

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6103766号
(P6103766)

(45) 発行日 平成29年3月29日 (2017.3.29)

(24) 登録日 平成29年3月10日 (2017.3.10)

(51) Int.Cl.	F I
G 0 6 F 17/30 (2006.01)	G 0 6 F 17/30 2 1 O A
	G 0 6 F 17/30 2 2 O Z

請求項の数 10 (全 19 頁)

(21) 出願番号	特願2013-146836 (P2013-146836)	(73) 特許権者	000004226
(22) 出願日	平成25年7月12日 (2013.7.12)		日本電信電話株式会社
(65) 公開番号	特開2014-241122 (P2014-241122A)		東京都千代田区大手町一丁目5番1号
(43) 公開日	平成26年12月25日 (2014.12.25)	(74) 代理人	100108855
審査請求日	平成27年6月18日 (2015.6.18)		弁理士 蔵田 昌俊
(31) 優先権主張番号	特願2013-105414 (P2013-105414)	(74) 代理人	100103034
(32) 優先日	平成25年5月17日 (2013.5.17)		弁理士 野河 信久
(33) 優先権主張国	日本国 (JP)	(74) 代理人	100075672
			弁理士 峰 隆司
		(74) 代理人	100179062
			弁理士 井上 正
		(72) 発明者	高橋 公海
			東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 行動プロセス抽出方法及び行動プロセス抽出装置

(57) 【特許請求の範囲】

【請求項 1】

プロセッサを備える行動プロセス抽出装置が実行する行動プロセス抽出方法であって、
前記行動プロセス抽出装置が、自然言語で記述されたデータ集合から、入力されたキーワードまたは前記データ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する第1のステップと、

前記行動プロセス抽出装置が、前記抽出した特徴語の集合と前記キーワードとの共起度を算出し、前記共起度に基づいて前記抽出した特徴語の集合の中から該当する特徴語を選択する第2のステップと、

前記行動プロセス抽出装置が、前記データ集合の中から、前記選択した特徴語に対する動作を表現する語を抽出することで前記キーワードに対する複数の行動プロセスを生成し、これらの行動プロセスが前記データ集合中に先に出現する確率に基づいて当該複数の行動プロセスの前後関係を決定する第3のステップと

を備えることを特徴とする行動プロセス抽出方法。

【請求項 2】

前記行動プロセス抽出装置が実行する前記第2のステップは、前記抽出した特徴語の集合と前記キーワードとの共起度を算出し、前記共起度が予め定めた閾値または統計に基づく算出値より低い特徴語をノイズとしてフィルタリングすることを特徴とする請求項1記載の行動プロセス抽出方法。

【請求項 3】

10

20

プロセッサを備える行動プロセス抽出装置が実行する行動プロセス抽出方法であって、
前記行動プロセス抽出装置が、自然言語で記述されたデータ集合から、入力されたキーワードまたは前記データ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する第1のステップと、

前記行動プロセス抽出装置が、前記データ集合の中から、各特徴語に対する動作を表現する語を抽出する第2のステップと、

前記行動プロセス抽出装置が、前記第2のステップで抽出された動作と特徴語とからなる行動の組について、前記データ集合の中で先に出現する確率を算出し、算出結果に基づいて行動間の前後関係を決定することにより行動プロセスを抽出する第3のステップと
を備えることを特徴とする行動プロセス抽出方法。

10

【請求項4】

プロセッサを備える行動プロセス抽出装置が実行する行動プロセス抽出方法であって、
前記行動プロセス抽出装置が、自然言語で記述されたデータ集合から、入力されたキーワードまたは前記データ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する第1のステップと、

前記行動プロセス抽出装置が、前記抽出した特徴語の集合と前記キーワードとの共起度を算出し、前記共起度に基づいて前記抽出した特徴語の集合の中から該当する特徴語を選択する第2のステップと、

前記行動プロセス抽出装置が、前記データ集合の中から、前記選択した特徴語に対する動作を表現する語を抽出する第3のステップと、

20

前記行動プロセス抽出装置が、前記第3のステップで抽出された動作と特徴語とからなる行動の組について、前記データ集合の中で先に出現する確率を算出し、算出結果に基づいて行動間の前後関係を決定することにより行動プロセスを抽出する第4のステップと
を備えることを特徴とする行動プロセス抽出方法。

【請求項5】

前記行動プロセス抽出装置が実行する前記第2のステップは、前記抽出した特徴語の集合と前記キーワードとの共起度を算出し、前記共起度が予め定めた閾値または統計に基づく算出値より低い特徴語をノイズとしてフィルタリングすることを特徴とする請求項4記載の行動プロセス抽出方法。

【請求項6】

30

自然言語で記述されたデータ集合から、入力されたキーワードまたは前記データ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する特徴語抽出手段と、

前記抽出した特徴語の集合と前記キーワードとの共起度を算出し、前記共起度に基づいて前記抽出した特徴語の集合の中から該当する特徴語を選択する特徴語選択手段と、

前記データ集合の中から、前記選択した特徴語に対する動作を表現する語を抽出することと前記キーワードに対する複数の行動プロセスを生成し、これらの行動プロセスが前記データ集合中に先に出現する確率に基づいて当該複数の行動プロセスの前後関係を決定する行動プロセス生成手段と

を備えることを特徴とする行動プロセス抽出装置。

40

【請求項7】

自然言語で記述されたデータ集合から、入力されたキーワードまたは前記データ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する特徴語抽出手段と、

前記データ集合の中から、各特徴語に対する動作を表現する語を抽出する動作抽出手段と、

前記動作抽出手段で抽出された動作と特徴語とからなる行動の組について、前記データ集合の中で先に出現する確率を算出し、算出結果に基づいて行動間の前後関係を決定することにより行動プロセスを抽出する前後関係算出手段とを備えることを特徴とする行動プロセス抽出装置。

50

【請求項 8】

自然言語で記述されたデータ集合から、入力されたキーワードまたは前記データ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する特徴語抽出手段と、

前記抽出した特徴語の集合と前記キーワードとの共起度を算出し、前記共起度に基づいて前記抽出した特徴語の集合の中から該当する特徴語を選択する特徴語選択手段と、

前記データ集合の中から、前記選択した特徴語に対する動作を表現する語を抽出する動作抽出手段と、

前記動作抽出手段で抽出された動作と特徴語とからなる行動の組について、前記データ集合の中で先に出現する確率を算出し、算出結果に基づいて行動間の前後関係を決定することにより行動プロセスを抽出する前後関係算出手段とを備えることを特徴とする行動プロセス抽出装置。

10

【請求項 9】

前記行動プロセス抽出装置が、前記抽出される行動プロセスについて、前記データ集合中で起こり得る確率を算出し、前記起こり得る確率が予め定めた閾値または統計に基づく算出値より低い行動プロセスをノイズとして除去するステップを含むことを特徴とする請求項 1 から 5 のいずれか 1 項に記載の行動プロセス抽出方法。

