結果として得られた正しい動作139であると考えることができる。意味パーサ121において構文解析をランク付けするためのパラメータまたはダイアログコントローラ130によって計画をランク付けするためのパラメータなどの構成可能なパラメータは、ダイアログ状態の進化の注釈またはイベントの処理の他の内部の態様を必ずしも必要とすることなく、イベントおよび動作に最も適合するように最適化される。いくつかの例では、イベントインタープリタ120およびダイアログコントローラのコンポーネントのパラメータを最適化することに加えて、イベントから生成されたイベントというよりむしろ、オリジナルの入力音声などのオリジナルの入力となるように動作とペアにされたイベントを考慮することによって、音声統合コンポーネント111（例えば、自動音声認識およびテキスト/音声合成）のパラメータも最適化することができる。

30

【0062】

一般に、訓練のための手法は、訓練の例の集合体である入力を使用することを伴う。各訓練の例は、ブレット付きのリストで説明されるイベントのシーケンスおよび任意のメタデータである。訓練の例の集合体に対応するカードライブラリも使用される。訓練予定のすべてのモデル（例えば、パーサ、エージェント/ダイアログマネージャ、テキスト生成、自動音声認識または合成）に対するパラメータは、例えばパラメータのランダムまたはゼロ数値ベクトルとして、最初に初期化される。いくつかの事例では、パラメータは、「いつも承認ばかりしない」または「パーサは文章内のほとんどの言葉を使用すべきである」などの態様を指定するなど、デフォルト条件で知らされる値に初期化されうる。

40

【0063】

次いで、パラメータの値の収束または反復回数の限度などの停止条件に至るまで、様々なコンポーネントのパラメータの決定が反復して実行される。各訓練の例（または1組の訓練の例の「ミニバッチ」）に対し、入力イベントは、現在のパラメータを使用して処理され、最良の重み付け実行追跡が選ばれる。その最良の追跡と関連付けられた各決定に対

50

し、現在のモデルパラメータを用いて選ばれた選択が最良の選択より劣っている場合は、パラメータは、学習レートおよび選ばれた選択と最良の選択との差に従って漸増する。

【0064】

最良の重み付け実行追跡の決定は、ビーム探索、粒子フィルタリング、 A^* 、または、現在のモデルパラメータを用いる他の関連技法を使用する。これは、その例に対する重み付け実行追跡の1組の集合体の出力をもたらす。重みは、各決定からおおよび各誤差関数演算からの集計スコアの総和である。追跡の各ステップは、システムが行わなければならない1組の「決定」（決定は、構文解析、計画、生成テキストなどを含む）を含む。各決定は、実際に選ばれた選択と共に、選択を行うことと関連付けられた使用される特徴を説明するベクトルの集合体である（これは、<http://www.aclweb.org/anthology/N10-1083>と同様である）。

10

【0065】

生成的確率モデル形成に基づく手法では、学習の目標は、入力において調節された出力の確率を最大化することである。使用するモデルの1つは、隠れマルコフモデル（HMM）であり、記録のタイムラインを時間ステップのシーケンスに分割する。各時間ステップでは、システムは、その時刻に（および任意選択によりその時刻になるまで）入力において調節されたイベントの時刻に起こる出力におけるイベントを生成する。期間にわたるイベントの場合は、ちょうど開始時刻または終了時刻にモデリングすること、イベントが起こるその時間帯に繰り返し起こるものとしてイベントをモデリングすることもできる（あるいは、隠れセミマルコフモデルを使用し、単一の時点というよりむしろ、イベントの開始時刻および終了時刻を予測することができる）。

20

【0066】

HMMの隠れバックボーンの場合は、システムの状態はまさしくダイアログマネージャの状態（すなわち、顕著な特徴リストおよびカード）である。HMMの移行モデルは、単に、ダイアログマネージャ自体である（すなわち、ユーザ発声に対する音声認識装置出力から発声を構文解析し、電話イベントに反応する）。HMMは、ダイアログマネージャの状態を使用して各時間ステップにおいてイベントを「生成する」。イベントタイプは、システムによって生成された出力、ダイアログ状態から生成された注釈およびアプリケーションイベントを含む。

【0067】

30

テキスト生成器は、発声にわたる分布を生み出し、モデルは、その分布の下で観測された発声の確率を最大化するように訓練される。その上、観測された発声の尤度の演算に実際に使用される分布は、生成器によって出力されたものと全く同じである必要はない。分布は、代わりに、テキスト生成器出力の関数であり得る。具体的には、尤度演算に使用される分布は、例えば「uhs」の挿入などを可能にする編集距離トランスデューサを使用することによって、吃音または音声認識アーチファクトに対してよりロバストになるようにすることができる。説明される実施形態では、自動的に推定されたパラメータを有する言葉にわたる編集距離トランスデューサが使用される（EMを介して推定される）。

