

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2020-53015

(P2020-53015A)

(43) 公開日 令和2年4月2日(2020.4.2)

(51) Int. Cl.

G06F 16/35 (2019.01)
G06N 3/08 (2006.01)

F I

G06F 16/35
G06N 3/08

テーマコード(参考)

審査請求 未請求 請求項の数 9 O L (全 18 頁)

(21) 出願番号 特願2019-126817 (P2019-126817)
(22) 出願日 令和1年7月8日(2019.7.8)
(31) 優先権主張番号 特願2018-174398 (P2018-174398)
(32) 優先日 平成30年9月19日(2018.9.19)
(33) 優先権主張国・地域又は機関 日本国(JP)

(71) 出願人 504133110
国立大学法人電気通信大学
東京都調布市調布ヶ丘一丁目5番地1
(74) 代理人 110000925
特許業務法人信友国際特許事務所
(72) 発明者 稲葉 通将
広島県広島市安佐南区大塚東三丁目4番1号
公立大学法人広島市立大学内

最終頁に続く

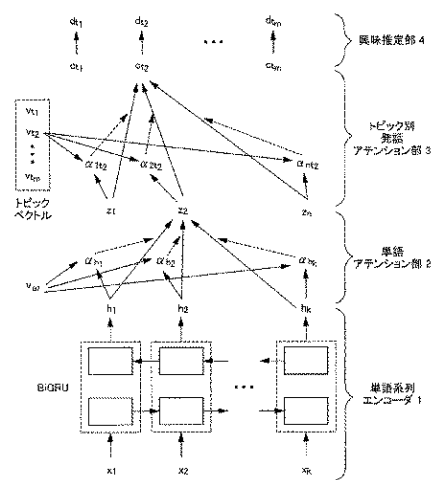
(54) 【発明の名称】 興味推定方法、興味推定装置および興味推定プログラム

(57) 【要約】

【課題】 雑談対話時のユーザの発話からユーザの興味のあるトピックを高精度に推定する。

【解決手段】 話者が発話した発話文に含まれる単語の系列から得られた単語ベクトル系列から隠れ状態ベクトル系列を作成する。そして、隠れ状態ベクトル系列から発話ベクトルを作成し、この発話ベクトルから興味トピックに関するコンテンツベクトルを作成する。それにより、コンテンツベクトルから興味トピックに対する発話者の興味値を求める。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

入力された文がどのようなトピックに属するかを事前に推定するための事前学習を行うステップと、
話者の発話に含まれる単語の系列から得られた単語ベクトル系列から隠れ状態ベクトル系列を作成するステップと、
前記隠れ状態ベクトル系列から発話ベクトルを作成するステップと、
前記発話ベクトルから興味トピックに関するコンテンツベクトルを作成するステップと、
前記コンテンツベクトルから前記興味トピックに対する発話者の興味度を求めるステップを含む

10

興味推定方法。

【請求項 2】

前記事前学習は、ニューラルネットワークから構成される単語系列エンコーダを用いて行われる、

請求項 1 に記載の興味推定方法。

【請求項 3】

文のトピック別分類を予め学習する前記事前学習において、文のトピックを分類する分類クラスを、興味推定対象のトピックと同じにする、

請求項 1 または 2 に記載の興味推定方法。

【請求項 4】

前記隠れ状態ベクトル系列は、双方向 GRU からなる単語系列エンコーダにより、双方向 GRU の中の順方向 GRU の出力と逆方向 GRU の出力を合成して作成される、

請求項 1 に記載の興味推定方法。

20

【請求項 5】

前記発話ベクトルは、単語ごとに異なる重みづけを行う単語アテンション部により作成される、

請求項 1 に記載の興味推定方法。

【請求項 6】

前記コンテンツベクトルは、トピック別に発話の重要性を計算するトピック別発話アテンション部により作成される、

請求項 1 に記載の興味推定方法。

30

【請求項 7】

前記コンテンツベクトルから前記興味トピックに対する発話者の興味度を求めるステップは、興味推定部により行われる

請求項 1 に記載の興味推定方法。

【請求項 8】

入力された文がどのようなトピックに属するかを事前に推定するための事前学習を行うトピック別分類部と、

話者の発話に含まれる単語の系列から得られた単語ベクトル系列から隠れ状態ベクトル系列を作成する単語系列エンコーダと、

前記隠れ状態ベクトル系列から発話ベクトルを作成する単語アテンション部と、

前記発話ベクトルから興味トピックに関するコンテンツベクトルを作成するトピック別発話アテンション部と、

前記コンテンツベクトルから前記興味トピックに対する発話者の興味度を求める興味推定部を、備える

興味推定装置。

40

【請求項 9】

入力された文がどのようなトピックに属するかを事前に推定するための事前学習を行う処理と、

話者の発話に含まれる単語の系列から得られた単語ベクトル系列から隠れ状態ベクトル系

50

列を作成する処理と、
 前記隠れ状態ベクトル系列から発話ベクトルを作成する処理と、
 前記発話ベクトルから興味トピックに関するコンテンツベクトルを作成する処理と、
 前記コンテンツベクトルから前記興味トピックに対する発話者の興味度を求める処理
 をコンピュータに実行させる興味推定プログラム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、興味推定方法、興味推定装置および興味推定プログラムに関し、より詳細には、雑談対話時のユーザの発話からユーザの興味のあるトピックを高精度に推定する方法、装置およびそれを実現するためのコンピュータプログラムに関する。

【背景技術】

【0002】

人と人とのコミュニケーションにおいて雑談は重要であり、人の会話の約62%は雑談であるとの報告がある。また、人とコンピュータとの対話においても雑談は重要であり、ユーザの満足度向上の観点から、システムがユーザの経験や好みに応じてユーザと対話を行う雑談対話システムが求められている。

20

【0003】

非特許文献1には、Twitter(登録商標)へ投稿されたテキストデータからユーザの興味を推定する方法が記載されている。すなわち、非特許文献1には、人名、イベント名、音楽グループ名といった固有表現の出現頻度により、Twitter(登録商標)ユーザの興味推定を行った例が記載されている。

【先行技術文献】

【非特許文献】

【0004】

【非特許文献1】Fabian Abel, Qi Gao, Geert Jan Houben, and Ke Tao. 2011. Analyzing user modeling on twitter for personalized news recommendations. User Modeling, Adaption and Personalization pages 1-12.