【請求項 10】

前記抽出される行動プロセスについて、前記データ集合中で起こり得る確率を算出し、前記起こり得る確率が予め定めた閾値または統計に基づく算出値より低い行動プロセスをノイズとして除去するフィルタリング手段を備えることを特徴とする請求項 6 から 8 のいずれか 1 項に記載の行動プロセス抽出装置。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、実世界における人間の行動プロセスを自動的に抽出する技術であり、高度な検索・推薦に応用可能な行動プロセス抽出方法及び行動プロセス抽出装置に関する。

【背景技術】

【0002】

実世界における人間の行動をモデル化することは、様々な有用なアプリケーションを可能にする技術であると長い間考えられてきた。特に、人間の行動や出来事の繋がり（行動プロセス）をモデル化することにより、人間の行動や思考の予測が可能となり、行動提示や発話生成といった応用が見込まれる。従来の技術は大きく分けて 3 つあり、それぞれの概要と課題について以下で述べる。

30

【0003】

（a1）状況を限定し人手で行動プロセスをモデル化する技術

予めいくつかの状況を想定し、人手で行動プロセスをモデル化した技術としてタスクオントロジ（非特許文献 1）が挙げられる。タスクオントロジは、ユーザが実世界で認識する問題（タスク）に対して、どのような行動プロセスで解決していくかを示した知識ベースである。「移動」「食事」「遊ぶ」「買い物」「宿泊」「旅行」という 6 つの状況において、起こり得る問題を洗い出し、それぞれの問題を解決する行動プロセスを人手で記述している。

40

【0004】

しかし、タスクオントロジでは、状況や記述されている行動プロセスが非常に限定的であり、かつオントロジの構築には多大な労力を要するという 2 つの課題がある。

【0005】

（a2）限定的な状況における行動プロセスを自動的に抽出する技術

料理レシピサイトのように、行動のプロセスが順に追って明確に書かれている形式化された文書を対象とし、大量の行動プロセスを自動的に生成する技術として Perkowitzm らの手法（非特許文献 2）が挙げられる。しかし、シンプルなテキスト解析で行動プロセスを

50

生成しているための、「step1, step2…」といったように、形式的に書かれた料理レシピなどの文書にした対応しておらず、抽出可能な行動プロセスは限定的である。

【0006】

(a3) 多様な状況における行動プロセス抽出技術

大量かつ多様な人間の行動プロセスを抽出するためには、形式化された文書だけでなく、個人が実世界においてどのような行動をとったかが記述された非構造的な文書（blogやTwitter、質問応答サイトなど）から行動プロセスを抽出する技術が必要である。アプローチとしては、Nguyenらのように言語的な手がかりを利用した手法（非特許文献3）や、倉島らの相関の高いイベントや行動・主観をマイニングする技術（非特許文献4）が挙げられる。

10

【0007】

Nguyenらは、条件付き確率場と自己教師あり学習を用いて、文に現れる行動を構成する語をラベル付けし、1つの文中における行動間の遷移を精度良く抽出する手法を提案している。しかし、この手法では、1つの文中に「の前に」「の後で」といった行動間の遷移を示す言語的な手がかりが存在する場合しか行動のプロセスを抽出することができない。また、どのような状況で成立する行動プロセスなのかも意識されていない。

【0008】

倉島らは、文中に言語的な手がかりとなる語が存在しない場合でも、例えば「ディズニーランド」は「パレードを見る」という行動や、「喜び」という主観と相関が高いという知識を抽出する手法を提案している。しかし、Nguyenらと同様に、どのような状況で成立する知識なのかは意識されておらず、前後関係など行動間の関連性も考慮されていない。

20

【先行技術文献】

【非特許文献】

【0009】

【非特許文献1】笹嶋 宗彦，古谷 孝一郎，來村 徳信，深澤 佑介，長沼 武史，倉掛 正治，溝口 理一郎，“実規模モバイルサービス向けタスク指向型メニューの開発と評価” 人工知能学会全国大会，2009.

【非特許文献2】Perkowitz M., Philipose, M., Fishkin, K., Patterson, D.J., “Mining models of human activities from the web” Proc. 13th international conf. on World Wide Web, pp.573-582, New York, USA, May 2004.

30

【非特許文献3】グエンミンテイ，川村隆浩，中川博之，田原康之，大須賀昭彦，“Webからの自己教師あり学習を用いた人間行動マイニング” 電子情報通信学会人工知能と知識処理研究会，AI2009-22，2009.

【非特許文献4】倉島健，藤村考，奥田英範，“大規模テキストからの経験マイニング” 第19回データ工学ワークショップ（DEWS2008），2008.

【非特許文献5】高橋 公海，佐藤 進也，松尾 真人，“Webからの効率的な行動プロセス抽出方法の検討” 第4回データ工学と情報マネジメントに関するフォーラム（DEIM2012），2012.

【非特許文献6】Blei, D.M., Ng, A.Y. and Jordan, M.I., “Latent Dirichlet Allocation” Journal of Machine Learning Research 3, pp.993-1022, 2003.

40

【非特許文献7】Cilibrasi, R. L. and Vitanyi, P. M. B. “The Google Similarity Distance” IEEE Transactions on Knowledge and Data Engineering, 19(3), 370-383, 2007.

【非特許文献8】北研二，津田和彦，獅々堀正幹：情報検索アルゴリズム，共立出版（2002）.

【発明の概要】

【発明が解決しようとする課題】

【0010】

抽出した行動プロセスを人間への行動提示などに応用する場合、状況によって取り得る行動が変わるため、提示する対象の人間がどのような状況下にあるかを考慮に入れること

50

が望ましい。上記（a1）、（a2）、（a3）の技術はいずれも、抽出した知識がどのような状況で成立するものかを考慮していない。

【0011】

本発明の目的は、状況を意識した行動プロセスの獲得を図り得る行動プロセス抽出方法及び行動プロセス抽出装置を提供することにある。

【課題を解決するための手段】

【0012】

上記目的を達成するために本発明に係る行動プロセス抽出方法は、プロセッサを備える行動プロセス抽出装置が、自然言語で記述されたデータ集合から、入力されたキーワードまたはデータ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する第1のステップと、前記プロセッサを備える行動プロセス抽出装置が、抽出した特徴語の集合とキーワードとの共起度を算出し、共起度に基づいて抽出した特徴語の集合の中から該当する特徴語を選択する第2のステップと、前記プロセッサを備える行動プロセス抽出装置が、データ集合の中から、選択した特徴語に対する動作を表現する語を抽出することで前記キーワードに対する複数の行動プロセスを生成し、これらの行動プロセスが前記データ集合中に先に出現する確率に基づいて当該複数の行動プロセスの前後関係を決定する第3のステップとを備えるようにしたものである。

10

【0013】

このように構成すると、特徴語の集合で表される状況自体を文書集合から自動的に抽出し、さらにノイズとなる特徴語をフィルタリングすることにより、予め状況を限定せずとも、状況に応じた行動プロセスを出力することができる。

20

【0014】

また、本発明に係る行動プロセス抽出方法は、プロセッサを備える行動プロセス抽出装置が、自然言語で記述されたデータ集合から、入力されたキーワードまたはデータ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する第1のステップと、前記プロセッサを備える行動プロセス抽出装置が、データ集合の中から、各特徴語に対する動作を表現する語を抽出する第2のステップと、前記プロセッサを備える行動プロセス抽出装置が、第2のステップで抽出された動作と特徴語とからなる行動の組について、データ集合の中で先に出現する確率を算出し、算出結果に基づいて行動間の前後関係を決定することにより行動プロセスを抽出する第3のステップとを備えるようにしたものである。

