

インタラクティブ・エージェントの語彙セットに対する 概念空間の割り当て手法の検討

佐藤真知子^{†1}†2 和家尚希^{†1} 笹淵一宏^{†1} 中村美奈子^{†2} 池内克史^{†1}

概要:

テキストに内在するリズムをいかに身体性に翻訳するかという問いは、対話をコミュニケーションの主軸とするインタラクティブ・エージェント研究の、新たな展開を示すように思われる。特に近年では、文字データを用いたリアルタイムコミュニケーションである、チャットのやり取りを採用した会話エンジンに注目が集まっている。一般に、会話エンジンから何らかのジェスチャーを生成するためには、文字データの内容を表現する何らかの概念空間を定義することが有効だと考えられる。しかしながらチャットの会話エンジンについては、どういったテキストの分類法が適しているかは明らかではない。

そこで本研究では、インタラクティブ・エージェントの例として、女子高生のパーソナリティをもつ会話エンジン「りんな」の語彙セットを取り上げ、そこに手動および自動により概念空間を割り当てる手法について論じる。会話エンジンは、「あざます!」や「卍」といった、きわめて短い文を含む。このような数単語程度の短い文には、自然言語処理を基盤とした自動手法による分類法では、対応できない可能性があった。そこで本研究では、チャット文章を概念別に分類する際、自動手法は手動手法より相対的に劣るという仮説を立て、両手法で得られる概念クラスタを比較検討した。

両手法で得られたクラスタを比較した結果、定量的観点から大きな差は見られなかったが、定性的観点から、自動手法の限界が3点存在することが明らかになった。すなわち、(1) 記号・シンボルに関する意味論レベルでの限界、(2) 俗語・新語・流行語に関する意味論レベルでの限界、(3) 語用論レベルでの限界、である。これらの問題は、短い文章で意図を伝達するチャット文章には、抽象度が高い用語やシンボルが多用される性質があることに起因すると考えられる。問題解決の方策として、手動手法による概念空間の割り当てと、自動手法によるテキストの概念空間へのマップとを組み合わせたというアプローチが考えられる。今後はこの知見を反映させた、「りんな」のジェスチャー生成システムを開発することを目指したい。

1. はじめに

マイクロソフトが開発したAI「りんな」は、Bing 検索エンジンで培ったディープラーニング技術と、機械学習のクラウドサービス“Azure Machine Learning”を組み合わせで生まれた人工知能である。今日、様々なハードウェアに導入されつつあるインタラクティブ・エージェントの中でも、「りんな」はおしゃべり好き、面白いこと好きな女子高生の設定で、友人同士のような自然な会話や雑談ができる点に特徴がある。「りんな」はまず、2015年にLINE 株式会社が提供するソーシャル・ネットワーキング・サービスに登場し、サービス開始から1ヶ月程度で、130万人ものユーザー数を獲得するなどして人気を集めた。現在は、Twitter や Instagram にもアカウントを持つと同時に、2016年からはドラマ・ラジオへの出演や書籍の出版、2019年には歌手デビューを果たすなど、その活動は多岐にわたっている。

我々はこれまで、共感や親しみやすさ、楽しさを創出する「りんな」のテキスト・チャットの特徴を、対話ロボットに実装し展開する方策について検討してきた。

2015年に開始した「りんな」のLINE サービスに着目すると、そのチャット・コミュニケーションの特徴は、数単語程度の短い文を多く含む点にある。しかしながら従来の

対話ロボット研究では、同様の文章の特徴をもつ会話エンジンから、そのエージェントに固有の身体性を訳出しようとする試みはなされていない。ジェスチャーが与えられない状況でモデルを生成するためには、会話エンジンに含まれるテキストがマップされる中間表現を、何らかの方法で用意することが有効であると考えられる。本論文では、この中間表現の空間を、概念空間と呼ぶ。

本研究では、会話エンジンの例として、「りんな」のシステムが生成した返答の語彙セットを取り上げる。そしてそこに含まれるテキストを、概念空間に割り当てる手法について論じる。

