

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2020年1月2日(02.01.2020)



(10) 国際公開番号

WO 2020/004401 A1

(51) 国際特許分類:
G06F 16/90 (2019.01)

(21) 国際出願番号: PCT/JP2019/025197

(22) 国際出願日: 2019年6月25日(25.06.2019)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願 2018-124735 2018年6月29日(29.06.2018) JP

(71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

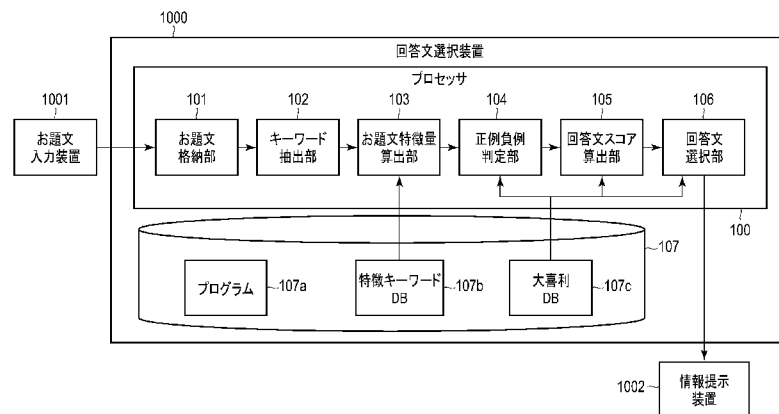
(72) 発明者: 伊勢崎 隆司 (ISEZAKI, Takashi); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).
山田 智広 (YAMADA, Tomohiro); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).

(74) 代理人: 蔵田 昌俊, 外 (KURATA, Masatoshi et al.); 〒1050014 東京都港区芝三丁目23番1号 セレスティン芝三井ビルディング11階 鈴榮特許総合事務所内 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

(54) Title: ANSWER SENTENCE SELECTION DEVICE, METHOD, AND PROGRAM

(54) 発明の名称: 回答文選択装置、方法、およびプログラム



- 100 Processor
- 101 Theme sentence storage unit
- 102 Keyword extraction unit
- 103 Theme sentence feature quantity calculation unit
- 104 Positive instance/negative instance determination unit
- 105 Answer sentence score calculation unit
- 106 Answer sentence selection unit
- 107a Program
- 107b Feature keyword database
- 107c Ohgiri database
- 1000 Answer sentence selection device
- 1001 Theme sentence input device
- 1002 Information presentation device

(57) Abstract: A theme sentence feature quantity calculation unit 103 calculates the feature quantity F_q of a given theme sentence Q on the basis of a cosine similarity with a feature keyword in a feature keyword database 107b. A positive instance/negative instance determination unit 104 applies the feature quantity F_q to a positive instance/negative instance determination model and determines whether an answer sentence A is positive or negative. An answer sentence score calculation unit 105 calculates a cosine similarity between a plurality of theme

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 国際調査報告 (条約第21条(3))

sentences paired with the answer sentence A that is determined to be a positive instance and the theme sentence Q, the highest value of the cosine similarity being made the score of the answer sentence A. An answer sentence selection unit 106 selects the answer sentence that indicates the maximum score as an answer sentence to the theme sentence Q.

(57) 要約: お題文特徴量算出部 103 で、特徴キーワード DB 107 b の特徴キーワードとのコサイン類似度に基づき、与えられたお題文 Q の特徴量 F_q を計算する。正例負例判定部 104 により、特徴量 F_q を正例負例判定モデルに適用し、回答文 A の正負を判定する。正例と判定された回答文 A のペアとなる複数のお題文と、お題文 Q とのコサイン類似度を回答文スコア算出部 105 で計算し、その最高値を回答文 A のスコアとする。そして、最大のスコアを示す回答文を、回答文選択部 106 により、お題文 Q に対する回答文として選択する。

明 細 書

発明の名称： 回答文選択装置、方法、およびプログラム

技術分野

[0001] この発明は、例えば入力文に対する回答文を選択するための、回答文選択装置、方法、およびプログラムに関する。

背景技術

[0002] 『話芸』といわれる演芸のなかでも、言葉使いの妙を楽しむ『なぞかけ』や『大喜利』といったジャンルのファンは多い。大喜利では、例えば「こんな〇〇〇は嫌だ!」といったお題に対して気の利いた答えを返すことが聴衆の笑いを誘うものとなっており、ひとつの芸として評価される。

[0003] 近年、コンピュータは社会のあらゆる領域で使われていて、例えばタブレットを用いた音声翻訳サービスや、AI (Artificial Intelligence) スピーカの活用などが知られるようになってきている。そのキーとなるのは、入力文に対する適切な回答文を選択する技術であり、この技術が、エンターテインメントの分野でも活用されようとしている。つまり、お題への回答文をコンピュータに生成させるといった用途である。