30

【0015】

このように構成すると、データ集合を統計的に解析し、どちらの行動が先に出現する確率が高いかを算出することで、形成的に書かれていない文書や文中に言語的な手がかりが存在しない場合でも、行動間に順序関係のある行動プロセスを出力することが可能となる。

【0016】

さらに、本発明に係る行動プロセス抽出方法は、プロセッサを備える行動プロセス抽出装置が、自然言語で記述されたデータ集合から、入力されたキーワードまたはデータ集合に含まれるキーワードについて、状況の特徴付ける特徴語の集合を抽出する第1のステップと、前記プロセッサを備える行動プロセス抽出装置が、抽出した特徴語の集合とキーワードとの共起度を算出し、共起度に基づいて抽出した特徴語の集合の中から該当する特徴語を選択する第2のステップと、前記プロセッサを備える行動プロセス抽出装置が、データ集合の中から、選択した特徴語に対する動作を表現する語を抽出する第3のステップと、前記プロセッサを備える行動プロセス抽出装置が、第3のステップで抽出された動作と特徴語とからなる行動の組について、データ集合の中で先に出現する確率を算出し、算出結果に基づいて行動間の前後関係を決定することにより行動プロセスを抽出する第4のステップとを備えるようにしたものである。

40

【0017】

このように構成すると、非構造な文書集合から、状況の特徴付ける特徴語の集合のうち

50

ノイズがない信頼度の高い特徴語と、各特徴語に対する動作を示す語を抽出することで行動を構成し、文書集合から行動間の前後関係を決定することにより、多様な状況における自動的な行動プロセス抽出を実現でき、さらに行動プロセス抽出に関する信頼性を向上できる。

【0018】

さらに、本発明に係る行動プロセス抽出方法は、プロセッサを備える行動プロセス抽出装置が、抽出される行動プロセスについて、データ集合中で起こり得る確率を算出し、起こり得る確率が予め定めた閾値または統計に基づく算出値より低い行動プロセスをノイズとして除去するステップを含むようにしたものである。

【0019】

このように構成すると、抽出される行動プロセスについて、データ集合中で起こり得る確率を算出し、起こり得る確率の低い行動プロセスをノイズとしてフィルタリングすることにより、行動プロセス抽出精度の向上を図ることができる。

【発明の効果】

【0020】

本発明によれば、従来技術では対象としていないテキストからも行動プロセス抽出を行うことが可能となるため、大量かつ多様な人間の行動プロセスを生成することができる。また、入力する文書集合を上手く限定することにより、特定のセグメント（場所・年代・各個人など）に特化した行動プロセスを抽出することも可能である。

【図面の簡単な説明】

【0021】

【図1】本発明の第1の実施形態に係る行動プロセス抽出方法を実現する装置の機能ブロック図。

【図2】同第1の実施形態において、文書集合からの特徴語抽出部の動作を示すフローチャート。

【図3】同第1の実施形態において、特徴語集合から行動プロセスを生成する行動プロセス生成部の動作を示すフローチャート。

【図4】本発明の第2の実施形態に係る行動プロセス抽出方法を実現する装置の機能ブロック図。

【図5】同第2の実施形態において、フィルタリング機能の動作を示すフローチャート。

【図6】同第2の実施形態において、縦軸に割合、横軸に行動の出現回数をとった、行動の出現頻度分布図。

【発明を実施するための形態】

【0022】

以下、本発明の実施形態について図面を参照して詳細に説明する。

（第1の実施形態）

図1は、本発明の第1の実施形態に係る行動プロセス抽出方法を実現する装置の機能ブロック図である。図1において、状況の特徴付ける特徴語抽出部10及び行動プロセス生成部20は、例えば、CPU、メモリ、ハードディスクを備える計算機に備えられる。また、図1において、入力部31及び出力部32が備えられる。入力部31は、ユーザによるキーワードWの入力を受け付け、またweb（図示せず）から送られるデータを入力するためのものである。出力部32は、特徴語抽出部10及び行動プロセス生成部20による処理結果を表示し、さらに特徴語抽出部10及び行動プロセス生成部20による処理結果を出力するためのものである。なお、実施形態では、テキストファイルやDBに格納する形で出力しているが、出力はテキストやDBに限定されるものではない。

【0023】

状況の特徴付ける特徴語抽出部10は、特徴語抽出機能11とノイズとなる語をフィルタリングする特徴語選択機能12から構成される。行動プロセス生成部20は、特徴語に対する動作を抽出する動作抽出機能21と、行動（特徴語と動作の組）間の前後関係を算出し、どちらの行動が先に出現する確率が高いかを判定する前後関係算出機能22とから

10

20

30

40

50

構成される。

【0024】

入力部31では、自然言語で記述された文書集合が入力される。文書集合を絞り込む任意のキーワードWの入力は必須ではないが、キーワードが入力された場合には、そのキーワードに関連する状況における行動プロセスを抽出することが可能である。任意のキーワードWが入力されない場合には、文書集合に含まれる全ての名詞をキーワードと考え、同様に全てのキーワードに関連する状況の行動プロセスを抽出する。なお、本発明では、入力された任意のキーワードまたはデータ集合中に含まれる全てのキーワードや、全ての動作と特徴語とからなる行動の組を対象とすることが可能であるが、全てを対象としなくともよい。

10

【0025】

例えば入力される任意のキーワードWが「就職活動」であった場合に、「新卒採用」「転職」「天下り」「服装選び」といった状況ごとに、それぞれの状況の特徴付ける特徴語を得る。「新卒採用」という状況の特徴付ける語の集合としては、「面接、採用、書類、内定、エントリー」といった語が得られる。

【0026】

状況に応じた特徴語を得る手法として、従来技術（非特許文献5：高橋 公海，佐藤 進也，松尾 真人，“Webからの効率的な行動プロセス抽出方法の検討”第4回データ工学と情報マネジメントに関するフォーラム（DEIM2012），2012.）が存在するが、一見関連性のない特徴語も得られるため、ノイズが多くそのまま適用することはできない。そこで、キーワードとの共起度を検索エンジン等を利用して算出することで、ノイズを減らすことが可能である。これが特徴語選択機能12である。

20

【0027】

行動プロセス生成部20の動作抽出機能21では、各特徴語に対する動詞を抽出する。例えば、「面接」という特徴語であれば「受ける」という動詞を抽出し、特徴語と組み合わせ「面接を受ける」という行動を得る。

【0028】

前後関係算出機能22では、全ての行動間の前後関係を算出する。「面接を受ける」と「エントリーをする」という2つの行動について、文書集合中でどちらの行動が先に出現する確率が高いかを算出し、「エントリーをする → 面接を受ける」という行動プロセスを生成する。対象（特徴語）と動作の組を行動とし、それらを並べることで行動プロセスを表現する形式自体は、従来と同様である。