【0068】

クリックストリームイベントを生成する分布の形態は、イベントタイプにわたるカード特有の多項分布のように簡単なものであり得る。成功する可能性が高い手法は、イベントタイプにわたる構造化された分布を使用し、例えば、「第1のEメールを未読にする」および「第3のEメールを削除する」のようなイベント間の共通性を学習することである。それに加えて、いくつかのイベントは、「ノイズ」分布によって（エージェントが誤ってストレイクリックするかまたはウェブページが何らかの自動バックグラウンドプロセスによって更新される（例えば、新しい広告が出現する））、最もうまく説明することができる。説明される実施形態では、訓練システムは、バックグラウンド「ノイズ」分布と動作および調節環境（すなわち、カード）に関する特徴を有する特徴豊富な記録の線形モデルとの混合分布を使用する。この分布に対するパラメータは、残りのシステムと共同で推定される。

40

50

【 0 0 6 9 】

事実上、システムのすべてのコンポーネントは、この手法を使用して訓練することができる。テキスト生成器は、記述およびエージェントの音声から学習する。ダイアログマネージャおよび意味パーサは、その分布が正しいクリックストリームイベントならびに正しい「発言」および「生成」イベントを生成するカードを予測することによって学習する。

【 0 0 7 0 】

上記で述べられるように、上記で説明される訓練手順の特性の1つは、記録されたデータの多くがヒューマンまたは混合ヒューマン/自動データ収集モードで自己注釈付けされることである。すなわち、ユーザの発声およびヒューマンエージェントの発声は、エージェントの動作の記録と共に、メタデータで増大される。この豊富なデータは、システム性能を向上させるため、自動化への新しい経路を見出すため、情報回収タスクをカスタマイズするため、および、ダイアログプロセスを分類するために、様々な機械学習プロセスへの入力として使用される。いくつかの例を以下に続ける。

【 0 0 7 1 】

ユーザからのオーディオデータは、音声認識システムで自動的に認識される。次いで、意味パーサで構文解析され、動作のためにエージェント（人間または機械にかかわらず）に送られる。動作は、データをユーザに送り返すこと、タスクを続行するために必要な情報のいくつかの部分の明確化を依頼すること、または、問題についてユーザに知らせることであり得る。

【 0 0 7 2 】

意味パーサと自動音声認識システムは両方とも、統計機械であり得る。すなわち、意味パーサおよび自動音声認識システムは、特定の成果の確率の演算の結果としての結果を生成し、結果は、それらの結果の正確性の確率と共に作成される。結果は、正しいものでも間違っただけのものでもあり得、このステータスは、情報を受信した際のユーザの動作によってまたはユーザからの受信音声に従って行動する際のヒューマンアシスタントの動作によって判断される。音声認識および意味構文解析の正確性は、より多くの「正しい」出力を作成し、かつ、「間違っただけ」出力を最小化するようにシステムを調整するために、機械学習アルゴリズムへの入力として使用することができる。ASR結果、意味構文解析出力、エージェントの動作および正確性測定はすべてユーザ入力文章の「注釈」である。

【 0 0 7 3 】

ユーザ発声は、入力チャネルとして使用されているデバイスからのメタデータを有する。携帯電話は、ジオロケーション計測、速度データ、加速度データ、温度測定、高度計測、周囲騒音測定、携帯電話アイデンティティ、電話番号、および、通話の間にネットワークに伝達される他のインジケータを有する。これらのデータは、発声のメタデータとして使用することができ、例えば、話し手の可能性のあるアクセントまたは言語は何か、ユーザがトラフィックの対応に忙しいかどうかまたは他の識別特性を示す。また、これらのデータは、ある変数の存在下で特定の応答の確率を調整することができる機械学習アルゴリズムへの生の入力として使用することもできる（例えば、銃の展示会についての質問は、ミシシッピ州よりもマサチューセッツ州の方が少ない）。

【 0 0 7 4 】

より粗いレベルでは、現在のシステムの「カード」は、タスクを完了するために行わなければならないすべてのものを定義する。カードは、他のサブ要素を参照する。例えば、「Eメールを書く」というカードは、宛先、差出人、件名およびテキストフィールドを定義するためのサブ要素を有する。機械学習アルゴリズムは、特定のユーザの場合、宛先の値がホワイトハウスであり、差出人がこの特定のユーザであり、件名が議決権であることや、音声認識装置が否定的な言葉よりむしろ肯定的な言葉寄りの姿勢であるべきであることを学習することができる。この事例では、サブカードの値は今のところ、カード自体のメタデータであり、カードを完成させるというシステムの性能を最適化するために使用することができる。