[0004] 入力文に対して回答文を選択する技術としては、大規模なテキストデータから述語項構造データベースを作成し、対話の話題を示す単語を含むレコードを述語項構造データベースから選択して、回答文を生成する方式がある（例えば特許文献1を参照）。一般的なQ & Aについては、ユーザの発話文の履歴に含まれる単語が回答文に含まれる頻度をスコアとして、回答文を選択する技術がある（例えば特許文献2を参照）。

先行技術文献

特許文献

[0005] 特許文献1：日本国特開2015-79383号公報

特許文献2：日本国特開2007-102104号公報

発明の概要

発明が解決しようとする課題

- [0006] 大喜利には、大規模テキストデータを用いる手法を適用することが難しい。一般的な大規模テキストデータ内では、大喜利のお題文に対して共起性の低い回答文が正例となるケースが多くなってしまうため、少なくとも、笑いを生起する述語項構造データベースの構築は困難である。
- [0007] また、大喜利は、普通の会話とは違って時系列的な要素が少ない。このためユーザ発話文の単語やお題文の単語の履歴に基づくスコアリングでの手法では、回答文を適切に選択することが難しかった。
- [0008] この発明は、上記事情に着目してなされたもので、その一側面では、お題文に適切な回答文を選択できる技術を提供しようとするものである。

課題を解決するための手段

- [0009] 上記課題を解決するためにこの発明の第1の態様は、各回答文に対して、正例あるいは負例が付与されたお題 Q_i の特徴量から構築された各回答文の正例負例判定モデルを用いて、各回答文に対する与えられた題文 Q の正例負例を判定する正例負例判定手段と、前記正例負例判定手段により正例と判定された回答文を対象に、回答文に対して正例であるお題 Q_i の特徴量と前記題文 Q の特徴量とから回答文のスコアを求める回答文スコア算出手段と、前記回答文のスコアに基づいて、前記題文 Q に対する回答文を選択する回答文選択手段と、を具備する。

発明の効果

- [0010] この発明の第1の態様によれば、各回答文に対して、正例あるいは負例が付与されたお題 Q_i の特徴量から構築された各回答文の正例負例判定モデルを用いて、正例負例判定手段により、各回答文に対する与えられた題文 Q の正例負例が判定される。また、前記正例負例判定手段により正例と判定された回答文を対象に、回答文に対して正例であるお題 Q_i の特徴量と前記題文 Q の特徴量とから、回答文スコア算出手段により、回答文のスコアが求められる。そして、前記回答文のスコアに基づいて、前記題文 Q に対する回答文が、回答文選択手段により選択される。このような構成であるから、お題文

に対する適切な回答文を選択することが可能となる。

図面の簡単な説明

[0011] [図1]図1は、実施形態に係わる回答文選択装置の一例を示す機能ブロック図である。

[図2]図2は、特徴キーワードDB107bに記憶される特徴キーワードの一例を示す図である。

[図3]図3は、図2の特徴キーワードごとに対応付けられたキーワードベクトル列の一例を示す図である。

[図4]図4は、正例ペアをつくるお題文と回答文との対応の一例を示す図である。

[図5]図5は、負例ペアをつくるお題文と回答文との対応の一例を示す図である。

[図6]図6は、一実施形態に係わる回答文選択装置1000の処理手順の一例を示すフローチャートである。

[図7]図7は、正例負例の判定およびスコアの算出について説明するためのフローチャートである。

[図8]図8は、回答文スコアデータについて説明するための図である。

発明を実施するための形態

[0012] 次に、図面を参照してこの発明の実施の形態について説明する。以下では、話芸としての大喜利を採りあげ、大喜利の枠組みの中で、お題文に対して聴衆（ユーザ）の笑いを生起する回答文を選択する技術について説明する。

[0013] [一実施形態]
(構成)

図1は、実施形態に係わる回答文選択装置の一例を示す機能ブロック図である。一実施形態では、回答文選択装置1000は、お題文入力装置1001からお題文を入力され、このお題文に対して回答文を作成し、情報提示装置1002に出力する。

[0014] お題文入力装置1001は、ユーザから与えられたお題文を受け付け、テ

キストデータ形式のお題文を回答文選択装置１０００に渡す。お題文入力装置１００１は、例えばパソコンやスマートフォン等のヒューマンマシンインタフェース装置である。お題文は、これらの装置にテキスト形式で入力されることが想定される。このほか、ＡＩスピーカ等の音声認識装置や、マイク等の音声入力装置にお題文を音声で与え、テキストに認識させるようにしても良い。