30

【0029】

なお、前後関係算出機能22では、例えば行動Aと行動Bの前後関係を判定する際に、行動Aが先に出現する可能性が50%、行動Bも同じく50%の場合、「同じ状況で発生しうる行動の組だが、前後関係は無い」ものとして扱う。この場合も、これらの行動の組はチェックリストとして利用可能である。

【0030】

第1の実施形態として、文書集合にweb上のblog記事、前後関係の特定に各blog記事の投稿時刻を利用し、行動プロセスを抽出する例について示す。

40

【0031】

(b1) 文書集合からの特徴語抽出

図2は、文書集合からの特徴語抽出部10の動作を示すフローチャートである。まず、キーワードWが入力されると（ステップST2a）、特徴語抽出部10は、解析対象のデータを読み込み（ステップST2b）、本文に任意のキーワードWを含む記事のみを解析対象として絞り込む（ステップST2c）。また、特徴語抽出部10は、blog記事を書いた著者のidと本文、投稿時刻を取得する（ステップST2d）。

【0032】

次に、特徴語抽出部10は、本文からbag-of-words（以降、BOWと称する）ファイルを作成し（ステップST2e）、BOWから特徴語抽出を行う（ステップST2f）。B

50

OWとは、単語の集合のことであり、本実施形態では本文を形態素解析し、名詞（形式名詞は除く）と未知語を特徴語の候補として抽出しBOWを作成する。例えば、「風邪なので薬を飲んだ」という文からは、「風邪」「薬」という単語が特徴語の候補として抽出され、それらの単語の集合がBOWファイルに記述される。

【0033】

BOWからの特徴語抽出手法として、トピックモデル（非特許文献6：Blei, D.M., Ng, A.Y. and Jordan, M.I., “Latent Dirichlet Allocation” Journal of Machine Learning Research 3, pp.993-1022, 2003.）を利用する。これは、各トピックが状況に対応しており、それぞれの状況の特徴付ける語を得られるためである。ただし、状況の特徴付ける語を得られるのであれば、LSIなどの別の手法でもよい。例えば、任意のキーワードとして「結婚式」を入力し、キーワードを含む文書の本文からBOWを作成しトピックモデルで特徴語抽出を行った場合、表1のような結果が得られる。

【表1】

トピック 番号	特徴語
0	ドレス、服装、ショール、ネックレス、白、黒、靴、バッグ、ワンピース、マナー
1	気持ち、前、時、自分、今、気、後、何か、一番、時間
2	事、時、人、後、今回、中、事情、頭、現在、訳
3	ご祝儀、額、会費、夫婦、披露宴、相場、式、品物、会費制
4	自分、今、話、先日、人、結果、その日、人間、内容、去年

【0034】

例えば、トピック番号0は、結婚式に参加するため服装や小物を準備している状況、トピック番号3は披露宴のご祝儀を用意している状況に対応した特徴語集合が得られている。一方で、トピック番号1, 2, 4のように、特徴語を一見しただけではどのような状況か分からないトピックも存在している。特徴語集合に含まれる各語を見ると、どのような状況か分からないトピックには、キーワードとの関連性が低い語が多く含まれる傾向がある。そこで、特徴語選択機能12では、キーワードと各特徴語との共起度を算出し（ステップST2g）、それらの平均値を各トピックの共起度とする。共起度の値が閾値以上のトピックを残すことにより、特徴語の選択を行う。

【0035】

フィルタリングの閾値は、経験的に予め定めた閾値または統計に基づく算出値により決定する。

【0036】

本第1の実施形態では経験的な知見から、各トピックを共起度の高い順に並べ、上位20%にあたるトピックの共起度を閾値として用いた。

【0037】

しかしこれに限らず、統計値に基づく算出値、例えば、平均値、中央値、最頻値、指数平均値、平均値から標準偏差の数倍以上の外れ値（極端な値）を除いて処理する調整平均、移動平均、平均値から標準偏差の数倍より低い値、その他過去の事例から得られた予め定めた値等を用いてもよい。

【0038】

キーワードとの関連性を測る方法としては、Googleのヒット数を用いて意味的な関わりの度合いを測る手法（非特許文献7：Cilibrasi, R. L. and Vitanyi, P. M. B. “The Google Similarity Distance” IEEE Transactions on Knowledge and Data Engineering, 19(3), 370-383, 2007.）を利用する。関連性の高いトピック順に並べると、表2のような結果が得られる（ステップST2h）。

【表 2】

トピック 番号	特徴語
55	披露宴、親族、式、雰囲気、出席者、食事、挙式、余興、入場、写真
37	衣装、色、ショップ、白無垢、人前式、 タキシード、花嫁、店員、白、ドレス
0	ドレス、服装、ショール、ネックレス、白、 黒、靴、バッグ、ワンピース、マナー
3	ご祝儀、額、会費、夫婦、披露宴、相場、式、品物、会費制
81	二次会、幹事、会費、ゲーム、ビンゴ、 人形、景品、参加者、携帯、会場

10

【0039】

(b2) 行動プロセスの生成

図3は、特徴語集合から行動プロセスを生成する行動プロセス生成部20の動作を示すフローチャートである。

【0040】

まず、行動プロセス生成部20は、上記特徴語抽出部10で得られる特徴語集合が入力されると(ステップST3a)、一時変数*i*を初期化し(ステップST3b)、*i*番目のトピックがあるか否かの判断を行う(ステップST3c)。ここでは、トピックが存在するものとし(存在)、行動プロセス生成部20は、0番目のトピックの特徴語を読み込み(ステップST3d)、そして、0番目のトピックにおける特徴語集合とキーワードを含む記事を読み込み(ステップST3e)、動作抽出機能21で特徴語に対する動作を抽出する(ステップST3f)。

20

【0041】

ここでは、係り受け解析を行い、文中で特徴語に係る動作を抽出するが、文中で特徴語に係る動作を抽出するが、文中に共起する動作と組み合わせるなど別の方法で抽出してもよい。例えば、トピック番号0について、「青いドレスを着て、ショールをはおった。」という文から動作を抽出する場合、「ドレス」に対して「着る」、「ショール」に対して「はおる」という動作をそれぞれ得る。

30

【0042】

次に、行動プロセス生成部20は、特徴語と動作からなる全ての行動の組について、本文中にどちらが先に出現する確率が高いかを前後関係算出機能22により算出する(ステップST3g)。各記事内、さらに記事を書いたユーザ毎投稿順に並べた場合に、どちらの行動が先に出現しているかをカウントし、行動間の前後関係を決定する。

【0043】

行動間の前後関係を決定した場合、行動プロセス生成部20は、ステップST3a及びステップST3bの処理を実行する。そして、ステップST3cにおいて、トピックが存在しない場合(不在)、行動プロセス生成部20は行動プロセスの集合を出力し、ユーザに提示する(ステップST3i)。

40

【0044】

生成される行動プロセスとしては、例えばトピック番号0については「ドレスを着る → ショールをはおる」「ネックレスをつける → ショールをはおる」、トピック番号3については「式に呼ばれる → ご祝儀を渡す」「額を減らす → ご祝儀を渡す」、トピック番号37については「ドレスを見る → 衣装を借りる」「ドレスを選ぶ → ドレスを着る」「人前式を行う → 白無垢が似合う」といったものが挙げられる。