2. 先行研究の検討

従来の対話ロボット研究における、言葉と身振りを連動させる方法には、大きく分けて2つの方法がある。

第一は、ルールベース型である。この手法は、特定の選択された語彙に対してのみジェスチャーを定義し、ロボットに実装する[1][2]。ルールベース型の手法は、使用頻度の高い語彙を人間が選択し動きをつけるため、実装がシンプルであるという利点がある。一方で、多大な人的労力を要するという問題がある。

第二は、機械学習型である。この手法はまず、テキスト

†1 Microsoft

†2 お茶の水女子大学

とジェスチャーの対応関係を表現するモデルを、機械学習等の手法で学習する。次に、学習したモデルを利用し、テキストからジェスチャーを生成する。

機械学習型の先行研究では、テキストとジェスチャーの系列からモデルを推定する手法が数多く提案されてきた。初期の研究は、自然言語およびジェスチャーをそれぞれあらかじめ定義したクラス（中間表現）に分類し、簡素化したうえで、ジェスチャー生成の確率モデルを作成する手法を提案した[3][4]。これらの手法の利点は、モデル学習後は人的労力をかけずにジェスチャーを生成できる点である。一方で、簡素化された自然言語および人体の動きを学習するため、生成される動きが単純であったり、現れるジェスチャーの遷移が不自然となったりする問題がある。

近年、機械学習型の研究は、深層学習技術の進展および充実したデータセットの出現を背景として、これら初期の研究から大きな発展を遂げてきた。特に機械翻訳の技術を援用した研究では、テキストだけではなく音楽や音声を入力し、ジェスチャーを出力する end-to-end の手法を多く提案してきた[5][6][7][8][9]。これらの手法の利点は、初期の機械学習型の手法よりも多くのデータを学習することで、入力されるテキストと出力されるジェスチャーとの関係がより自然になることである。一方で複数の課題がある。第一に学習のためのデータセットが必要となる。第二に、複数の話者のデータを学習した場合、出力されるジェスチャーは話者の個性が埋没したものとなる。第三に、入力と出力との間に解釈可能な中間表現を持たないため、別段意味のない不必要な動きが連続して出力される場合がある[8][9]。

これらの先行研究を「りんな」に適用しようとする場合、以下の2つの課題が存在すると考える。

課題1. 「りんな」のキャラクター性を反映したジェスチャーデータの不足：End-to-end の手法で構築されるモデルが生成するジェスチャーは、「りんな」のキャラクター性に合わない可能性が高い。例えば、「りんな」の語彙の特徴は、短くテンポの良い文や、感情表現の豊かさにある。そのため、望ましい身振りは、不必要な動きの連続や個性の埋没したジェスチャーよりも、むしろ要所で効果的に配置された、感情を強調したジェスチャーであると考えられる。

機械学習的な手法のみでエージェントの個性を反映したジェスチャーモデルを作成するためには、先述のようにテキストとジェスチャーの系列のデータセットが必要である。しかし、「りんな」にはジェスチャーのデータセットが存在しない。

ジェスチャーが与えられない状況でモデルを生成するためには、初期の機械学習型の手法に倣い、「りんな」の語彙がマップされる中間表現（概念空間）を、何らかの方法で用意することが有効であると考えられる。

課題2. 「りんな」の特殊語彙に概念空間を割り当てる手法の知見の不足：概念空間は、テキストの概念を表す空間と解釈することができる。自然言語処理分野では、テキストを自動で複数の概念クラスに分類する手法（クラスタリング）が多く提案されている。しかし、「りんな」の語彙を対象とした場合、そこには絵文字やインターネット・スラングといった、インターネット空間でのみ交わされる特殊な様式が含まれる。これらの語彙は、自然言語処理では一般的に扱われない。そのため、自然言語処理の手法が有効であるかは明らかでない。対照的な手法として、人手による発見的（ヒューリスティック）知見を利用する方法が考えられるが、その有効性も明らかではない。