【 0 0 7 5 】

当然ながら、メタデータは、複合注釈として使用することもできる。音声認識テキスト材料はすべて、音声認識装置を最適化するために使用することができる。実際のテキストをより可能性が高いものにすることに加えて、そのような訓練は、未知の言葉（音声認識辞書にはないもの）を識別することおよび音声認識知識ベースに未知の言葉を追加することを試みることができる。

【0076】

ユーザの全人口にわたってカードと関連付けられた活動を追跡することにより、以前のステップを考慮して、完了すべき次のステップのより良い予測が可能になる。このタイプの予測は、よりふさわしいガイダンスをヒューマンエージェントに提供することができ、十分予想可能な場合は、以前はヒューマンアシスタントによって行われていたシステム活動の自動化を可能にすることができる。

10

【0077】

条件およびユーザ人口は変化するため、世界は進化し続けるため、ならびに、システム自体がユーザとアシスタントとの間の対話をどのように最適化するかを学習するため、注意深い補助システムによって提供される活動および注釈の豊富なセットは、時間の経過と共に変化する。各対話のデータの完全な記録により、特定のいかなるデータ要素も注釈として扱うことができる。

【0078】

図1に示されるシステムは、ユーザに通信補助を提供するためのものを含めて、様々な状況において使用することができる。例えば、外部の統合セクション110は、電話システムを統合するコンポーネントを含みうる。それにより、通話受信に関連するイベントは、受信した通話の処理に関連するユーザとのダイアログを開始することができる。システムの使用のための別の状況は、例えば、ユーザによるシステムの呼び出しや、電話システムに結合された音声統合コンポーネントを介してシステムによって処理される電話音声対話を伴う、電話注文または問い合わせへの自動電話応答におけるものである。多くの状況では、アプリケーション統合コンポーネントは、ユーザの代わりに情報を得るかまたは動作（例えば、ピザの注文）を引き起こすために、ウェブベースのサービスとインタフェースを取ることができる。

20

【0079】

システムの実装形態は、1つまたは複数のコンピュータを制御する指令（非一時的な機械可読媒体上に格納される）を含むソフトウェアを使用することができる。例えば、図1に示される機能は、単一のコンピュータ上で実行することも、例えばデータネットワーク上で通信する複数のコンピュータ上で分割して実行することもできる。いくつかの複数のコンピュータの実装形態では、あるコンポーネントは、コンピュータでまたはユーザの場所における他のコンピューティングデバイス（例えば、音声統合コンポーネント）でホストすることができる一方で、他のコンポーネントは、ユーザから離れた1つまたは複数の場所におけるサーバコンピュータ上でホストすることができる。上記で説明される訓練の機能は、さらなる他のコンピュータ上でホストすることができ、必ずしもユーザに関与するダイアログの実施に関与する必要はない。単一のユーザの環境下で説明されているが、複数のユーザおよび複数のダイアログを同時にサポートし、ダイアログを独立して効果的に操作するために適切な分離を維持するようにシステムを構成してもよい。いくつかのシステムは、単一のダイアログで、複数のユーザとの対話をサポートすることができる。

30

40

【0080】

前述の説明は、例示することを意図し、本発明の範囲を制限することを意図せず、本発明の範囲は、添付の請求項の範囲によって定義されることを理解されたい。他の実施形態は、以下の請求項の範囲内にある。