【0045】

ここでは、2つの行動間の前後関係を算出し出力するが、3つ以上の行動を含むプロセスを抽出する方法としては、推移律を利用して「行動A → 行動Bかつ行動B → 行動C」な

50

らば「行動A→行動C」が成り立つため、「行動A→行動B→行動C」という行動プロセスを抽出することも可能である。また、prefixspanなどの系列抽出手法を利用してもよい。

【0046】

(b3) 行動プロセスを利用したサービス例

行動プロセスを示すグラフを利用することにより、人が今度とる行動や思考を推定することや、行動プロセスを利用したナビゲーションや失敗の回避、マニュアルの自動作成、行動拡張や新しい方法の発見等のサービスに応用可能である。

【0047】

例えば、「風邪」という状況で「熱がある → 会社に電話 → 病院に行く → 薬をもらう」という一連の行動プロセスがあった時、センサ等の何等かの方法で平熱よりも体温が高い状態を検知できれば、プロセスを用いて、その後「会社に電話」「病院に行く」といった行動をとることは推測できる。また、旅行の感想などが書かれた文書を入力すると、観光ルートが行動プロセスとして出力されるため、観光のナビゲーションにも利用することができる。さらに、シーケンシャルなパターンではなく分岐が存在するグラフとなった行動プロセスを提示することで、別の選択肢や新しい方法を発見することも可能である。

10

【0048】

以上のように上記第1の実施形態によれば、特徴語抽出部10に特徴語抽出機能11の他に、特徴語選択機能12を備えるようにしているので、特徴語の集合で表される状況自体を例えばweb上のサイトに存在する文書集合から自動的に抽出し、さらにノイズとなる特徴語をフィルタリングすることができ、これにより予め状況を限定せずとも、状況に応じた行動プロセスを出力することができる。

20

【0049】

また、上記第1の実施形態によれば、行動プロセス生成部20に動作抽出機能21の他に、前後関係算出機能22を備えるようにしているので、例えばweb上から収集した文書集合を統計的に解析し、どちらの行動が先に出現する確率が高いかを算出することができ、これにより形成的に書かれていない文書や文中に言語的な手がかりが存在しない場合でも、行動間に順序関係のある行動プロセスを出力することが可能となる。

【0050】

さらに、上記第1の実施形態によれば、blogやTwitterなどから収集した非構造的な文書集合から、状況の特徴付ける特徴語の集合のうちノイズがない信頼度の高い特徴語と、各特徴語に対する動作を示す語を抽出することで行動を構成し、文書集合から行動間の前後関係を決定することにより、多様な状況における自動的な行動プロセス抽出を実現でき、さらに行動プロセス抽出に関する信頼性を向上できる。

30

【0051】

また、上記第1の実施形態であれば、blog記事を書いた著者のidと本文、投稿時刻を利用すれば、特定のセグメント(場所・年代・各個人など)に特化した行動プロセスを抽出することも可能である。

【0052】

(第2の実施形態)

先の第1の実施形態では、状況と関連性の無い行動プロセスも多数出力されることがある。

40

【0053】

例えば、キーワード「かぼちゃ」に関連する状況として「冬至」や「ハロウィン」「離乳食を作る」といった状況が挙げられる「冬至」に関連する特徴語集合を手掛かりに行動プロセス集合を生成すると、表3のような結果が得られる。

【表 3】

行動プロセス集合
かぼちゃを食べる→ゆず湯に入る、何かを忘れる→ゆず湯に入る、 かぼちゃを食べる→風邪を予防、風邪をひく→かぼちゃを食べる、 冬至という→自分で決める、風邪を引く→地平線に一致、 野菜が途切れる→風邪をひく、花を紹介する→かぼちゃを食べる、 お風呂に入る→ゆずを入れる 等

10

【0054】

プロセスと、「風邪を引く→地平線に一致」のように、連続して発生することが滅多にない行動プロセスとが混在して出力される。実際には、出力される行動プロセス集合のうち、状況と関連性の高い行動プロセスは極僅かしか存在せず、連続して起こることが殆どない行動プロセスが多数を占める。

【0055】

そこで、第2の実施形態では、連続して発生する確率が低い行動プロセスをフィルタリングし、状況の特徴付けるプロセスに重み付けすることでノイズを低減させるようにしている。

20

【0056】

図4は、本発明の第2の実施形態に係る行動プロセス抽出方法を実現する装置の機能ブロック図である。図4において、上記図1と同一部分には、同一符号を付して、詳細な説明を省略する。

【0057】

第2の実施形態では、上記出力部32の前段に、フィルタリング機能41を追加する。フィルタリング機能41は、例えば、上記特徴語抽出部10及び行動プロセス生成部20と同様に、CPU、メモリ、ハードディスクを備える計算機に備えられ、上記前後関係算出機能22で抽出された複数の行動プロセスそれぞれについて、文書集合中で起こり得る確率を算出し、起こり得る確率が予め定めた閾値または統計に基づく算出値より低い行動

30

【0058】

特徴語抽出部10及び行動プロセス生成部20は、先の第1の実施形態と同様であるが、本第2の実施形態は、フィルタリング機能41で行動プロセスが起こり得る確率を算出し、確率が低いものを除去することで行動プロセス抽出精度向上が可能となる。

図5は、フィルタリング機能41の動作を示すフローチャートである。

【0059】

まず、フィルタリング機能41は、上記前後関係算出機能22で得られる行動プロセス集合が入力されると(ステップST5a)、一時変数*i*を初期化し(ステップST5b)、*i*番目のトピックがあるか否かの判断を行う(ステップST5c)。ここでは、トピックが存在するものとし(存在)、フィルタリング機能41は、0番目のトピックの特徴語を読み込み(ステップST5d)、そして、0番目のトピックにおける特徴語集合を含む記事を読み込み(ステップST5e)、行動プロセスの特徴量を算出する(ステップST5f)。

40

【0060】

行動プロセスの特徴量を算出するステップでは、文書集合中の行動プロセスの頻度などの特徴を用いて、各行動プロセスについて起こり得る確率を算出する。文書集合中の行動プロセスの頻度を算出する方法としては、例えば文書集合中においてある行動プロセスが出現する数*n*を、文書集合中の全ての行動プロセス数*m*で除したもの(n/m)が挙げられる。本第2の実施形態では、行動プロセスの頻度以外に重要語句を抽出する技術である

50

残差 idf[非特許文献 8：北研二，津田和彦，獅々堀正幹：情報検索アルゴリズム，共立出版（2002）.]を利用しているが、検索エンジンのヒット数を用いて行動とキーワードとの関連度合いを測る手法[非特許文献 7：Cilibrasi, R. L. and Vitanyi, P. M. B. “The Google Similarity Distance” IEEE Transactions on Knowledge and Data Engineering, 19(3), 370-383, 2007.]で算出された数値なども特徴量として有効である。なお、特徴量や行動プロセスが起こり得る確率を数値化する方法は上記に限定されず、新たな統計理論等を用いて適宜修正することができる。