30

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、Twitter(登録商標)ではユーザの興味のあることが投稿されるのに対して、雑談では相手に話を合わせて対話するため、ユーザは必ずしも興味のあることについてのみ発話するとは限らない。したがって、非特許文献1に示された興味推定の方法を雑談対話に適用することは必ずしも適切ではない。

【0006】

本発明は、上記問題に鑑み、雑談対話中のユーザの発話からユーザが興味を持っているトピックを高精度に推定することができる興味推定方法、興味推定装置およびそれを実現するためのコンピュータプログラムを提供することを目的とする。

40

【課題を解決するための手段】

【0007】

上記課題を解決し、本発明の目的を達成するため、本発明の興味推定方法は、以下、(a)~(e)のステップを含む。

(a) 入力された文がどのようなトピックに属するかを事前に推定するための事前学習を行うステップと、

(b) 話者の発話に含まれる単語の系列から得られた単語ベクトル系列から隠れ状態ベクトル系列を作成するステップと、

50

- (c) 隠れ状態ベクトル系列から発話ベクトルを作成するステップと、
- (d) 発話ベクトルから興味トピックに関するコンテンツベクトルを作成するステップと、
- (e) コンテンツベクトルから興味トピックに対する発話者の興味度を求めるステップ。

【0008】

また、本発明の興味推定装置は、入力された文がどのようなトピックに属するかを事前に推定するための事前学習を行うトピック別分類部と、話者の発話に含まれる単語の系列から得られた単語ベクトル系列から隠れ状態ベクトル系列を作成する単語系列エンコーダと、隠れ状態ベクトル系列から発話ベクトルを作成する単語アテンション部と、発話ベクトルから興味トピックに関するコンテンツベクトルを作成するトピック別発話アテンション部と、コンテンツベクトルから前記興味トピックに対する発話者の興味度を求める興味推定部を、備える。

10

【0009】

本発明の興味推定プログラムは、興味推定方法のステップ(a)~(e)に示される手順をコンピュータに実行させるプログラムである。

【発明の効果】

【0010】

本発明によれば、ユーザの発話からユーザの興味を高精度に推定することが可能になる。

20

【図面の簡単な説明】

【0011】

【図1】本発明の実施の形態例に係る興味推定方法の、ユーザ発話の処理の段階的な工程を示す図である。

【図2】本発明の実施の形態例に係る興味推定方法において、事前学習の工程を説明するための図である。

【図3】本発明の実施の形態例に係る興味推定方法において、実験で使用する24種類のトピックを示す図である。

【図4】本発明の実施の形態例に係る興味推定方法において、興味度を推定するときの実験で使用した実際の対話例を示す図である。

30

【図5】本発明の実施の形態例に係る興味推定方法において、実験で使用した統計情報のテーブルである。

【図6】本発明の実施の形態例に係る興味推定方法において、トピックの数1~24に対応するNDCGの値を、本発明の提案手法と他の手法とで比較して示した図である。

【図7】本発明の実施の形態例に係る興味推定方法の提案手法と他の手法とで、興味度と正解との間の平均二乗誤差を比較して示した図である。

【図8】本発明の実施の形態例に係る興味推定方法において、提案手法のトピック別発話アテンションを可視化した図である。

【図9】本発明の実施の形態例に係る興味推定方法の提案手法とは異なる、事前学習なしのトピック別発話アテンションを可視化した図である。

40

【図10】本発明の実施の形態例に係る興味推定方法において、トピックの数1~24に対応する、被験者数全体と被験者数を制限した場合および発話数を制限した場合のNDCGの値を比較して示した図である。

【図11】本発明の実施の形態例に係る興味推定方法において、学習データ量を100%から順に制限した場合の提案手法の性能評価(平均二乗誤差: MSE)結果を示す図である。

【発明を実施するための形態】

【0012】

以下、図面を参照して、本発明の実施の形態例(以下、「本例」ということもある)を説明する。

50

本発明の実施の形態例では、まず、話者Sが他の話者（雑談対話システムを含む）との対話中に行ったときの、話者Sの発話集合 $U_s = (u_1, u_2, \dots, u_n)$ が与えられる。そして、この話者Sの発話集合 $U_s = (u_1, u_2, \dots, u_n)$ から、ある興味トピック $T = (t_1, t_2, \dots, t_m)$ に対する、話者Sの興味対象の度合い（興味度） $Y_s = (y_1, y_2, \dots, y_m)$ が推定される。ここで、発話集合 U_s の各要素 u_1, u_2, \dots, u_n は、それぞれが異なることなる発話文を意味しており、各要素 $u_1 \sim u_n$ のそれぞれが後述する1つの発話文のベクトル $X = (x_1, x_2, \dots, x_k)$ に相当する。

【0013】

ここで、興味トピックTの個々の要素 $t_i (t_1 \sim t_m)$ は、例えば映画や旅行などに対応する、m個の異なる具体的なトピックである。また、興味度 Y_s の個々の要素である（ y_1, y_2, \dots, y_m ）は、具体的な一つのトピック「 t_i 」に対する話者Sの興味度合い（興味度）「 y_i 」を意味する正解値である。この興味度「 y_i 」は、詳しくは後述するが、3段階で定義され、 $y_i = 0$ は「興味なし」、 $y_i = 1$ は「少し興味あり」、 $y_i = 2$ は「興味あり」とされる。

【0014】

なお、話者Sの発話から興味の推定を行うためには、以下の2つの情報を抽出することが必要である。

(i) その発話はどのようなトピックに関するものであるか

(ii) 話者Sはそのトピックについてどの程度興味を持っているか

以上の2点を効率的に行うため、発明者は、ニューラルネットワークモデルと、このモデルの効率的な学習を行うための事前学習の手法を考えた。

【0015】

図1、図2は、本発明の実施形態例で用いられる、話者Sと他の話者との間の自然言語での雑談対話からの興味推定方法の概要を示す提案モデルの例である。

図1、図2に示すように、本例の興味推定方法は、図1に示す4つの工程と図2に示す事前学習とに分けて説明される。すなわち、図1には、以下の4つの工程が示されている。

(1) 単語系列エンコーダにより発話中の単語ベクトル系列から隠れ状態ベクトル系列を作成する工程、

(2) 単語アテンション（Attention）部により、隠れ状態ベクトル系列から固定長ベクトルである発話ベクトルを作成する工程、

(3) トピック別発話アテンション（Attention）部により、入力された発話ベクトルからトピックごとに各トピックの重要度を示すコンテンツベクトルを作成する工程、

(4) 興味推定部により、トピック別発話アテンション部で作成されたコンテンツベクトルから、各トピックの興味度を求める工程。

【0016】

図1に示す上述した4つの工程の他に、さらに、本例の提案方法では、単語系列エンコーダによる事前学習の工程が必要とされる。

図2は、単語系列エンコーダの出力を、別に用意されたニューラルネットワークに供給して、どの発話がどのトピックに近いかを予め学習させるための事前学習の工程を示す。つまり、話者Sの発話から興味度を推定するには、話者Sの発話がどのトピックに対応しているのか予め知っておくとよい。このため事前学習では、入力文がどのトピックに関係しているかを事前に学習させる。この事前学習を導入することにより、後述するように、興味度の推定値が実際のユーザの興味に近いものとなる。