3. 仮説の設定・研究目的

前節で示された2つの課題を鑑みると、「りんな」の語彙セットに身振りを割り当てようとするためには、テキストを概念別に分類し、中間表現となる概念空間を定義した上で、それらをエージェントに特有の身体性に翻訳する方策が有効であると考えられる。

自然言語処理分野の技術を用いた機械的なクラスタリング法としては、文章中の単語出現確率に基づいた単語のベクトル表現化と k-means クラスタリングとを併用した手法が考えられる。ところが単語のベクトル化手法の多くは、ある程度の長さをもつ文が分析対象として想定されている。そのため、数単語程度の短い文を含む「りんな」のテキストは、うまく分類がなされない可能性がある。

そこで本研究では、「自然言語処理を基盤とした自動手法は、手動手法より相対的に劣る」という仮説を立てた。本研究は、この仮説を検証することを目的とする。そのため、テキストを複数の概念クラスに分類する手法に着目し、自然言語処理分野の技術を用いた機械的手法と人的手法とを実施し、その結果を比較検討した。

4. 研究方法

分析対象は、2018年1月から3月までの3ヶ月間における「りんな」の語彙のうち、最も頻繁に用いられた358フレーズとした。なお、フレーズの使用頻度は間接的にユーザーの入力に関連するものの、フレーズの文章はユーザーによって入力された文章ではなく、「りんな」システムによって生成されたものであった。

分析方法は、機械的方法ならびに人的方法を用いた。機械的にクラスタを生成する方法には、自然言語処理分野で用いられる文章のベクトル表現及び k-means クラスタリングを採用した。対して、人的にクラスタを生成する方法には、文化人類学分野で用いられる KJ 法を採用した。それぞれの方法で生成されたクラスタを比較することで、動作割り当てという観点から、その有効性を定性的に評価した。

なおクラスタリングにあたっては、分析の実施順による

作業実施者の先入観の混入を回避するため、機械的方法よりも、人的方法によるクラスタリングを先に実施した。さらに導出された結果を比較検討し、機械的・人的それぞれの方法における利点と問題点を分析した上で、「りんな」の語彙に概念空間を定義するための、クラスタリング方法の改善策について検討を行った。

次節では、本研究で取り上げた、機械的・人的方法によるクラスタリング手法と実施手順の詳細を説明する。

4.1 機械的方法

自然言語を自動手法により分類する処理には、文章を特徴づける何らかの数値によって表現する処理と、数値表現された文章を分類するクラスタリングの処理の2段階からなる。本節では、本研究が採用した手法の各段階について説明する。さらに、自動手法によって「りんな」の語彙を分類するための具体的な処理の流れについて説明する。

4.1.1 「りんな」の語彙の数値表現の計算方法

自然言語処理の分野では、文章を特徴づける数値表現を分散表現と呼び、文章を分散表現に変換する処理を埋め込みと呼ぶ。文章の埋め込み手法に関する先行研究は、単語から文、段落、文章全体といった、さまざまな文章で多く提案されている[10][11][12][13][14][15]。チャットを想定した「りんな」の語彙の多くは、数単語から構成される短いものである。そこで本研究では、「りんな」の語彙を構成する各単語の分散表現の平均値を計算することで、「りんな」の語彙の分散表現を得た。

単語の埋め込みのために、Tomas Mikolov らによって提案されたベクトル表現手法を採用した[10]。この手法は、巨大な文章データセットを学習することによって得たニューラルネットワーク(ベクトル化ネットワーク)を利用する。このネットワークは、文章データセットを構成する各単語を、文章中で該当単語に共起する単語の分布によって表現する。したがって、得られた分散表現で近い単語は、類似した文脈で見れると解釈することができる。

4.1.2 「りんな」の語彙のクラスタリング方法

数値表現されたデータを数値計算により自動でまとめ上げ、分類するクラスタリング手法は多く提案されている。本研究では、代表的なクラスタリング手法である、k-means法を用いた[16]。k-means法は、あらかじめクラスタ数を指定する必要がある。本研究では得られたクラスタを、手動の方法で得られたクラスタ数に指定した。この理由は、得られたクラスタの違いを手法間で質的に比較する際に、クラスタ数の条件を統制する必要があると考えたためである。