【0061】

フィルタリング機能 41 は、上記ステップ S T 5 f で算出された行動プロセスと特徴量を記憶媒体に格納し（ステップ S T 5 g）、次のトピックを設定し（ステップ S T 5 h）、全ての行動プロセスについて各特徴量を算出し記憶媒体に記憶する。

10

【0062】

そして、ステップ S T 5 c において、トピックが存在しない場合（不在）、フィルタリング機能 41 は記憶媒体に格納された行動プロセス集合から閾値以上のものを出力し（ステップ S T 5 i）、記憶媒体を空にし（ステップ S T 5 j）、閾値以上の行動プロセスの集合をユーザに提示する（ステップ S T 5 k）。

【0063】

ここで、第 2 の実施形態の一例として、Web から収集した blog 記事集合を対象に、先の第 1 の実施形態で行動プロセス抽出を行い、フィルタリング機能 41 によりノイズを低減した例を示す。

20

【0064】

実際には例えば「かぼちゃ」を含む blog 記事約 2000 件を対象とした場合でも、取り得る全ての行動プロセスは 70 万以上、そのうち「冬至」のトピックに関連する文書集合から得られる行動プロセスは 2 万程度存在する。

【0065】

本第 2 の実施形態では、「行動の組 $\langle X, Y \rangle$ があったとき、ある状況において X が起きた際、それに続いて Y が生じる確率が高い行動の組」の抽出を目的としている。このため、次の 2 つの仮定をもとに行動パターンに重み付けを行う。

【0066】

c 1. 行動の組 $\langle X, Y \rangle$ が文書集合中の多くの文書で $X \rightarrow Y$ という順に出現している場合、X に続いて Y が生じる確率が高い。

30

c 2. 文書集合全体よりもクラスタ内の文書集合に偏って出現する行動パターンは、クラスタが対応する状況に依存している。

【0067】

仮定（c 1）を基に、文書集合全体における行動パターンの文書頻度（Document Frequency: df）を指標として利用する。行動は文中から係り受け関係にある名詞・格助詞・動詞の 3 つを自動的に抽出しているが、ブログのような非構造のテキストから記述される文章は文法的に正しくない文も多いため、日本語として不自然な行動も抽出されてしまう。例えば、「お粥を作る」は行動として生じるものだが、「食事に組立てる」は不自然であり生じることは滅多にない。後者のような行動は多くの文書集合で出現するものではないため、行動パターンの文書頻度を指標とすることで、日本語として不自然な行動を含むパターンはフィルタリングすることができる。

40

【0068】

閾値の設定のため実際に各行動について、何件の blog 記事に出現しているか出現頻度の調査を行ったところ、1 ～ 数回程度しか出現しない行動が全体の 8 割程度を占める傾向があった。

【0069】

図 6 は、キーワード「かぼちゃ」を含む記事から生成した行動の頻度分布を示したものである。縦軸が行動全体に占める割合、横軸が出現回数である。例えば、出現回数 1 回の行動は、全体の 7 割以上になる。出現回数 1 ～ 2 回の行動は全体の 8 割以上を占めている

50

。そこで、行動パターンを文書頻度の高い順に並べ上位20%を残し、残りをフィルタリングする。

【0070】

次に、仮定(c2)に基づき残差idfをベースとした手法を行動パターンの重み付け指標として利用する。残差idfとは、ポアソン分布を利用して単語のidf値を推定し、実際のidf値との差を測ることで、一般語よりも内容語に重み加わるようにした重要語抽出技術である。一般には文書中の単語の重要度を算出するために使われるが、今回は行動パターンを1つの単語とみなし、各行動パターンについて残差idf値を算出する。

【0071】

また、実際のidf値は状況に対応するクラスタ内におけるidf値、推定したidf値はキーワードを含む文書集合全体から算出することで、状況に対応するクラスタ内に偏って出現する行動パターンにより高い重みを与えるようにしている。実際のidf値から推定したidf値の差をとったとき、その値が大きいということは、行動パターンが文書集合全体よりもクラスタ内の文書に偏って出現していることを示している。つまり、差分が大きい行動パターンは特定の状況に依存している可能性が高いことを意味している。

【0072】

具体的な算出方法を以降に示す。ポアソン分布は文書において語がランダムに生起する場合の生起回数を確率的に表現するモデルである。

【0073】

ポアソン分布はkを行動パターンの生起回数、λを期待値として次の式で表される。

【数1】

$$P(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (1)$$

【0074】

ここで、 F_i を行動パターンiの大域的頻度、nをキーワードを含む全文書数とすると、ある文書中で行動パターンiが1回以上出現する確率pは次の式で表される。

【数2】

$$p = 1 - P(0; \frac{F_i}{n}) = 1 - e^{-\frac{F_i}{n}} \quad (2)$$

【0075】

さらに idf_{ij} を、クラスタリング後のクラスタj内での行動パターンiの文書頻度の逆数とし、 n_j をクラスタjに属する全文書数とする。

このとき、残差idf値 $ridf_{ij}$ は次の式で求められる。

【数3】

$$ridf_{ij} = idf_{ij} - \log(1 - e^{-\frac{F_i}{n}}) \quad (3)$$

【0076】

文書内にスパムブログ（行動や経験が書かれておらず、商品の宣伝等を目的とした記事など）が含まれていない場合には、上記の式で算出した残差idf値を利用して値の大きい上位の行動パターンのみに着目すると精度良く抽出することが可能である。しかし、実際には収集したblog記事の中にそのような記事が含まれており、特定の文書内に繰り返し出現する行動パターンがあった場合、残差idf値が大きくなることがある。そこで、トピックモデルで文書を分類した際に得られた特徴語を含む行動パターンのみを残す。また、各特徴語に対して最も関連する格助詞・動詞を組合せて行動を構成し、その行動を含む行動

10

20

30

40

50

パターンで残差idf値が高いものを抽出すると精度は向上する。

【0077】

実際にキーワード「札幌」を含むブログ記事約12,000件を解析した結果を示す。特徴語抽出部10で名詞と未知語のBOWを作成してLDAを適用、トピック数は50に設定した場合、表4のような結果が得られる。

【表4】

トピック 番号	特徴語
0	北海道、車、バス、ホテル、駅、写真、 小樽、温泉、旅行、列車、函館
1	ラーメン、店、北海道、スープ、カレー、麺、 キャラメル、ビール、ジンギスカン、味噌
2	ドーム、選手、日本ハム、試合、楽天、 ファイターズ、巨人、観戦、プロ野球
3	アイヌ、文化、民族、調査、地域、支援、 事業、生活、歴史、音楽、演奏
4	選手、サッカー、コンサドーレ札幌、監督、 節、ゴール、ホーム、得点、試合
...	...