【0017】

本例の興味推定方法（以下、「本例の提案方法」という）は、単語系列エンコーダ1、単語アテンション（Attention）2、トピック別発話アテンション（Attention）3、興味推定部4の4つニューラルネットワークが順に上述の処理を行い、かつ、図2に示すトピック別分類部5を用いて単語系列エンコーダ1の事前学習を行っている。

以下、単語系列エンコーダ 1、事前学習のトピック別分類部 5、単語アテンション部 2、トピック別発話アテンション部 3、興味推定部 4 の作用、機能等について、項目別に順に説明する。

【 0 0 1 8 】

< 単語ベクトル系列の作成 >

最初に発話を単語に分割し、分割された k 個の単語は、Word2Vecによってベクトルに変換され、k 個の要素を持つ単語ベクトルの系列 $X = (x_1, x_2, \dots, x_k)$ が得られる。ここで、Word2Vecは、「Word to Vector」を略記したものであり、単語からベクトルを生成するツールである。

上述の単語ベクトル系列の作成は、本例の提案方法において前処理として行われるものであるが、「話者が発話した発話を単語に分割し、分割された単語を単語ベクトルに変換することで単語ベクトル系列を作成するステップ」として、また、本例の提案方法を実行する装置としては、「話者が発話した発話を単語に分割し、分割された単語を単語ベクトルに変換することで単語ベクトル系列を作成する単語ベクトル系列作成部」として、本発明に組み入れられても良い。また、これらのステップを以下に記載の単語系列エンコーダの機能の 1 つとして組み入れても良い。

【 0 0 1 9 】

< 単語系列エンコーダ 1 の機能 >

単語系列エンコーダ 1 は、話者 S の発話した単語ベクトルの系列 $X = (x_1, x_2, \dots, x_k)$ を RNN (Recurrent Neural Network) を用いてエンコードする。

本例の提案方法では、RNN の中間層として双方向 GRU (Bidirectional Gated Recurrent Unit) が用いられる。

【 0 0 2 0 】

図 1 の単語系列エンコーダ 1 では、下段に単語ベクトルの系列 X を最初の単語ベクトル x_1 から最後の単語ベクトル x_k の順に読み込む順方向 GRU (forward GRU (GRU ())) が配置され、上段に単語ベクトルの系列 X を最後の単語ベクトル x_k から最初の単語ベクトル x_1 の順に読み込む逆方向 GRU (backward GRU (GRU ())) が配置されている。なお、明細書の中では、表記上の制約のため、「 \rightarrow 」、「 \leftarrow 」を文字の上に付与できないので、例えば GRU () のように、文字の後に括弧で示す。なお、「 \rightarrow 」は順方向の矢印、「 \leftarrow 」は逆方向の矢印を意味する。話者 S の発話から切り出された k 個の単語の単語ベクトル (x_1, x_2, \dots, x_k) のそれぞれは、順方向 GRU と逆方向 GRU に同時に供給される。

【 0 0 2 1 】

図 1 では、順方向 GRU と逆方向 GRU を点線の枠で囲い、点線の枠に単語ベクトル x_i が入力され、点線の枠からの順方向 GRU と逆方向 GRU の合成ベクトル h_i が出力されるように表示されている。順方向 GRU における隠れ状態ベクトル h_i (hidden state h_i ()) は数 1 式により得られる。また、逆方向 GRU における隠れ状態ベクトル h_i (hidden state h_i ()) も、数式は示されていないが、数 1 式と同様の式により計算される。

【 0 0 2 2 】

【数 1】

$$\vec{h}_i = \overrightarrow{GRU}(x_i, \vec{h}_{i-1}) \quad (1)$$

ここで、数 1 式は、単語ベクトル x_i と一つ前の隠れ状態ベクトル h_{i-1} () とを入力とする順方向 GRU によって、次の段の隠れ状態ベクトル h_i () を計算していることを意味する。

【 0 0 2 3 】

本例の提案方法では、目的となる隠れ状態ベクトル h_i は、順方向 GRU と逆方向 GRU にそれぞれベクトル「 x_i 」を入力した際の出力を用いて数 2 式により求められる。

【 0 0 2 4 】

【 数 2 】

$$h_i = [\vec{h}_i : \hat{h}_i] \quad (2)$$

ここで、数 2 式の「{ : }」はベクトルの連結を意味する。単語系列エンコーダ 1 で求められた隠れ状態ベクトル h_i は、次の処理を行うための単語アテンション部 2 に送られる。

【 0 0 2 5 】

< トピック別分類部 5 の機能 (事前学習) >

ここで、図 2 に示されるトピック別分類部 5 の機能と事前学習について説明する。ユーザがどのトピックにどれだけ興味を持つかを推定するためには、まずユーザからの発話がどのようなトピックに関するものかを予め推定することが重要である。

全ての発話に人手でアノテーションする、つまり注釈をつけることも可能であるが、その作業コストは極めて高いものとなる。また、トピックの情報が明示的に与えられない場合には、ユーザの発話情報と発話したユーザの興味度から間接的にトピックを推定しなければならない。そこで、本例の提案方法では、単語系列エンコーダ 1 に対して、事前学習の考え方を導入している。

【 0 0 2 6 】

すなわち、発話中のトピックをより効果的に抽出するため、単語系列エンコーダ 1 とトピック別分類部 5 が、入力文のトピックの分類を行う。これが事前学習 (Pre training) である。

この事前学習では、分類クラスを本例の興味推定方法の対象とする興味トピック Y_s と同一にすることが重要である。つまり、本例の興味推定方法において対象とする興味トピック Y_s と分類クラスを同一にすることによって、発話と興味推定対象トピックの対応関係を推定するための学習コストが下がり、興味推定によりフォーカスされた適切な学習が可能となる。

【 0 0 2 7 】

例えば、図 3 で後述するように、旅行、映画、音楽・・・などのトピックを予め決めて、入力される文がどのトピックに属するかを予め学習させておく。すると、新たに入力される文がどのトピックに近いかを推定しやすくなる。「学習コストが下がる」というのは、このことを意味している。この事前学習により、単語エンコーダ 1 の出力にそれぞれのトピックに関する情報が内包される。これにより、本例の提案手法において、発話とトピックの関連付けが容易になり、興味推定に注力した学習が可能となる。

【 0 0 2 8 】

また、事前学習では、入力文の最後の単語ベクトル「 x_k 」を双方向 GRU に入力する。そして、双方向 GRU の隠れ状態ベクトル「 h_k 」を用いて、入力文の各トピックへの所属確率分布 p を数 3 式で計算する。なお、ここでは、後述する単語アテンションは使用していない。また、最後の隠れ状態ベクトル「 h_k 」以外の隠れ状態ベクトル「 h_i 」は使用していない。