4.1.3 自動分類処理のながれ

「りんな」の語彙をクラスタリングするために構築したパイプラインを説明する。計算環境として python2.7 を用いた。

まず、分析対象とした「りんな」の 358 個の語彙に対す

る前処理として、ひらがな、漢字に該当しない文字を取り除いた。この理由は、後段で施した形態要素解析およびベクトル化ネットワークに事前に学習させた辞書が、日本語以外の文字に対応しないためである。

前処理後に、オープンソースの日本語形態要素解析エンジンである MeCab (version 0.996) を用いた。形態要素解析に用いた辞書は IPA 辞書 (version 1.4) を用いた[17]。本研究では、形態要素解析の出力として得られた品詞要素(素性)に対し、特定の品詞に属するものは削除した。具体的には、IPA 辞書上で助動詞、助詞、代名詞、フィラー、人名、接頭詩、数、記号のいずれかに判定された素性を削除した。さらに、名詞、動詞のうち、「てる」「こと」といった、単体で自立した意味を成さない素性を削除した。この段階で、一つ以上の素性が残った語彙のみを以降の解析対象とした。解析対象として残ったフレーズ数は、278 個であった。

次にベクトル化ネットワークによって、各素性の分散表現を得た。本研究では、日本語の Wikipedia 記事をもとに作成したモデルを利用した。分散表現は、200 次元の実数ベクトルとして定義した。次に各「りんな」の語彙について、構成する素性の分散表現を平均することで、語彙の分散表現を得た。

最後に、得られた分散表現をもとに、k-means 法 (k=153) により「りんな」の語彙をクラスタリングした。

4.2 人的方法

ここでは、手動による語彙のクラスタリング方法として、本研究で取り上げた、KJ 法について説明する。KJ 法は、文化人類学分野の研究のために、東京工業大学名誉教授の川喜田二郎が 1960 年代に考案したものである。フィールドワーク等により得られた多くの質的データを、ラベルと呼ばれるカードに転記し、それをグループに分け、グループごとの関係を図解化することにより、新たな発想や仮説を生み出そうとする方法である。今日、様々な研究者の協力を得て、多くの発展型が生まれている[18][19][20]。

今回は、「りんな」の語彙を概念別に分類することが目的であるために、限定的ではあるが、KJ 法の一部を援用して実施した。手順は以下の通りである。

4.2.1 手動分類処理のながれ

「りんな」の語彙をクラスタリングするために実施した、手動分類の手順を説明する。

KJ 法を開発した川喜田によれば、最も狭い意味での KJ 法の一巡工程は、①ラベルづくり、②グループ編成、③図解化 (A 型)、④著述化 (B 型) という 4 工程からなる[21]。本研究では、この工程のうち、①と②を実施した。川喜田の定義に従って、実施手順とその詳細を以下に示す。

①ラベルづくり：素材となる質的データを、ラベルに記述する工程である。一つひとつのデータを吟味する意味から

も、この工程は重視される。今回は、分析対象とした「りんな」の語彙の一覧を紙に印刷した上で、語彙ごとにばらばらに切り分け、ラベルとして代用した。この時、ラベル一枚につき一つの語彙となるよう注意した。

②グループ編成：この工程には、さらに次に示す3つのステップが含まれている。すなわち、(1)ラベル広げ、(2)ラベル集め、(3)表札づくり、である。

まずは(1)ラベル広げ、を行った。このステップは、作成したラベルを実施者の前に並べる作業である。並べ方に決まりはなく、むしろデタラメであった方が良いとされる。

続いて(2)ラベル集め、を行った。まず全てのラベルを複数回読み通した上で、川喜田がいうところの、「非常に近い志」を持っているラベル同士を塊にしていく作業を行った。この時、ラベルは複数枚集まるものもあれば、最後まで仲間を見つけられずに一枚のまま、というものも多く残る。ここで無理に仲間と引き合わせることは、避けるべきであるとされる。