10

20

【0078】

次に、行動プロセス生成部20は各トピックに最も関連する文書から行動パターン候補を生成する。つまり、仮定(c1)に基づき文書頻度を算出し、上位にくる行動パターンを表5に示す。

【表5】

トピック 番号	文書頻度の高い行動パターン
0	スピードを出す→写真を撮る、前で撮る→写真を撮る、 北海道へ行く→飛行機に乗る、話になる→写真を撮る
1	本場という→北海道で囲む、北海道を訪問→札幌を訪問、 観光で訪れる→札幌を訪れる、北海道に遠征→味覚を楽しむ
2	散歩に行く→味噌汁を作る、試合を見に行く→試合を見る、 札幌ドームで開幕→散歩に行く、札幌ドームに行く→家に帰る
3	調査を行う→場所で予測、アイヌ民族を認める→先住民族と認める、 道が委託→アイヌ民族協会に委託、生活を強いる→損害賠償を求める
4	ボールを持つ→選手が孤立、歴史に刻む→名を刻む、 ゴールで先制→シュートを打つ、ボールを奪う→試合を見る
...	...

30

40

【0079】

そして、行動プロセス生成部20は、仮定(c2)に基づき残差idf値を算出し、上位にくる行動パターンを表6に示す。

【表 6】

トピック 番号	残差idf値の高い行動パターン
0	峠を越える→写真を撮る、温泉に入る→ホテルに戻る、 自転車に乗る→写真を撮る、ホテルに戻る→舌鼓を打つ
1	本場という→北海道で囲む、バスで移動→トイレに行く、 ジンギスカンを食べる→ガイド本に載る、 店に入る→味噌ラーメンを注文
2	札幌ドームに行く→応援に行く、本拠地を移る→北海道に来る、 日本シリーズを見る→日本ハムが勝つ、試合を見る→テレビで見る
3	地域で暮らす→ルールを守る、 アイヌ民族を認める→先住民族と認める、 生活を支える→制度を確立、アイヌ文化を発信→理解を得る
4	選手を集める→対策を打つ、コンサドーレ札幌に移籍→穴を埋める、 ゴール前で転ぶ→ボールを奪う、試合を見に行く→試合を見る
...	...

10

【0080】

20

表5と比較して表6の方が具体的な行動プロセスを抽出できていることが分かる。例えば、トピック番号1において「北海道を訪問→札幌を訪問」という行動パターンは実際に生じることであり、間違ったパターンであるとは言えない。しかし、行動パターンを観光ナビゲーションなどに応用する場合、「店に入る→味噌ラーメンを注文」のように、より具体的な行動パターンの方が一般的すぎる知識よりも有用であると考えられる。なお、表6の各トピックについて上位10件ずつ計50件のパターンを評価したところ、正解率は64%であった。他のキーワードについても正解率は概ね65%前後となった。

【0081】

さらに、トピックモデルで抽出した特徴語を含む行動プロセス以外をフィルタリングすると表7のようになり、例えばトピック番号1における「バスで移動→トイレに行く」はフィルタリングされる。

30

【表 7】

トピック 番号	特徴語を含み、かつ残差idf値の高い行動パターン
0	峠を越える→写真を撮る、温泉に入る→ホテルに戻る、 自転車に乗る→写真を撮る、電車に乗る→札幌に向かう
1	本場という→北海道で囲む、札幌に行く→スープカレーを食べる、 ジンギスカンを食べる→ガイド本に載る、 店に入る→味噌ラーメンを注文
2	札幌ドームに行く→応援に行く、本拠地を移る→北海道に来る、 日本シリーズを見る→日本ハムが勝つ、試合を見る→テレビで見る
3	地域で暮らす→ルールを守る、 アイヌ民族を認める→先住民族と認める、 生活を支える→制度を確立、アイヌ文化を発信→理解を得る
4	選手を集める→対策を打つ、コンサドーレ札幌に移籍→穴を埋める、 ゴール前で転ぶ→ボールを奪う、試合を見に行く→試合を見る
...	...

40

【0082】

50

精度をさらに高める方法として、単に特徴語を含む行動プロセスを残すのではなく、各特徴語を含む文書頻度の高い行動を含むプロセスを残すことも一定の効果があるが、抽出される行動パターンが限定的になることも考えられる。トピック番号1における各特徴語について最も文書頻度の高い行動を表8に示す。

【表8】

トピック 番号	特徴語を含む行動
1	ラーメンを食べる、店に行く、北海道に行く、 スープカレーを食べる、麺を食べる、キャラメルを買う、 ビールを飲む、ジンギスカンを食べる、味噌ラーメンを食べる

10

【0083】

表8の行動を含み、かつ残差idf値の高い行動パターンを表9に示す。人手で各トピックについて上位10件ずつ計50件のパターンを評価したところ、正解率は68%である。他のトピックやキーワードについても概ね70%前後の正解率でパターンを抽出することができる。

【表9】

トピック 番号	表8の行動を含む残差idf値の高い行動パターン
1	ジンギスカンを食べる→ガイド本に載る、 札幌に行く→スープカレーを食べる、 キャラメルを買う→お土産に配る、 店に行く→味噌ラーメンを食べる、 空港に向かう→ラーメンを食べる、 ビールを飲む→ガイド本に載る

20

【0084】

行動プロセス生成部20で生成され、フィルタリング機能41に入力される行動プロセス集合からランダムにプロセスを100個選択し、同様に人手で評価を行ったところ、正解率は0~2%である。すなわち、フィルタリング機能41によりノイズを減らし、行動プロセス抽出の精度を向上させることができている。

30

【0085】

以上のように上記第2の実施形態によれば、行動プロセス生成部20により抽出された複数の行動プロセスそれぞれについて、フィルタリング機能41によりデータ集合中で起こり得る確率を算出し、起こり得る確率の低い行動プロセスをノイズとしてフィルタリングすることにより、行動プロセス抽出精度の向上を図ることができる。

【0086】

(その他の実施形態)

なお、この発明は上記各実施形態に限定されるものではない。例えば、特徴語抽出部10に特徴語抽出機能11の他に、特徴語選択機能12を備えるようにし、行動プロセス生成部20に動作抽出機能21の他に、前後関係算出機能22を備える例について説明したが、例えば、特徴語抽出部10に特徴語抽出機能11を備え、行動プロセス生成部20に動作抽出機能21及び、前後関係算出機能22を備えるものであってもよい。また、例えば、特徴語抽出部10に特徴語抽出機能11及び特徴語選択機能12を備え、行動プロセス生成部20に動作抽出機能21を備えるものであってもよい。

40

【0087】

また、上記第2の実施形態では、出力部32の前段に、フィルタリング機能41を追加

50

する例について説明した。出力部 32 の前段に、フィルタリング機能 41 を追加することが好適であるが、フィルタリング機能 41 は特徴語抽出機能 11 から出力部 32 のどこに追加しても効果がある。例えば、特徴語抽出部 10 に特徴語抽出機能 11 を備え、行動プロセス生成部 20 に動作抽出機能 21 及び、前後関係算出機能 22 を備えるものである例についても、フィルタリング機能 41 は特徴語抽出機能 11 から出力部 32 のどこに追加しても同様の効果がある。さらに、例えば、特徴語抽出部 10 に特徴語抽出機能 11 及び特徴語選択機能 12 を備え、行動プロセス生成部 20 に動作抽出機能 21 を備えるものである例についても、フィルタリング機能 41 は特徴語抽出機能 11 から出力部 32 のどこに追加しても同様の効果がある。