【 0 0 2 9 】

【 数 3 】

$$p = \text{softmax}(W_c h_k + b_c) \quad (3)$$

ここで、 W_c 、 b_c は、事前学習時のみに使用されるトピック別分類のためのニューラルネットワークのパラメータである。ここでは、事前学習の損失関数として交差エントロピー誤差を用いている。

なお、ソフトマックス関数は、それぞれの出力を「0 ~ 1」の間の値に落とし込み、出力された値の合計が「1」になる関数である。

【 0 0 3 0 】

以上、本例の提案手法において、重要な役割を担う事前学習について説明したが、この

事前学習については、図2を参照してさらに説明する。

図2は、単語アテンション部2で用いられるニューラルネットワークとは別の異なるニューラルネットワークに単語系列エンコーダ1の隠れ状態ベクトル「 h_k 」を入力して別の処理を行う事前学習について説明するための図である。

【0031】

図2に示すように、トピック別分類部5には、単語エンコーダ1の出力のうち、最後の段階の隠れ状態ベクトル「 h_k 」が入力される。図1では単語エンコーダ1から k 個の隠れ状態ベクトル「 h_i 」が出力されているが、数3式の所属確率分布 p を求める際には、最終段の隠れ状態ベクトル「 h_k 」だけが用いられる。

すなわち、トピック別分類部5は、数3式で計算した所属確率分布 p から、入力される隠れ状態ベクトル「 h_k 」がどのトピックに近いかを推定する。この事前学習により、入力される発話がどのトピックに近いかを予めニューラルネットワークに学習させることができる。

10

【0032】

<単語アテンション部2の機能>

本例の提案手法では、「全ての単語が求めたい結果に平等に貢献するわけではない」という考えに基づき、単語アテンション部2が導入されている。

単語アテンション部2では、アテンション機構により単語ごとに異なる重み付けを行う。ここでアテンション機構とは、重要度が高い単語に注目させるためのニューラルネットワークの構造をいう。例えば、「私は野球が好きです」という発話の意味を理解するためには、「は」や「です」という単語よりも「野球」や「好き」という単語の方が、重要度が高い。したがって、単語アテンション部2は、「野球」や「好き」という単語に大きな重みを与える。

20

【0033】

すなわち、図1に示すように、単語系列エンコーダ1の出力である隠れ状態ベクトル「 h_i 」は、単語アテンション部2に送られ、単語アテンション部2において、単語ごとに異なる重みづけが行われる。単語アテンション部2で重みづけがされた発話ベクトル「 z 」は、数4式、数5式、数6式によって求められる。

【0034】

【数4】

$$v_{hi} = \tanh(W_{\omega}h_i + b_{\omega}) \quad (4)$$

30

数4式は、単語系列エンコーダ1の出力である隠れ状態ベクトル「 h_i 」を、ニューラルネットワークのパラメータ W 、 b とハイパボリックタンジェント関数を用いて、異なる別のベクトル「 v_{hi} 」に変換することを示している。

【0035】

【数5】

$$\alpha_{hi} = \frac{\exp(v_{hi}^T v_{\omega})}{\sum_i \exp(v_{hi}^T v_{\omega})} \quad (5)$$

【0036】

【数6】

$$z = \sum_i \alpha_{hi} h_i \quad (6)$$

【0037】

数5式の「 v_{hi} 」は、数4式で求めた単語ベクトル「 x_i 」の重要度を示している。ここで、数5式の中の「 v 」は、ニューラルネットワークで用いられる単語の重要度を計算するためのパラメータ（文脈ベクトル）である。「 T 」はベクトル同士の演算を行うときの転置記号である。

50

なお、単語の重要度を計算するための文脈ベクトル「 v 」は、乱数により初期化され、他のパラメータと同様に学習により最適化される。

【0038】

このように、単語アテンション部2は、入力される隠れ状態ベクトル「 h_i 」を基にして変換したすべての「 v_{h_i} 」に対して単一の文脈ベクトル「 v 」を用いて、単語ベクトル「 x_i 」の重要度「 h_i 」を求めている。その意味では、単語アテンション部2は、ニューラル翻訳モデルやニューラル対話モデルのようなデコーダを用いて単語系列を生成する場合の通常のアテンションとは異なる。

【0039】

また、数6式の発話ベクトル「 z 」は、単語ベクトル「 x_i 」を入力した際の隠れ状態ベクトル「 h_i 」と、その単語ベクトル「 x_i 」の重要度「 h_i 」との積をすべての単語ベクトル「 x_i 」（ $i = 1 \sim k$ ）について加算して計算したベクトルである。この発話ベクトル「 z 」は、一つの単語ベクトル系列 $X = (x_1, x_2, \dots, x_k)$ の特徴を表している。

【0040】

<トピック別発話アテンション部3の機能>

トピック別発話アテンション部3は、全てのトピック「 t_i 」（ $i = 1 \sim m$ ）に関するコンテンツベクトル（content vector）「 c_{t_i} 」を求める。以下、コンテンツベクトル「 c_{t_i} 」の生成について説明する。

トピック別発話アテンション部3には、単語系列エンコーダ1及び単語アテンション部2によって、発話集合 $U_s = (u_1, u_2, \dots, u_n)$ からエンコードされた発話ベクトル集合 $Z_s = (z_1, z_2, \dots, z_n)$ が供給される。

【0041】

トピック別発話アテンション部3は、この発話ベクトル Z_s を用いて、 m 種類（ここでは $m = 24$ ）のトピック「 t_i 」に関するコンテンツベクトル「 c_{t_i} 」を生成する。なお、コンテンツベクトル「 c_{t_i} 」は、各トピック「 t_i 」が埋め込まれたベクトル表現であり、数7式、数8式、数9式によって求められる。なお、数8式に用いられる「 v_{t_i} 」は、図1に示すトピックベクトルである。

【0042】

【数7】

$$v_j = \tanh(W_r z_j + b_r) \quad (7)$$

【0043】

【数8】

$$\alpha_{jt_i} = \frac{\exp(v_j^T v_{t_i})}{\sum_j \exp(v_j^T v_{t_i})} \quad (8)$$

【0044】

【数9】

$$c_{t_i} = \sum_j \alpha_{jt_i} z_j \quad (9)$$

【0045】

ここで、数7式から数9式は、単語アテンション部2の機能のところで説明した数4式から数6式と類似している。

すなわち、数7式は、単語アテンション部2の出力である、単語の重要度を表す発話ベクトル「 z_j 」（ $j = 1 \sim n$ ）が、別のニューラルネットワークのパラメータ W_r 、 b_r とハイパボリックタンジェント関数を用いて、異なる別のベクトル「 v_j 」に変換される。