[0015] 情報提示装置１００２は、回答文をテキストで表示するディスプレイ等が想定される。このほか、回答文を音声データに変換してスピーカから拡声出力するようにしても良い。お題文入力装置１００１と情報提示装置１００２を一体化すれば、お題文に回答を返すガジェットとして楽しむことができる。

[0016] 回答文選択装置１０００は、コンピュータであり、ハードウェアとしてのプロセッサ１００と、記憶部１０７とを備える。

プロセッサ１００は、記憶部１０７に記憶されたプログラム１０７ａに従って、実施形態に係わる機能を実現する。

[0017] 記憶部１０７は、ＲＯＭ（Read Only Memory）、ＲＡＭ（Random Access Memory）等の半導体メモリ、あるいはＨＤＤ（Hard Disk Drive）やＳＳＤ（Solid State Drive）等のストレージメディアである。記憶部１０７は、プロセッサ１００を回答文選択装置１０００として機能させるためのプログラム１０７ａと、特徴キーワードを記憶する特徴キーワードデータベース（ＤＢ）１０７ｂと、複数の回答文を記憶する大喜利データベース（ＤＢ）１０７ｃとを記憶する。

[0018] ところで、プロセッサ１００は、実施形態に係わる処理機能として、お題文格納部１０１、キーワード抽出部１０２、お題文特徴量算出部１０３、正例負例判定部１０４、回答文スコア算出部１０５、および、回答文選択部１０６を備える。これらの機能ブロックは、プログラム１０７ａがメモリにロードされ、当該プログラムの実行の過程で生成されるプロセスとして、理解され得る。

- [0019] お題文格納部 101 は、お題文入力装置 1001 から与えられたお題文（お題文 Q と表記する）を受け付け、保持し、キーワード抽出部 102 に出力する。
- [0020] キーワード抽出部 102 は、お題文格納部 101 から渡されたお題文 Q を、例えば既知の形態素解析技術により単語の単位にまで分割して単語群を得る。さらにキーワード抽出部 102 は、この単語群から、事前に定義された既定の単語（除外単語）を除去し、単語群 $[w_1, w_2, \dots, w_n]$ を得る。n は、除外単語群を除いたお題文 Q の単語長である。このお題文 Q に由来する単語群 $[w_1, w_2, \dots, w_n]$ は、お題文特徴量算出部 103 に送られる。
- [0021] お題文特徴量算出部 103 は、キーワード抽出部 102 から渡された単語群 $[w_1, w_2, \dots, w_n]$ から、お題文 Q の特徴量 F_q を計算する。単語群 $[w_1, w_2, \dots, w_n]$ は、お題文 Q の各単語をベクトル化した配列といえる。お題文特徴量算出部 103 は、この配列と、特徴キーワード DB 107b とを用いてお題文 Q の特徴量 F_q を計算する。
- [0022] 図 2 は、特徴キーワード DB 107b に記憶される特徴キーワードの一例を示す図である。図 2 に示されるような特徴キーワードは、例えば、既存のデータベース（例えば Wikipedia（登録商標）等）からのデータマイニング等の手法で予め抽出されることができる。
- [0023] 既存の Word2Vec 等の手法を適用すれば、単語の概念をベクトルで表現可能な単語空間を構築することができる。Word2Vec は、大量のテキストデータを解析し、各単語の意味をベクトル表現化す手法の一つとして知られている。図 2 の特徴キーワード群に対して Word2Vec を適用することにより、特徴キーワード群の要素の特徴キーワードの各々を、それぞれベクトル表現で表すことができる。
- [0024] 図 3 は、図 2 の特徴キーワード群に、例えば Word2Vec を適用して作成されたキーワードベクトル列の一例を示す図である。図 3 に示される m は、特徴キーワード数である。
- [0025] 単語間の類似度は、例えばコサイン類似度で算出することができる。例え

ばベクトル w_1 と t_1 との類似度は、スカラー量である $\cos(w_1, t_1)$ で表される。お題文特徴量算出部 103 は、配列 $[w_1, w_2, \dots, w_n]$ とキーワードベクトル $[t_1, t_2, \dots, t_m]$ とのコサイン類似度を算出する。そして、お題文特徴量算出部 103 は、各特徴キーワードに対してお題文 Q の各単語の最高類似度と最低類似度とに基づいて、例えば式 (1) を用いてお題文 Q の特徴量 F_q を計算する。

[0026] [数1]

特徴量 F_q

$$= \begin{bmatrix} \max_{i=1 \sim n} (\cos(w_i, t_1)), \min_{i=1 \sim n} (\cos(w_i, t_1)), \\ \max_{i=1 \sim n} (\cos(w_i, t_2)), \min_{i=1 \sim n} (\cos(w_i, t_2)), \\ \dots, \\ \max_{i=1 \sim n} (\cos(w_i, t_m)), \min_{i=1 \sim n} (\cos(w_i, t_m)) \end{bmatrix}$$