【0088】

10

また、特徴語抽出機能 11、特徴語選択機能 12、動作抽出機能 21、前後関係算出機能 22、フィルタリング機能 41 で説明した個々の処理は、コンピュータプログラムによってソフトウェア処理することが可能である。

要するにこの発明は、上記実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、上記実施形態に開示されている複数の構成要素の適宜な組み合わせにより種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。さらに、異なる実施形態に亘る構成要素を適宜組み合わせてもよい。

【符号の説明】

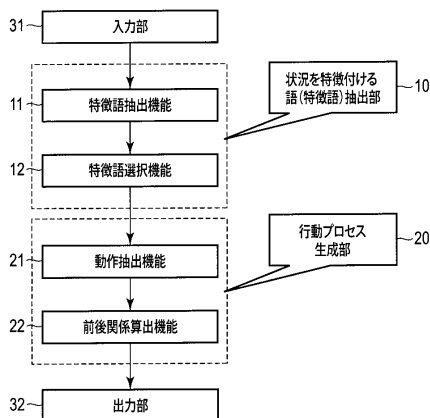
【0089】

20

10…特徴語抽出部、11…特徴語抽出機能、12…特徴語選択機能、20…行動プロセス生成部、21…動作抽出機能、22…前後関係算出機能、31…入力部、32…出力部、41…フィルタリング機能。

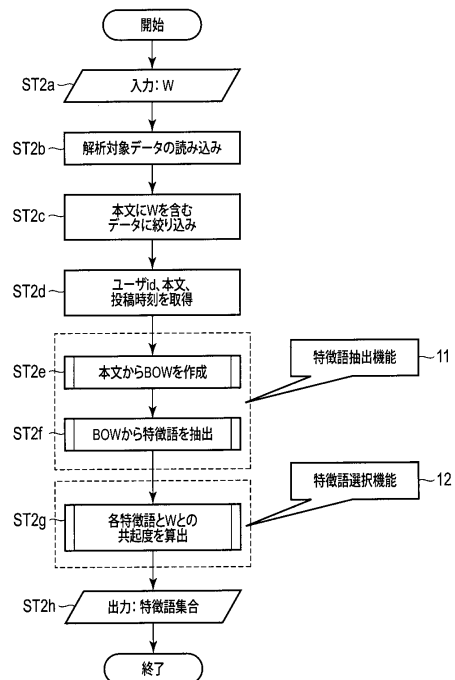
【図 1】

図 1



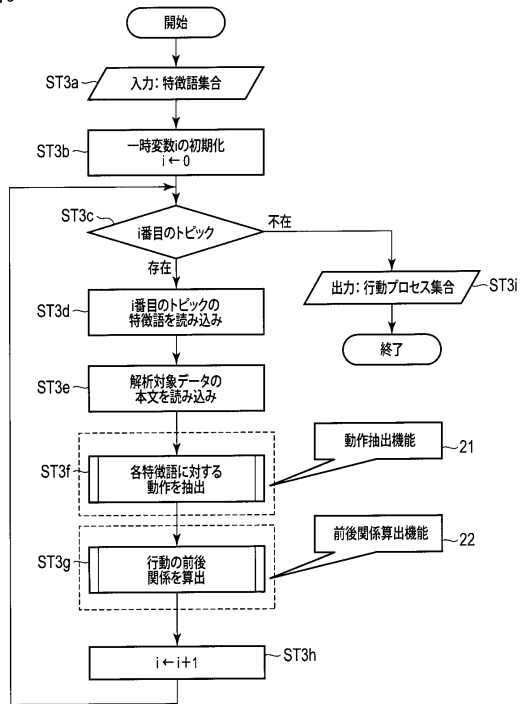
【図 2】

図 2



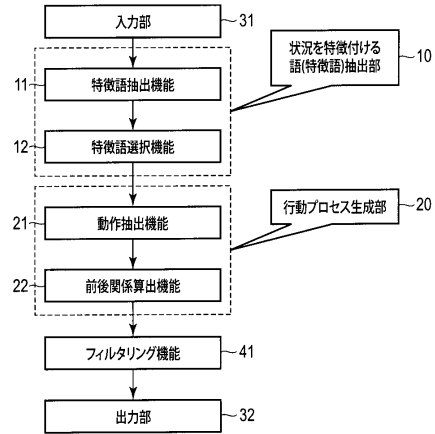
【図 3】

図 3



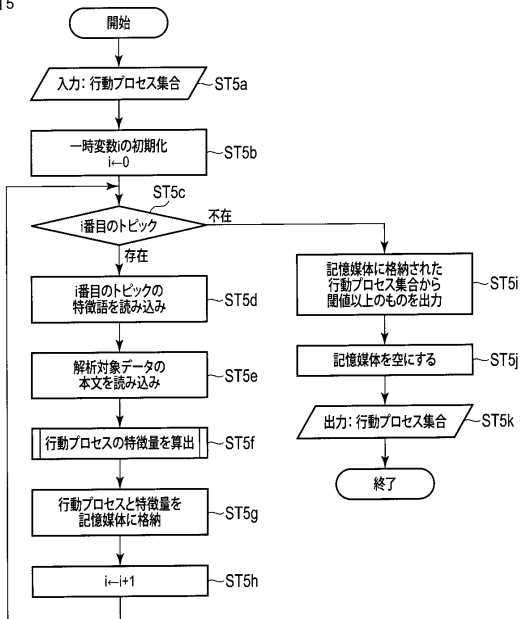
【図 4】

図 4



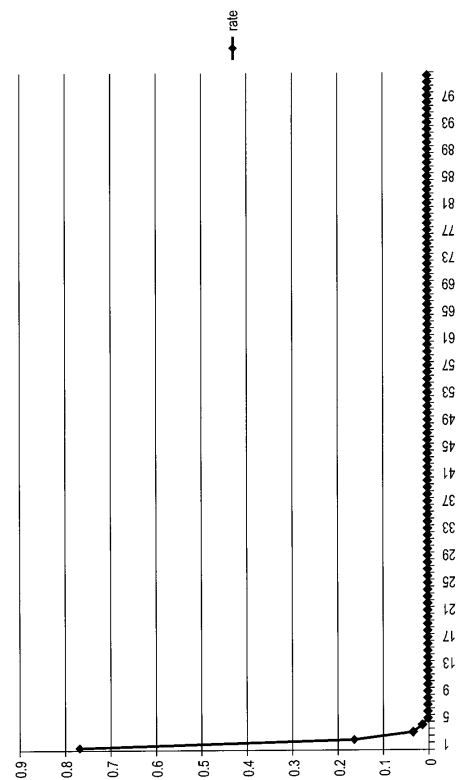
【図 5】

図 5



【図 6】

図 6



フロントページの続き

- (72)発明者 佐藤 進也
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
- (72)発明者 松尾 真人
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

審査官 田中 秀樹

- (56)参考文献 米国特許出願公開第2012/0296931 (US, A1)
特開2012-022530 (JP, A)
高橋 公海, Webからの行動プロセス抽出手法の提案, 電子情報通信学会技術研究報告, 日本
, 一般社団法人電子情報通信学会, 2012年11月19日, 第112巻, 第319号, P.31-35
- (58)調査した分野(Int.Cl., DB名)
G06F 17/30