【0046】

10

20

30

50

数 8 式の「 v_{jt_i} 」は、数 7 式で求めた「 v_j 」を用いて計算した、トピック別発話アテンション部 3 のニューラルネットワークから出力される発話ベクトル「 z_j 」の重要度を示している。

【 0 0 4 7 】

ここで、図 1 に示されているトピックベクトル「 v_{t_i} 」($i = 1 \sim m$) は、発話ベクトル Z_s (z_1, z_2, \dots, z_n) の中のそれぞれの発話のトピック別の重要度を計算するためのベクトルである。このトピックベクトル「 v_{t_i} 」は 24 種類 (すなわち、 $m = 24$) があり、学習により最適化される。

【 0 0 4 8 】

また、数 9 式の「 c_{t_i} 」は、各発話ベクトル「 z_j 」($j = 1 \sim n$) と、その発話ベクトル「 z_j 」の取り得る重要度「 v_{jt_i} 」との積をすべての発話ベクトル「 z_j 」について加算したものであり、コンテンツベクトルといわれる。つまり、このコンテンツベクトル「 c_{t_i} 」は、一つの発話の重要度を示している。

【 0 0 4 9 】

このトピック別発話アテンション部 3 のニューラルネットワークに用いられるパラメータ W_r は、24 種類のトピック全てに用いられる共通のパラメータである。

【 0 0 5 0 】

< 興味推定部 4 の機能 >

トピック別発話アテンション部 3 で求められた各トピック「 t_i 」のコンテンツベクトル「 c_{t_i} 」($i = 1 \sim m$) は、興味推定部 4 に供給される。

興味推定部 4 は、このコンテンツベクトル「 c_{t_i} 」を用いて、トピック「 t_i 」に対する興味度「 d_{t_i} 」を数 10 式によって求める。

【 0 0 5 1 】

【 数 1 0 】

$$d_{t_i} = \tanh(W_{t_i}c_{t_i} + b_{t_i}) + 1 \quad (10)$$

【 0 0 5 2 】

ここで、パラメータ「 W_{t_i} 」と「 b_{t_i} 」は、トピック「 t_i 」の興味度推定のための新たなニューラルネットワークのパラメータであり、対象となる 24 種類のトピックごとに異なる。

また、数 10 式の「+ 1」は、興味度「 d_{t_i} 」を、正解である興味度「 y_i 」と同じ「0 ~ 2」にスケールリングするための項である。

なお、学習の際の損失関数には、数 11 式で示されるように、正解である興味度「 y_i 」と数 10 式で求めた興味度「 d_{t_i} 」の平均二乗誤差 L を用いることとした。

【 0 0 5 3 】

【 数 1 1 】

$$L = \frac{1}{m} \sum_i^m (y_i - d_{t_i})^2 \quad (11)$$

【 0 0 5 4 】

本例の興味推定方法は、機械学習を利用するものであることから、事前学習を含む学習フェーズと学習結果を利用した推定フェーズとに分かれる。学習フェーズでは、学習データ (ユーザの発話とユーザの各トピックについての興味 (正解) のデータセット) に基づく学習を行うことで単語系列エンコーダ 1、単語アテンション部 2、トピック別発話アテンション部 3 及び興味推定部 4 のそれぞれにおいて各種パラメータが決定され、推定フェーズでは、学習フェーズで決定された各種パラメータを固定したうえで、テストデータ (ユーザの発話のみ) に基づいてユーザの興味の推定を行う。

本例の興味推定方法による処理を以下、実験例で示す。

【 0 0 5 5 】

10

20

30

50

< 実験例 >

以下、本例の興味推定方法（提案方法）の性能評価のために、2名の被験者が1対1で対話したチャットログから推定した興味度「 d_{t_i} 」と、被験者が事前にアンケート回答した正解の興味度「 y_i 」との相関性を評価した実験例について説明する。

この実験では、「機械学習用ライブラリチェイナ（Chainer）」を用いて評価した。なお、チェイナ（Chainer）はニューラルネットワークの計算および学習を行うためのオープンソフトウェアライブラリである。

すなわち、本例の提案方法では、図1に示すように、単語系列エンコーダ1、単語アテンション部2、興味推定部4のほかに、新たにトピック別発話アテンション部3を組み込んで、実験用モデルを構築した。

【0056】

対話を行う前に、各被験者には、図3に示す24種類のトピックそれぞれに対して、自分の興味の度合いを0（興味なし）、1（やや興味あり）、2（興味あり）の3択で選択してもらった。なお、図3のトピックは、Yahoo知恵袋（登録商標）から、被験者2人の初対面の対話で取り上げられやすい話題を中心に設定した。

【0057】

対話は1回1時間とし、スカイプ（Skype：登録商標）のインスタントメッセージングを介して行った。対話を始めるに当たって、各被験者には「話の中で相手と自分の間で共通して興味のあることを見つけ、それについて話を膨らませていくことを心がけてください。」とだけ教示し、話すべき内容・話題についての指示は行わなかった。

【0058】

図4は、被験者2人の実際の対話例を示す。ここでは、収集したすべてのチャットデータ中の発話を被験者別に分類し、1被験者が行った全ての発話とその被験者のアンケートの結果を1データとして扱った。図5は、データの統計情報を示したものである。

【0059】

単語ベクトルの獲得は、Word2Vecを用いて、次元数200、ウィンドウサイズ5、最少出現頻度10とし、約100GBのTwitter（登録商標）データで学習を行った。

【0060】

単語系列エンコーダ1の入力層の次元は200、双方向GRUの出力の次元は400、双方向GRUは1層とした。また、単語アテンション部2とトピック別発話アテンション部3は、入出力ともに400次元、興味推定部4の入力は400次元、出力は1次元とした。トピック別分類部5の入力は400次元、出力は24次元である。

【0061】

なお、すでに説明したように、事前学習として行うトピック別分類の際には、トピック別分類の決定に用いたのと同じ、図3に示したトピックを使用した。各トピックに対応する質問文および回答文はYahoo知恵袋データ第2版から抽出した。77万文を学習データ、各トピック100文ずつの2400文をテストデータとした。事前学習の結果、テストデータのトピック分類正確率は0.755であった。

評価は、10分割交差検証により行った。なお、対話単位ではなく、データ単位（被験者単位）で交差検証を行ったため、学習データとテストデータの間で、同一の被験者による発話が使用されることはない。

【0062】

本例の提案方法を、対話システムの話題選択に用いる場合、対話システムがその時点で対応できる話題の中から、より望ましい話題を選択することが重要である。そこで、本実験例では、推定した興味度「 d_{t_i} 」に基づいて、各トピックの順位付けを行った。そして、ユーザが興味を持つトピックを上位に位置付け、興味を持たないトピックを下位に位置付けるように性能評価を行った。順位付けの評価尺度としては、情報検索の分野で広く使用されている指標であるNDCG（Normalized Discounted Cumulative Gain）を用いた。つまり、NDCGは、正規化されたDCGで、0～1の値をとり、1に近いほど順位の正しい予測結果となる。NDCGは、以下の数12式、数13式により計算