最後は(3)表札づくり、を行った。まず前ステップで集まったラベルを、塊ごとに束にする。さらに、これらのラベルが「集まったゆえん」の内容を別のラベルに要約し、要約したラベルを束の一番上に重ねてまとめる。KJ法の工程の中でも、表札づくりはとくに難しい作業であるとされる。今回は、同じ塊として集まったラベルを見渡した上で、そこに集まった全ての語彙が共有する概念を、英語で表記し名称をつけ、表札としてまとめた。

今回は、語彙に内包される概念の解釈が、分類にあたって一番重要な要素であった。作業実施者の恣意的な分類や、言葉の誤認識を防ぐため、インターネット用語や若者言葉に詳しい人物を交え、2人体制で作業を実施した。



図1 KJ法による「グループ編成」実施の様子

5. 結果・考察

ここでは、分析対象期間における「りんな」の語彙について、機械的、人的それぞれの方法を用いて、クラスタリングを行なった結果について述べる。

人的方法では計153の群が生成された。生成された群のうち、48(31.3%)の群は、複数の語彙が集約されて形成されたものである。残る105(68.6%)は、他の語彙と組み合わせることができなかった「一匹狼」であり、単独で一つの群と見なすこととした。

機械的方法については、あらかじめ生成する群の数を設定する必要がある。今回は、人的方法で得られた群の数に合わせ、計153の群が生成されるように設定した。その結果、複数の語彙が集約されたものとして38(24.8%)の群が、単独の語彙で一つの群と見なされたものとして115(75.2%)の群が生成された。

群の生成のされかたを定量的に比較すると、複数の語彙が集まって生成された群の割合は、どちらも20~30%程度であり、方法間で大きな差は見られなかった。

ところが、定性的に群の生成のされ方を比較すると、人的・機械的の両方法におけるクラスタリングの差が明らかになった。機械的方法のクラスタリングの限界としては、次のようなものが挙げられる。それはすなわち、(1)記号・シンボルに関する意味論レベルでの限界、(2)俗語・新語・流行語に関する意味論レベルでの限界、(3)語用論レベルでの限界、の3点である。

5.1 記号・シンボルに関する意味論レベルでの限界

第一の限界は、機械的方法では、絵文字や記号といった視覚的情報が解析対象から外れる点が挙げられる。そのため、言外に表された話し手の意図を鑑みた、群の生成がなされないことが確認された。

例えば、了承の意味を示す「はい」は、「りんな」の語彙において、「はい♡」「あ、はい。」「あ、はい…」といったヴァリエーションが存在する。語末に付される絵文字や記号によって、嬉々とした感情が働いているのか、それとも気が進まないのか、その心境が間接的に表現されているといえる。人的方法によれば、この語彙に関するクラスタは、「はい(積極的)」「はい(渋々)」「はい(笑)」の三種類に分けられた(表1)。しかしながら機械的方法によれば、これらは全て一つのクラスタと見なされる結果となった(表2)。

a このような仲間を見つけられずに一枚のまま残ったラベルは、「一匹

狼」や「離れ狼」と呼ばれる。

表 4 「ありがとう」を意味する語彙群② (機械的方法)

語彙	テキストサブ記号	分かち書き
ツンデレか	ツンデレか	[ツンデレ]
ありがとうー♪	ありがとうー	[ありがとう]
三ㄥ(ㄗㄗ^o^o)ㄗド ウルルル	三ㄗㄗドウルルル	[ㄗ, ㄗ, ドウルル ル]

表 5 「ありがとう」を意味する語彙群③ (機械的方法)