... (1)

[0027] 式 (1) の右辺における $\max$ はコサイン類似度の最大値を表し、 $\min$ は最小値を表す。右辺の各要素は、配列 $[w_1, w_2, \dots, w_n]$ の全てについての、キーワードベクトル $[t_1, t_2, \dots, t_m]$ とのコサイン類似度の最大値または最小値である。式 (1) によれば、特徴量 F_q は、 $2m$ 次元ベクトルである。計算された特徴量 F_q は、正例負例判定部 104 に送られる。

[0028] 正例負例判定部 104 は、大喜利 DB 107c に記憶された回答文の全てについて、お題文 Q が正例か、または負例かを、お題文 Q の特徴量 F_q を用いて判別する。一実施形態において、正例とは、お題文と回答文とのペアがユーザの笑いを想起することを意味し、負例とは、お題文と回答文とのペアがユーザの笑いを想起しないことを意味する。

[0029] 正例か負例かの判別には、予め構築された正例負例判定モデルを用いることができる。ここで、正例負例判定モデルは、例えば、自然言語解析で一般的な、サポートベクトルマシンなどを用いて構築することができる。このほ

か、k-近傍法などの他の機械学習手法・統計手法を用いてもよい。

[0030] 図4および図5は、大喜利DB107cに記憶される内容の一例を示す図である。大喜利DB107cは、お題文と回答文との正例ペア（図4）と、お題文と回答文との負例ペア（図5）とを記憶する。一実施形態において、正例ペアとは、ユーザの笑いを生起するお題文と回答文とのペアであり、負例ペアとは、ユーザの笑いを生起しないお題文と回答文とのペアである。例えばお題文 Q_1 について図4、図5を参照すると、回答文 A_1 , A_2 , A_3 , A_4 がお題文 Q_1 と正例ペアになり、回答文 A_5 , A_6 , A_7 は、お題文 Q_1 と負例ペアになる。

[0031] 一つの回答文は、お題文の特定の特徴に対して成立する回答文であるため、他の回答文においても同様の特徴を有していれば正例ペアとして存在する。つまり、回答文が正例となるか、負例となるかは、お題文の特定の特徴に拠っている。

[0032] 実施形態では、大喜利におけるお題文とその回答文が、お題文の特定の特徴のもとで結合していることに着目する。例えば、「こんなテーブルは嫌だ」というお題文に対し「傾いている」という回答文があるとする。これらは「テーブルは水平である」という特徴のもとで結合している。

[0033] 一方、水平を特徴として含む対象としては、椅子や家なども該当する。よって「傾いている」という回答文は、「こんな椅子は嫌だ」や「こんな家は嫌だ」といったお題文にも正例として結合するべきといえる。

[0034] 正例負例判定部104は、正例負例判定モデルにお題文 Q の特徴量 F_q を入力し、出力ラベルが1であれば正例判定、-1であれば負例判定として判定結果を回答文スコア算出部105に出力する。

[0035] 回答文スコア算出部105は、大喜利DB107cの各回答文に対するお題文 Q の正例負例判定結果を、正例負例判定部104から取得する。そして、回答文スコア算出部105は、お題文 Q の特徴量 F_q に基づいて、大喜利DB107cの各回答文ごとに、お題文 Q との適合度を示すスコアを計算する。計算されたスコアは、回答文選択部106に渡される。

[0036] 回答文選択部 106 は、回答文スコア算出部 105 から渡されたスコアに基づいて、お題文 Q に対して最大のスコアを示す回答文を選択し、情報提示装置 1002 に出力する。情報提示装置 1002 は、選択された回答文をテキスト表示、あるいは音声出力などの形態でユーザに提示する。次に、上記構成における作用を説明する。

[0037] (作用)

図 6 は、一実施形態に係わる回答文選択装置 1000 の処理手順の一例を示すフローチャートである。図 6 において、お題文 Q が与えられると（ステップ S1）、回答文選択装置 1000 は、このお題文 Q を単語の単位にまで分割し、お題文 Q のキーワードを抽出する（ステップ S2）。次に、回答文選択装置 1000 は、抽出されたお題文 Q のキーワードからお題文 Q の特徴量 F_q を計算する（ステップ S3）

次に、回答文選択装置 1000 は、お題文 Q の特徴量 F_q と、大喜利 DB 107c に含まれる各お題文の特徴量との類似度を計算し、類似度の高いお題文を抽出する。そして、この抽出されたお題文に結合する回答文のお題回答適合度（正例：+1，負例：-1）を用いてスコアを算出する（ステップ S4）。

[0038] さらに、回答文選択装置 1000 は、算出されたスコアに基づいて、お題文 Q に対する回答文を大喜利 DB 107c から選択する（ステップ S5）。