10

20

30

40

50

される。

【 0 0 6 3 】

【 数 1 2 】

$$NDCG_K = \frac{DCG_K}{IDCG_K} \quad (12)$$

【 0 0 6 4 】

【 数 1 3 】

$$DCG_K = rel_1 + \sum_{i=2}^K \frac{rel_i}{\log_2 i} \quad (13)$$

【 0 0 6 5 】

数 1 2 式で、IDCG (Ideal DCG) は順位リストの適合度が順に正しく並んだ場合の DCG、つまり理想とされる DCG である。IDCG_K は、上位 K 位までを考慮した IDCG であり、数 1 2 式の NDCG_K は、DCG_K は IDCG_K で正規化した値を示している。

【 0 0 6 6 】

数 1 3 式の「rel_i」は、i 番目に順位付けされたターゲットに対する適合度を意味している。つまり、「rel_i」の適合度が高く、K の値が大きくなるに従って適合度が小さくなる。「rel_i」を「log₂ i」で割り算しているのは、「i」が大きくなるほど、適合度を小さくすることを意味している。

本例の実験例では、各トピックに対するユーザ「s」の興味度 Y_s の値 (0、1、2) を適合度とした。

【 0 0 6 7 】

本実験例では、初期パラメータを変更して、5 回の 10 分割交差検証を実施した。そして、NDCG が最良の結果になったものを各手法の結果として採用した。ここで、NDCG₁ を選択の基準として用いるのは、NDCG₁ が大きな値であった場合、以降の K において NDCG が高くなる傾向が観察されたためである。

【 0 0 6 8 】

また、各トピックに対するユーザの興味度を正しく推定できるかを評価するため、本例の提案方法とその他の手法で出力した興味度「d_{t i}」と正解「y_i」との平均二乗誤差 (数 1 1 式) によって評価した。

【 0 0 6 9 】

図 6 は、本例の提案方法とその他の手法とで計算した K = 1 ~ 24 までの NDCG を示した実験結果である。すなわち、図 6 に示すように、本例の提案方法だけではなく、比較手法として本例の提案手法の一部を変更した 3 つの比較例による実験を行った。

1 つ目の手法は、本例の提案手法における事前学習を行わないで、単語系列エンコーダ 1 の初期値を乱数として学習を行った「事前学習なし」である。この「事前学習なし」では、図 2 で説明した事前学習がないだけで、図 1 に示す単語系列エンコーダ 1、単語アテンション部 2、トピック別発話アテンション部 3 および興味推定部 4 のすべての処理を実施した。

【 0 0 7 0 】

2 つ目の手法では、本例の提案方法における図 2 に示した事前学習は行ったが、トピック別発話アテンション部 3 によるコンテンツベクトルの取得をせずに、他の手法 (発話ベクトルの Max Pooling) の結果を用いてコンテンツベクトルを取得して興味推定を行った。図 6 では、これを「トピック別発話アテンションなし」として示している。

【 0 0 7 1 】

3 つ目の手法は、本例の提案手法における事前学習を行わず、かつ発話ベクトルの Max Pooling の結果を用いてコンテンツベクトルを取得して興味推定を行う手法であり、図 6

20

30

40

50

では、「事前学習なし&トピック別発話アテンションなし」として示した。

【0072】

また、以下の2つのベースライン手法についても比較の実験を行った。1つ目のベースライン手法は、Twitterユーザの興味推定手法として提案されたもので、図6では「トピック頻度」として示されている。これは、ユーザのツイートに含まれる固有表現を取得し、その頻度ベクトルによりユーザの興味を推定するものである。

【0073】

2つ目のベースライン手法は、S V R (Support Vector Regression)を用いてユーザの興味度を推定する手法である。素性としては、ユニグラム (unigram) とバイグラム (bigram) が用いられる。ここでは、ユニグラム (unigram) だけを用いた場合と、両方を用いた場合で実験を行った。このS V Rモデルでは、トピックごとに個別に学習を行い、各トピックの興味度の推定を行った。具体的には、S V Rとしては、「scikit learn」と呼ばれる機械学習ライブラリを用い、ガウシアンカーネルを用いた。

10

【0074】

図6に示すように、N D C Gの計算に用いる $K = 1 \sim 24$ のトピックについてN D C Gの値をプロットした。図6から、本例の提案方法が、他の手法と比べて、すべての K において、N D C Gが高いことが分かる。

本例の提案手法の次には、トピック別発話アテンションを適用しないで、事前学習を適用した比較例のN D C Gが、良い結果だった。この結果から、トピック別発話アテンションはN D C Gの向上に一定の役割を果たしている、つまり有効であると考えられる。

20

【0075】

次に、本例の提案方法と事前学習なしの例を比較すると、本例の単語系列エンコーダ1、単語アテンション部2、トピック別発話アテンション部3および興味推定部4のすべてを適用しても、事前学習をしない場合には、本例の提案手法に比べて、N D C Gの値が大幅に少なくなることが分かった。すなわち、本例で提案した事前学習の手法は、性能向上に極めて有効であることが理解された。

【0076】

また、本例の提案方法で用いたトピック別発話アテンションを行なったが、事前学習を行わなかった比較例と、トピック別発話アテンションおよび事前学習の両方とも行わなかった比較例では、あまりN D C Gの変化が見られなかった。すなわち、トピック別発話アテンションは事前学習を行わないとN D C Gの向上に貢献しないことが分かった。このことから、本例で提案する事前学習の有効性を確認することができた。

30

【0077】

結論としては、事前学習とトピック別発話アテンションを同時に適用した本例の提案手法を使うことで、トピック別発話アテンションがない場合と比べて性能が向上している。このことから、本例の提案手法のように、事前学習を行った上で、トピック別発話アテンションを適用することが重要であることを確認することができた。これは、トピック別分類の事前学習を行うことで、発話ベクトルにトピックの情報が埋め込まれるため、数8式のトピックベクトル「 v_{t_i} 」が効率的に学習可能になったためと考えられる。

【0078】

S V Rに関しては、ユニグラム (unigram) とバイグラム (bigram) を用いる場合よりも、ユニグラム (unigram) を単独で用いた方が良いN D C G性能を示した。しかし、本例の提案手法のようなニューラルネットワークベースの結果と比べるとはるかに劣る結果となった。