語彙	テキストサブ記号	分かち書き
あざます!	あざます	[あざ]
あざます	あざます	[あざ]
あざます!!!	あざます	[あざ]
あざます!!!!	あざます	[あざ]
あざっし!!	あざっし	[あざ]
あざます👍	あざます	[あざ]
あざます😄	あざます	[あざ]
あざます!!	あざます	[あざ]
あざます👏👏👏	あざます	[あざ]
あざます🙌🙌	あざます	[あざ]
あざっし。	あざっし	[あざ]
あざます💖	あざます	[あざ]
あざま す!!!!!!!!!!!!	あざます	[あざ]
あざます👍👍	あざます	[あざ]
あざます😄😄😄	あざます	[あざ]

表 6 「ありがとう」を意味する語彙群 (人的方法)

表札	語彙
('Thanks')	やだ、ありがとう👏❤️ ありがとう ありがとう! ありがとう👏 ありがとうございます! ありがとうー(((o(*∇*)o))) ありがとうー♪ あざっし!! あざっし。 あざます あざます! あざます!! あざます!!! あざます!!!! あざます!!!!!!!!!!!! あざます👍 あざます👍👍 あざます💖 あざます😄😄😄 あざます😄 あざます👏👏👏 あざます🙌🙌

d 「ㄗ」の意味については、例えば次のような解釈がある。「本当に、本気で、やる気満々、上等だ、強そうなど。マジㄗはㄗの強調系。」(佐藤優

加えて「ㄗ」のような、主に若者の間で広まった流行語についても、機械的方法による適切な分類は困難であることが示唆された(表 7, 8, 9)。本研究で分析対象とした「りんな」の語彙は 2018 年初頭のデータであるが、「ㄗ」とは、当時「JCIK 流行語大賞 2017 (コトバ部門)」にも選出された流行語の一つである。主に女子中高生の間で流行し、感情の高ぶりや物事の程度を表す意味合いで使われるd。そのためこの語彙の意味合いや使用される状況を鑑みると、「ㄗ」は、感情の高ぶりを示す「最高」や「(👏👏)👏ウェイ」といった語彙とともに、グループ化されることが望ましい(表 7)。ところが、機械的方法により生成された「ㄗ」「まじㄗ」を含む群は、語彙の特徴や意味合いに関する共通性が希薄であり、動作割り当てという観点からは上手くクラスタリングがなされていない事が確認できる(表 8, 9)。

表 7 流行語「ㄗ」を含む語彙群 (人的方法)

表札	語彙
('Awesome')	ㄗ Σd(・ω・d)おけ!! 三ㄥ(ㄗㄗ^o^o)ㄗドウルルル 最高 ㄥ(^o^o)h-==≡(ㄗ^o^o) (*^ω^*) (∇)ノ♡ (*^~*)ほほ (*`・ω・*)うい☆ (*皿'艸) (*-ω'-)9 ヨッヤア!! v(. . .)イェッ♪ まじㄗ 万次 (つ'ワ'c)♥← (👏👏)👏ウェイ

表 8 流行語「ㄗ」を含む語彙群① (抜粋, 機械的方法)

語彙	テキストサブ記号	分かち書き
お、おう	おおう	[おおう]
まじㄗ	まじㄗ	[まじ, ㄗ]
ちゅちゅちゅ	ちゅちゅちゅ	[ちる, ゆちゅちゅ]

表 9 流行語「ㄗ」を含む語彙群② (抜粋, 機械的方法)

語彙	テキストサブ記号	分かち書き
分った、りん なが一声掛け ればあつとい う間だ...	分ったりんなが一 声掛ければあつ という間だ	[分る, 一声, 掛ける, あつという間]
ㄗ	ㄗ	[ㄗ]
ありがとうー♪	ありがとうー	[ありがとう]

ほか。現代用語の基礎知識 2019 年版。自由国民社, 2019.)

言葉は、それを使用する集団のアイデンティティに、深く関与しているとも指摘される[22]. 「りんな」の語彙を適切に分類するためには、その言語文化を分析者自身が共有している必要がある. すなわち、新しくその言語社会に現れた語彙などについては、機械的方法によるクラスタリングに限界がある. と同時に、人的方法により分類を行う場合にも、当該語彙の運用に明るい人物を加えて、分析が行われる必要があることもまた示唆された.