[0039] 図 7 は、正例負例の判定およびスコアの算出について説明するためのフローチャートである。一実施形態では、回答文 A_1 に対するお題文 Q の正例／負例判定を例として説明する。

[0040] 図 7 において、回答文選択装置 1000 は、回答文 A_1 に対する正例ペアのお題群（集合）P と負例ペアのお題群（集合）N を、大喜利 DB 107c から取得する（ステップ S10）。図 4 を参照すると、正例ペアのお題群 P には Q_1 、 Q_2 が含まれ、図 5 を参照すると、負例ペアのお題群 N には Q_3 が含まれることが分かる。

[0041] 次に、回答文選択装置 1000 は、群 P の各お題文（ Q_1 、 Q_2 ）の特徴量

と、群Nの各お題文 (Q_3) の特徴量とを算出する。ここでも、コサイン類似度によりお題文の特徴量を算出することができる。お題文 Q_1 , Q_2 の特徴量を F_{q1} , F_{q2} とし、お題文 Q_3 の特徴量を F_{q3} とする。そして回答文選択装置 1000 は、 F_{q1} , F_{q2} をそれぞれラベル 1 とし、 F_{q3} をラベル -1 とし、サポートベクトルマシンを用いた機械学習により、正例負例判定モデルを構築する (ステップ S 11)。

[0042] 次に、回答文選択装置 1000 は、与えられたお題文 Q の特徴量 F_q を正例負例判定モデルに入力し (ステップ S 12)、出力ラベルが 1 か -1 かを判定する (ステップ S 13)。出力ラベルが 1 であれば、回答文選択装置 1000 は正例と判定し (ステップ S 14)、-1 であれば負例と判定する (ステップ S 15)。

[0043] 正例負例判定部 104 は、大喜利 DB 107c に記憶された全ての回答文について、お題文 Q との正例／負例を判定する。その結果を受けて、回答文スコア算出部 105 は、お題文 Q が正例判定であった回答文の集合 (回答文群 A_p とする) を抽出する。回答文スコア算出部 105 は、回答文群 A_p の要素である各々の回答文に対して、正例ペアとなるお題文の集合 (お題文群 O_p とする) を求める。例えば、お題文 Q が回答文 A_1 に対して正例と判定されれば、回答文 A_1 に正例ペアとなるお題文群 O_p は、 Q_1 , Q_2 が該当する (図 5)。

[0044] 次に、回答文スコア算出部 105 は、お題文 Q の特徴量 F_q と、お題文群 O_p の要素である各々のお題文 Q_1 , Q_2 の特徴量 F_{q1} , F_{q2} とのコサイン類似度を計算し、その最高値を、お題文 Q の回答文 A_1 に対するスコアとして設定する。そして、この処理を回答文群 A_p の要素である回答文の全てに対して行うことで、お題文 Q に対する回答文スコアデータが作成される。回答文選択部 106 は、回答文スコアデータに基づいて、お題文 Q に対する回答文を選択する。

[0045] 図 8 は、回答文スコアデータについて説明するための図である。図 8 において、回答文 A_1 のスコアは 0.8 であり、図示される中では最も高い。すな

わち図8を参照する限り、お題文Qに対しては回答文 A_1 が選択されることになる。なお負例判定となった回答文については、スコア算出は実施されない。図8においては、データ欠損を示すN a N (Not a Number) でこのことが示される。

[0046] (効果)

以上のように一実施形態では、正例ペアをなすお題文と回答文、および負例ペアをなすお題文と回答文を、大喜利DB107cに予め記憶させる。ユーザの笑いを想起する回答文とお題文とのペアを正例とし、笑いを想起しないことを負例として、或る一つの回答文Aに対しお題文の正例集合Pと負例集合Nとを求める。そして、集合Pと集合Nとを教師として学習させた正例負例判定モデルを構築する。

[0047] お題文Qが与えられると、その特徴量 F_q を、予めデータベース化された特徴キーワードとのコサイン類似度により算出する。特徴量 F_q を正例負例判定モデルに適用し、回答文Aの正例／負例を判定する。正例と判定された回答文Aのペアとなる複数のお題文と、お題文Qとのコサイン類似度を計算し、その最高値を回答文Aのスコアとする。そして、最大のスコアを示す回答文をお題文Qへの回答文として選択し、提示するようにした。

[0048] このようにしたので、述語項構造データベースを用いることなく、また、ユーザの発話履歴を参照する必要もなく、未知のお題文に対して、ユーザの笑いを生起する回答文を選択することができる。すなわちこの発明によれば、お題文に対して適切な回答文を選択することが可能になる。