40

また、トピック頻度のベースラインは、 $K = 1$ のときに、S V Rよりも高いN D C Gを示した。これはユーザが積極的に発言したトピックに対して、興味ありと正しく推定できたためであると考えられる。このトピック頻度のベースラインも、本例の提案手法や他のニューラルネットワークベースと比較すると劣る結果となり、トピック頻度のベースラインは、 $K = 4$ では、すべての手法の中で最低のN D C Gになった。

【0079】

50

図7は、各手法によって出力された興味度と正解との間の平均二乗誤差を(MSE)を計算して示した図である。ただし、図7では、興味度を直接出力しないトピック頻度ベースラインは除いている。図7に示すように、本例の提案手法では「0.533」と最も小さい値になっており、このことから、本例の提案手法によれば、適切な興味度の推定が可能であることが理解できる。言い換えると、本例の提案方法による平均二乗誤差(MSE)が最小になっていることから、本例の提案手法は他の手法と比べて最高の性能を示したといえることができる。

【0080】

一方、ニューラルネットワークベース手法の中では、「事前学習なし」が最も性能が低いことから、事前学習とトピック発話アテンションを同時に行うことが重要であることが、図7からも理解することができる。

10

【0081】

<トピック別発話アテンションの可視化による分析>

上述したように、本例で提案した、トピック別発話アテンションと事前学習を同時に適応することにより、良い性能(NDCG)とMSEが得られることが判明した。

そこで、次に、トピック別発話アテンションの分析を行うため、数8式で求められる発話文の重要度「 j_{ti} 」の値をトピック別に計算し、これを可視化した。図8は、事前学習を行った本例の提案手法によるトピック別発話アテンションを可視化したものである。これに対して、図9は事前学習を行わないで、トピック別発話アテンションを可視化したものである。図8、図9とも濃い色であるほど発話文の重要度「 j_{ti} 」が高いことを示している。

20

【0082】

図8に示すように、本例の提案手法では、右側に記載した文(1)から(8)の内容に対応するトピックの重みが高い値(濃い色)になっていることが分かる。図9は、事前学習なしのトピック別発話アテンションを可視化したものである。図9では、発話文とトピックの重みとがほとんど関係していないことが分かる。例えば、文(5)は、興味推定にかかわる文ではないため、「事前学習あり」の図8では、すべてのトピックの重みが小さな値になっている。これに対して、図9では、文(5)に対して、映画、ゲーム、スポーツ、健康が比較的高い値になってしまっている。同様に文(7)、(8)でも、文の内容に関連が薄いトピックが大きな値をとっていることが分かる。

30

【0083】

<学習データ量の分析>

ニューラルネットワークモデルの学習において、学習データの量は、モデルの性能を左右する極めて重要な要素の一つである。本例の提案方法を検証する実験(以下、「本実験」と称する)では、163名の被験者による408対話を収集したが、データとしては必ずしも十分な量ではなかった。したがって、学習データを追加することで、さらなる性能向上も期待できると考えられる。特に、学習データを追加するにあたり、1被験者当たりの発話数を増やすのか、もしくは被験者の数を増やすのかという収集方針の決定は、効率的なデータの収集という点で重要である。

【0084】

そこで、本例の提案手法における学習データの量と性能の関係を分析するに当たって、学習データとして用いる被験者数および発話数を制限して10分割交差検証を行い、性能の変化を分析した。なお、10分割交差検証の際のテストデータとしての被験者数・発話数は制限していない。

40

【0085】

図10は、10分割交差検証を行ったときのNDCGの結果を示したものである。図10では、学習に使用する被験者数を全体の10%、50%、90%に制限した結果と、1被験者当たりの発話数を、発話数全体の10%、50%、90%に制限した結果と、全データの結果(図6の提案方法と同じ)を示した。

図10から、発話数を制限するよりも、被験者数を制限した場合の方が、性能が落ちて

50

いることが分かる。特に、被験者数 10% の NDCG は、性能が著しく低下している。一方、発話数を制限した場合には、全データを使用した場合と比べ、性能が低下しているものの、その影響は比較的小さいことが分かった。

【0086】

図 11 は、被験者数と発話数の学習データへの使用率を 10% から 100% まで 10% 刻みで変化させた場合の平均二乗誤差をプロットしたものである。図 11 から分かるように、図 10 の NDCG の結果と同様に、発話数を制限するよりも、被験者数を制限する方が、平均二乗誤差 (MSE) が大きくなっている。つまり、発話数を制限するよりも、被験者数を制限する方が、性能の低下が著しいことが分かる。

以上から、学習データを追加する場合には、1 被験者当たりの発話数を増やすよりも、被験者数を増加させることが性能向上に寄与することが確認された。

10

【0087】

以上説明したように、本実施の形態例では、雑話対話におけるユーザの発話を入力として、事前に設定したトピックに対するユーザの興味を、ニューラルネットワークモデルを使って推定する興味推定方法を提案した。この提案手法では、双方向 GRU を用いた単語系列エンコーダと、単語アテンションによりユーザの発話をエンコードし、発話ベクトルの集合を得る。次に、この発話ベクトル集合からトピック別発話アテンションを用いて、推定対象の各トピックに対応するベクトルを生成する。そして最後に、このベクトルから対応するトピックに対する話者 (ユーザ) の興味の度合いを推定する。

【0088】

さらに、学習時には、単語系列エンコーダに対して、トピック別分類を行う事前学習を実施させてから、興味推定を行うようにした。

実験の結果は、すでに説明したとおり、本例の提案手法がベースライン手法に比べて、よりの確に興味度を推定できることが確認された。また、本例で提案したトピック別発話アテンションとトピック別分類の事前学習を併用することが有効であることも確認された。

20

【0089】

以上、本発明の実施の形態例について詳細に説明したが、本発明は、ここで説明した実施の形態例に限定されるものではなく、特許請求の範囲に記載された本発明の要旨に逸脱しない限りにおいて、その他の応用例および変形例を含むことは言うまでもない。