【図 1】

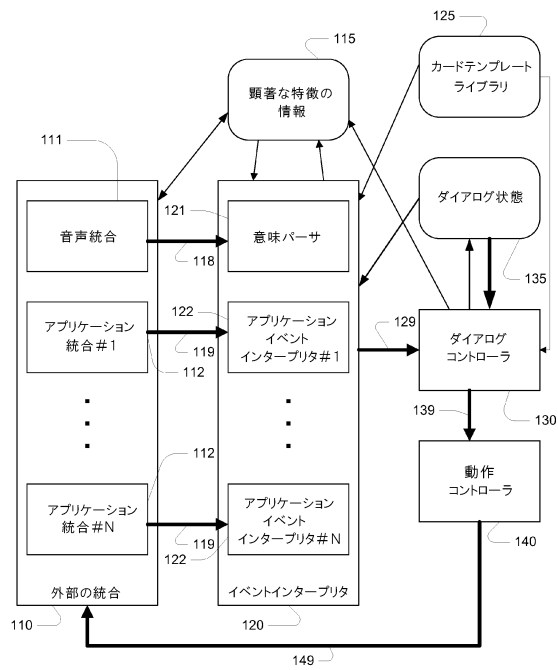


FIG. 1

【図 2】

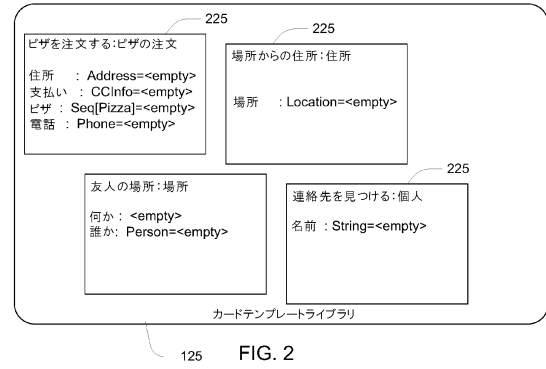


FIG. 2

【図 3】

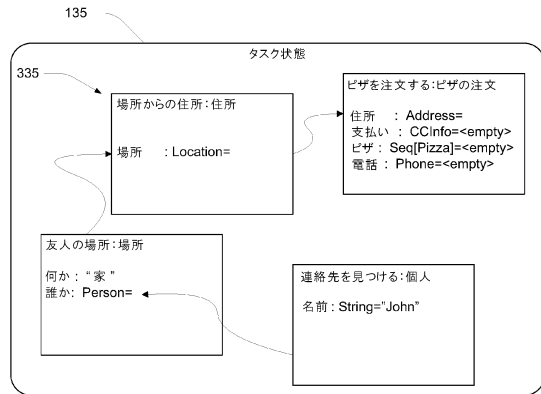


FIG. 3

フロントページの続き

- (74)代理人 100108213
弁理士 阿部 豊隆
- (72)発明者 アンドレアス, ジェイコブ
アメリカ合衆国 カリフォルニア州 94709 バークレー アーチ ストリート 1849
- (72)発明者 バークカークパトリック, テイラー ディー.
アメリカ合衆国 カリフォルニア州 94703 バークレー アpartment シー ドワイト
ウェイ 1713
- (72)発明者 チェン, ペンユ
アメリカ合衆国 カリフォルニア州 95014 クパチーノ 4 ピアードン ドライブ 10
228
- (72)発明者 コーエン, ジョーダン アール.
アメリカ合衆国 ノースカロライナ州 28449 クル ピーチ ノース 4番 アベニュー
209 エイ
- (72)発明者 ギリック, ローレンス エス.
アメリカ合衆国 マサチューセッツ州 02459 ニュートン エッツィ ファーム ロード
55
- (72)発明者 ホール, デビッド レオ ライト
アメリカ合衆国 カリフォルニア州 94710 バークレー 8番 ストリート 1904
- (72)発明者 クレイン, ダニエル
アメリカ合衆国 カリフォルニア州 94563 オリンダ フォーリン リーフ テラス 20
- (72)発明者 ニューマン, マイケル
アメリカ合衆国 マサチューセッツ州 02143 サマービル マウンテン アベニュー 12
- (72)発明者 ボールス, アダム ディー.
アメリカ合衆国 カリフォルニア州 94704 バークレー アpartment エイ ヘイスト
ストリート 1920
- (72)発明者 ロス, ダニエル エル.
アメリカ合衆国 マサチューセッツ州 02459 ニュートン マーシャル ストリート 16
- (72)発明者 ルサック, ジェシー ダニエル エスケス
アメリカ合衆国 マサチューセッツ州 02143 サマービル ユニット 1 カンブリア ス
トリート 26
- (72)発明者 ボルペ, アンドリュー ロバート
アメリカ合衆国 マサチューセッツ州 02210 ボストン ユニット 505 チャンネル
センター ストリート 35
- (72)発明者 ウェグマン, スティーブン エイ.
アメリカ合衆国 カリフォルニア州 94705 バークレー フロレンス ストリート 292
7

審査官 大野 弘

- (56)参考文献 特開平07-234694(JP,A)
特開平11-109991(JP,A)
特開2008-039928(JP,A)
特開2001-142484(JP,A)
特開2006-227611(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/22

G06F 3/16

G 0 6 F	4 0 / 2 0
G 0 6 F	4 0 / 3 0
G 1 0 L	1 5 / 1 0