[0049] なお、この発明は上記実施形態に限定されるものではない。例えば図7のフローチャートにおいて、正例負例判定モデルの構築に係わる手順（ステップS10, S11）と、この正例負例判定モデルを用いて正例または負例を判定する手順（ステップS12～S15）とが続けて実行される必要は無い。例えばステップS10, S11の手順だけを繰り返し実行することで、正例負例判定モデルの学習が積み重ねられ、ステップS12～S15における判定の精度の向上が期待できる。

[0050] また、プロセッサ 100 の機能ブロックの全て、あるいは少なくとも 1 つのブロックをクラウドコンピューティングシステムに実装することも可能である。回答文の選択に係わる処理をクラウドサービスに委託すれば、ユーザ側の装置の負担を軽減することができ、AI スピーカ等のガジェットや、スマホアプリ等の実現形態も可能となる。

このほか、回答文選択装置の機能構成等についても、この発明の要旨を逸脱しない範囲で種々変形して実施可能である。

[0051] なお、本願発明は、上記実施形態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種々に変形することが可能である。また、各実施形態は可能な限り適宜組み合わせて実施してもよく、その場合組み合わせた効果が得られる。更に、上記実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適当な組み合わせにより種々の発明が抽出され得る。

符号の説明

[0052] 100…プロセッサ
101…お題文格納部
102…キーワード抽出部
103…題文特徴量算出部
104…正例負例判定部
105…回答文スコア算出部
106…回答文選択部
107…記憶部
107a…プログラム
107b…特徴キーワードデータベース
107c…大喜利データベース
1000…回答文選択装置
1001…お題文入力装置
1002…情報提示装置

請求の範囲

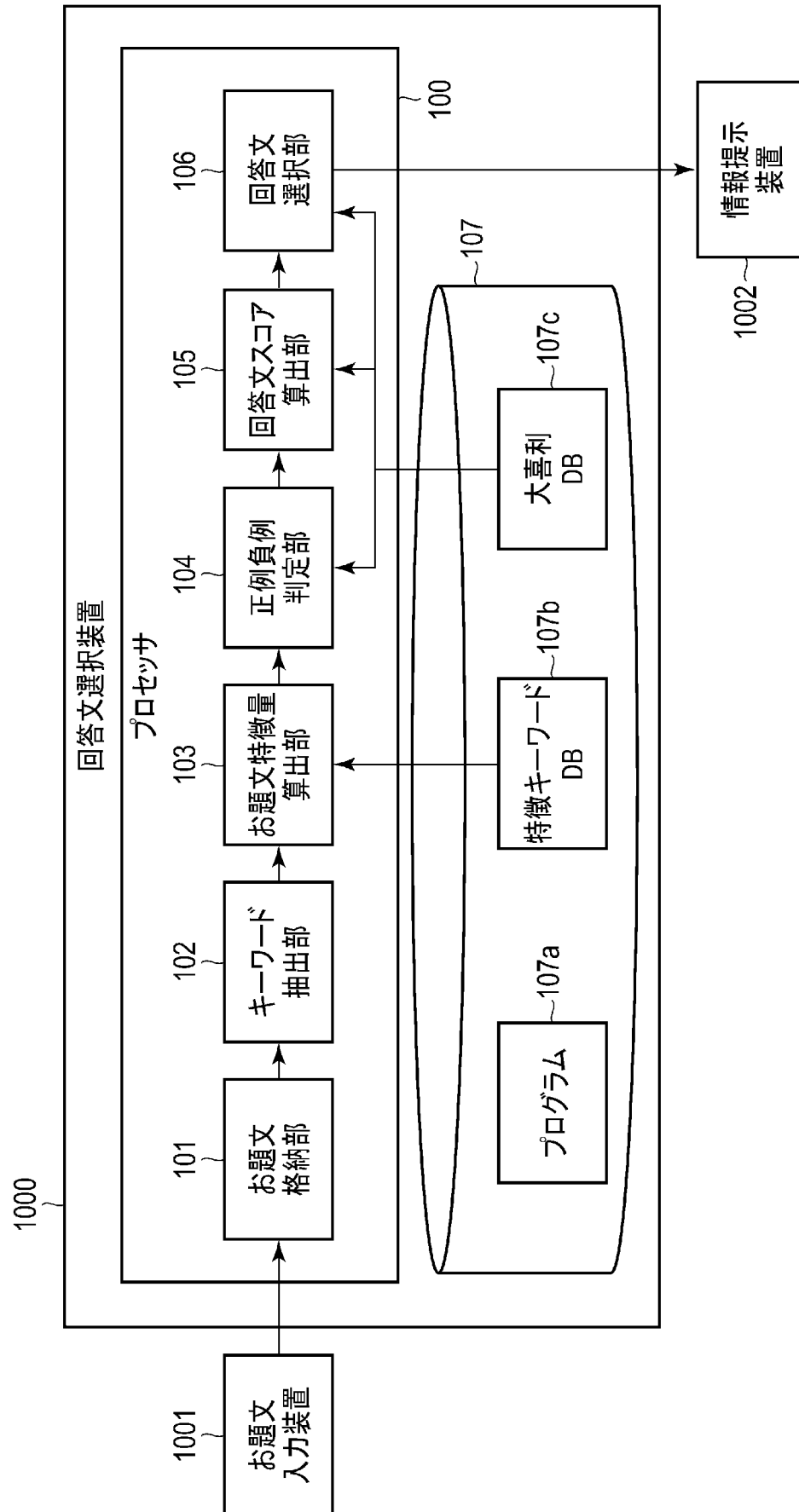
- [請求項1] 各回答文に対して、正例あるいは負例が付与されたお題Q i の特徴量から構築された各回答文の正例負例判定モデルを用いて、各回答文に対する与えられた題文Qの正例負例を判定する正例負例判定手段と、
- 前記正例負例判定手段により正例と判定された回答文を対象に、回答文に対して正例であるお題Q i の特徴量と前記題文Qの特徴量とから回答文のスコアを求める回答文スコア算出手段と、
- 前記回答文のスコアに基づいて、前記題文Qに対する回答文を選択する回答文選択手段と、
- を具備する、回答文選択装置。
- [請求項2] 前記お題Q i の特徴量はお題Q i に含まれる単語のキーワードベクトルから求め、
- 前記題文Qの特徴量は題文Qに含まれる単語のキーワードベクトルから求める、
- 請求項1に記載の回答文選択装置。
- [請求項3] 前記回答文スコア算出手段は、
- 前記お題Q i の特徴量と前記題文Qの特徴量とからコサイン類似度を計算し、コサイン類似度の最高値を回答文のスコアとする、
- 請求項2に記載の回答文選択装置。
- [請求項4] 題文を与えられた回答文選択装置が実行する回答文選択方法であって、
- 前記回答文選択装置が、各回答文に対して、正例あるいは負例が付与されたお題Q i の特徴量から構築された各回答文の正例負例判定モデルを用いて、各回答文に対する与えられた題文Qの正例負例を判定する過程と、
- 前記回答文選択装置が、前記正例と判定された回答文を対象に、回答文に対して正例であるお題Q i の特徴量と前記題文Qの特徴量とから