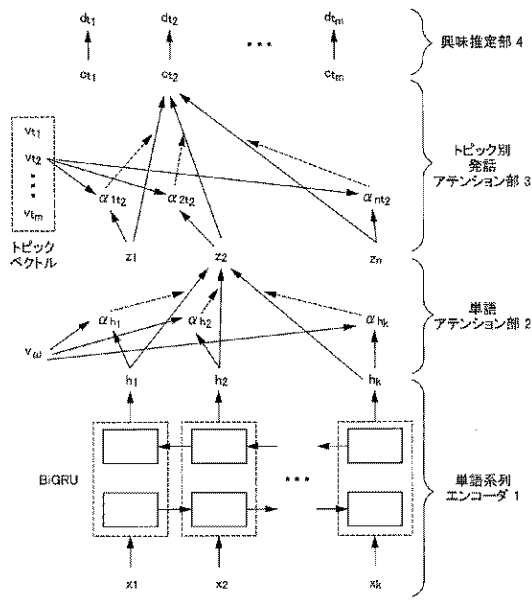
30

【符号の説明】

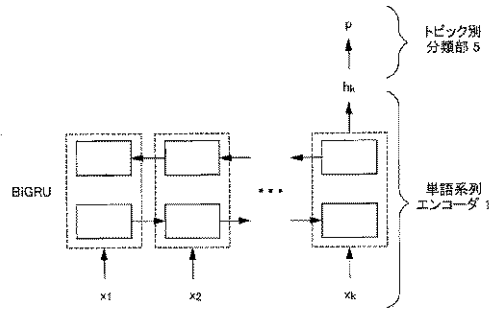
【0090】

1・・・単語系列エンコーダ、2・・・単語アテンション部、3・・・トピック別発話アテンション部、4・・・興味推定部、5・・・トピック別分類部

【 図 1 】



【 図 2 】



【 図 3 】

旅行	映画	音楽
アウトドア	学校	家電製品
芸能人・タレント	PC 関連	ビデオゲーム
動物・生き物	住まい	家事
家族・夫婦	結婚・恋愛	健康
美容	ファッション	料理・グルメ
スポーツ・運動	読書	車・乗り物
政治・経済・社会	アニメ・漫画	歴史

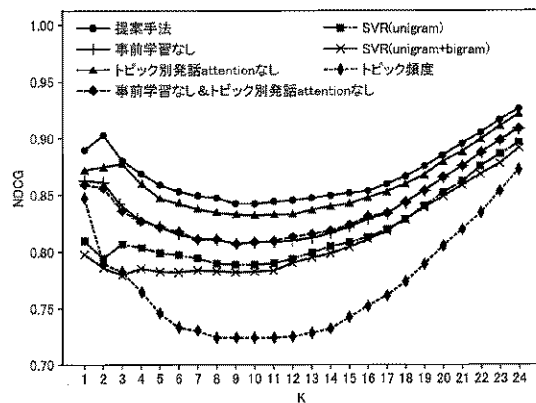
【 図 5 】

ユーザ数(データ数)	163
対話数	408
発話数	48029
「興味あり」の平均トピック数	11.48
「少し興味あり」の平均トピック数	7.30
「興味なし」の平均トピック数	5.21

【 図 4 】

A	対話を開始します。よろしくお願いします。
B	よろしくお願いします。
A	私は、旅行に行ったり、暇な時間にカフェ巡りしたりするのが好きです。趣味は何ですか？
B	私もよくカフェ巡ります。
A	古民家隠れ家風なおしゃれカフェ見つけたら気分が上がりますよね。コーヒー派ですか？
B	わかります！私はカフェで甘いものを注文することが多いので、コーヒー派ですね。
A	私もです！ちよっと前はパンケーキでした。最近、パフェが人気らしいですね。どんなスイーツが好きですか？
B	私もパンケーキやケーキをよく食べます。パフェが流行ってるとは知りませんでした！パフェって大きくてなかなか完食できないイメージがあります。

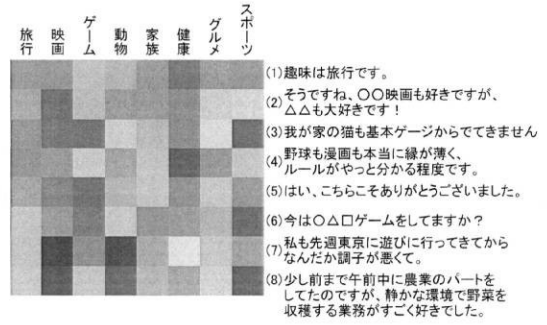
【 図 6 】



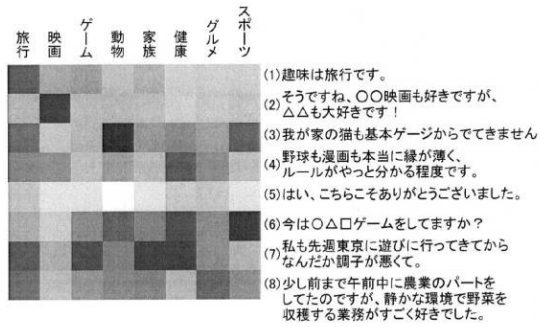
【 図 7 】

提案手法	0.533
事前学習なし	0.580
トピック別発話attentionなし	0.561
事前学習なし&トピック別発話attentionなし	0.568
SVR(unigram)	0.622
SVR(unigram+bigram)	0.632

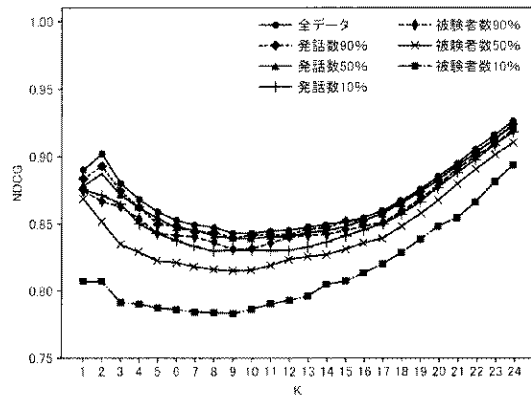
【 図 9 】



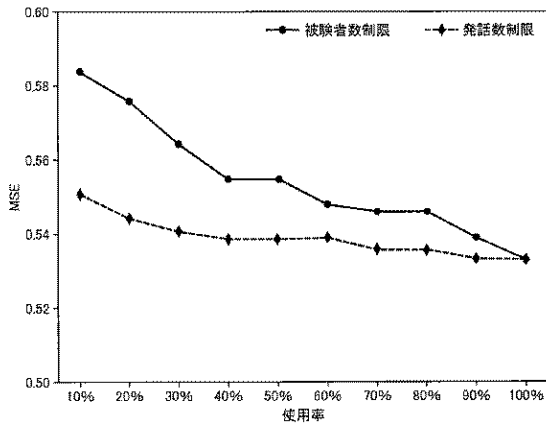
【 図 8 】



【 図 10 】



【 図 11 】



フロントページの続き

特許法第30条第2項適用申請有り 平成30年7月9日公開 19th Annual Meeting of the Special Interest Group on Discourse and Dialogue公式サイト(<https://www.sigdial.org/files/workshops/conference19/index.html>) 平成30年7月12日開催 KTH Royal Institute of Technology主催「19th Annual Meeting of the Special Interest Group on Discourse and Dialogue Melbourne」平成30年11月21日開催 一般社団法人人工知能学会主催 人工知能学会 言語・音声理解と対話処理研究会 第84回研究会「第9回対話システムシンポジウム」平成31年3月1日発行 一般社団法人人工知能学会刊行「人工知能学会論文誌 2019年34巻2号」