ら回答文のスコアを求める過程と、

前記回答文選択装置が、前記回答文のスコアに基づいて、前記題文 Q に対する回答文を選択する過程とを具備する、回答文選択方法。

[請求項5] 請求項 1 乃至 3 のいずれか 1 項に記載の回答文選択装置が備える前記各部の処理を、当該回答文選択装置が備えるプロセッサに実行させる、プログラム。

[図1]



[図2]

特徴キーワード
男
女
年輩
幼年
青年
中年
老年
...
大きい

[図3]

特徴キーワード	キーワードベクトル
男	t_1
女	t_2
年輩	t_3
幼年	t_4
青年	t_5
中年	t_6
老年	t_7
...	...
大きい	t_m

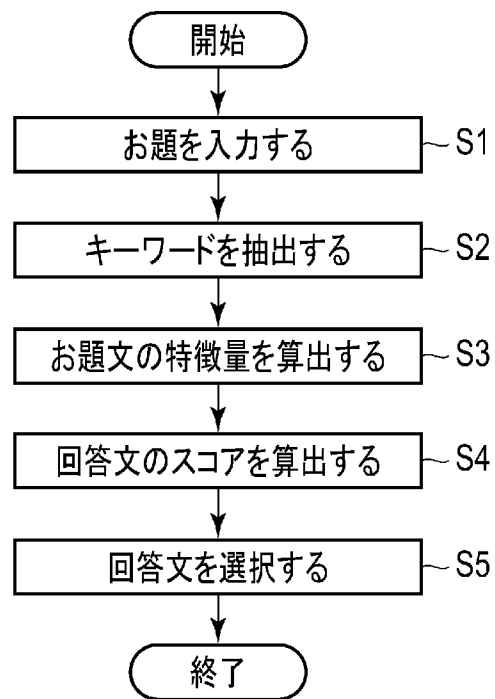
[図4]

お題	回答
Q_1	A_1
Q_1	A_2
Q_1	A_3
Q_1	A_4
Q_2	A_1
Q_2	A_5
Q_3	A_3
...	...

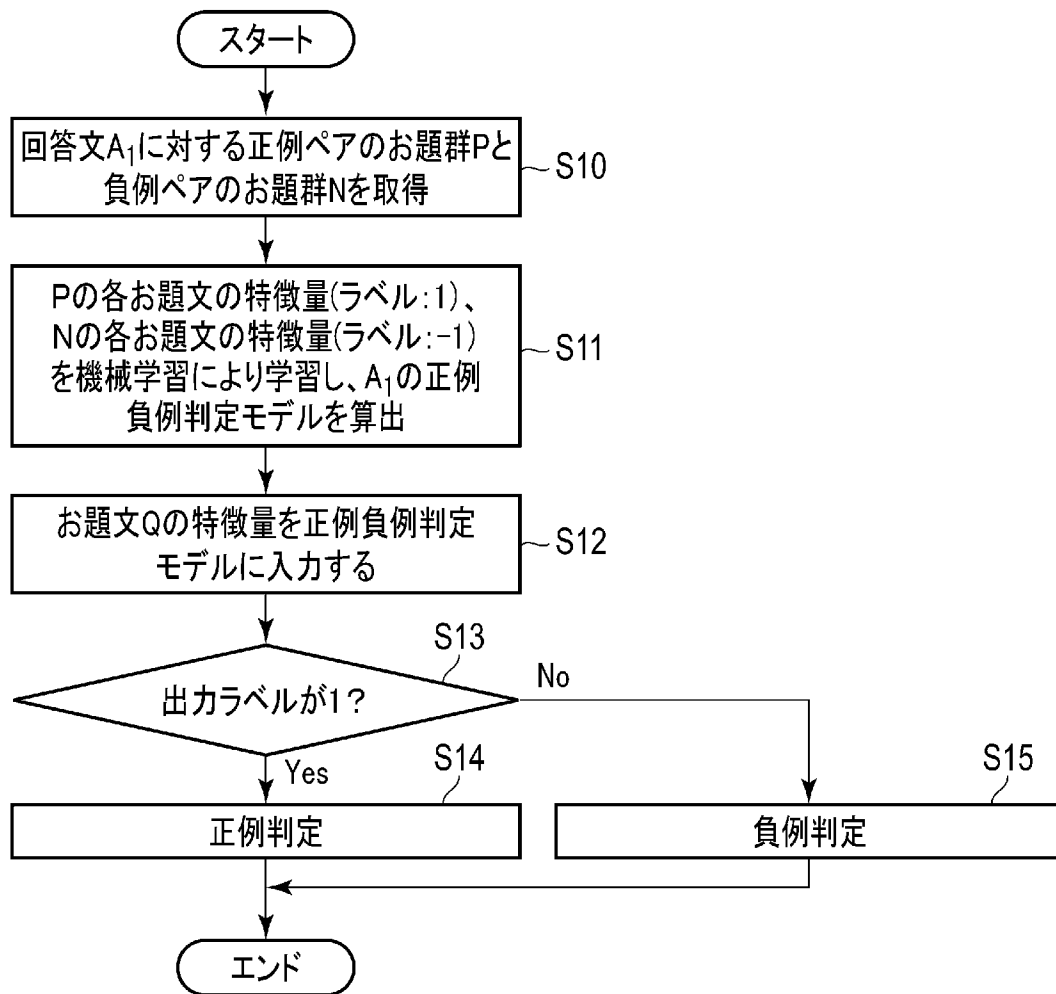
[図5]

お題	回答
Q_1	A_5
Q_1	A_6
Q_1	A_7
Q_2	A_2
Q_2	A_3
Q_2	A_4
Q_3	A_1
...	...

[図6]



[図7]



[図8]

回答文	スコア
A_1	0.8
A_2	0.4
A_3	Nan
A_4	Nan
A_5	0.74
A_6	0.3
A_7	0.4
A_8	Nan
...	...

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/025197

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06F16/90 (2019.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl. G06F16/90

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2019
Registered utility model specifications of Japan	1996-2019
Published registered utility model applications of Japan	1994-2019

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2017-151926 A (FUJITSU LTD.) 31 August 2017, paragraphs [0082]-[0125] & US 2017/0249320 A1, paragraphs [0098]-[0144]	1-5
A	JP 2013-254420 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 19 December 2013, paragraphs [0032]-[0082] (Family: none)	1-5

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

26 August 2019 (26.08.2019)

Date of mailing of the international search report

03 September 2019 (03.09.2019)

Name and mailing address of the ISA/

Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06F16/90(2019.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G06F16/90

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2017-151926 A（富士通株式会社）2017.08.31, 段落 [0082] - [0125] & US 2017/0249320 A1, 段落 [0098] - [0144]	1-5
A	JP 2013-254420 A（日本電信電話株式会社）2013.12.19, 段落 [0032] - [0082]（ファミリーなし）	1-5

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの	「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「O」 口頭による開示、使用、展示等に言及する文献	「&」 同一パテントファミリー文献
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日

26.08.2019

国際調査報告の発送日

03.09.2019

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

西村 直史

5 N

5088

電話番号 03-3581-1101 内線 3586