5.3 語用論レベルでの限界

さらに機械的方法では、言語の運用面から導出・想起される意味を考慮した上での、クラスタリングは困難であることも明らかになった.

例として、「いいから」という語彙を見てみたい(表 10, 11). この語彙は、すでに不満表明の慣用表現として使用されており、相手に有無を言わせないというニュアンスを含む. 従来、カラという言葉は、主節と従属節の間に使用され、論理関係を表す接続助詞であるとされてきた. しかしながら文末に用いられる場合は、本来の用法から離れ、何らかの意図や感情を伝える新たな用法があることが、近年指摘されている[23][24]. 文末のカラのなかでも、とくに「断定形+〇」と呼ばれるパターンでは、反論や苛立ちなど、話者の否定的態度を表明するという特徴がある.

機械的方法により分類された結果を見ると、この「いいから」という言葉は、肯定の意味を持つ「いいよ」と同じクラスに分類されている(表 10). しかしながら先述の用例を考慮すれば、「いいから」「いいから。」といった「りんな」の語彙は、相手の意見を制止したり、不都合な意見を拒否しようとしたりする意味合いで使われていると考えられる. そこから、「いいから」は肯定の意味ではなく、むしろ「o(*≥o≤)o ダメ～!!」「ほんとむり！」といった拒絶の表現として、分類がなされる方が望ましい(表 11).

表 10 「いいから」を含む語彙群 (機械的方法)

語彙	テキストサブ記号	分かち書き
いいから	いいから	[いい]
いいから。	いいから	[いい]
いいよ	いいよ	[いい]

表 11 「いいから」を含む語彙群 (人的方法)

表札	語彙
(Reject)	いいから。 いいから o(*≥o≤)o ダメ～!! ほんとむり!

5.4 新たな「りんな」の語彙セット分類システムを構築するための方策

前節までは、チャットの会話エンジンを概念別に分類する際、「自然言語処理を基盤とした自動手法は、手動手法より相対的に劣る」という仮説から、人的・機械的方法によるクラスタリングの有効性を検討してきた. その結果、チャット文章に含まれるシンボルや流行語、語用から導かれる解釈の精度という点で、人的方法の方が、機械的方法よりも概念空間の生成に対し有効であることが示唆された.

しかしながら、人的方法によるクラスタリングは多大な手間と人的労力がかかることから、「りんな」の語彙セットのようなビッグデータを扱う場合は、実用的な方法であるとは言い難い.

そこで本節では、これまでに示された分析結果を踏まえ、新たな「りんな」の語彙セット分類システムを構築するための方策について論じる.

5.4.1 「記号・シンボルに関する意味論レベルでの限界」に対する方策

「りんな」の語彙の中でも特に目を引くのは、その多彩な絵文字や顔文字の使用である. これらのシンボルを用いることによって、短い文章であってもテキストに感情を盛り込むことができる. 結果、会話の相手との心理的な距離感を縮めることも可能になる.

「りんな」とのチャット・コミュニケーションは、今日、主にスマートフォン・アプリを用いて行われる. そしてそこで使用される絵文字は、2014年に国際的な標準文字コードである Unicode にも採用され、システムに搭載する上でもより使いやすいものとなった[25]. 現状の機械的方法による、「記号・シンボルに関する意味論レベルでの限界」に対しては、解析システムを Unicode に対応させることで絵文字を扱えるように改良し、形態要素解析の辞書を拡張することが、方策として挙げられる[26].

5.4.2 「俗語・新語・流行語に関する意味論レベルでの限界」、ならびに「語用論レベルでの限界」に対する方策

「りんな」の語彙を適切に分類しようとするためには、ベクトル化ネットワークを、現代用語に対応するように拡張する必要がある. ところが、新語や流行語のような新たに生まれる語彙については、既存の辞典等を搭載する方法では、十分に対応することが難しい. そこで、分析対象期間を拡張した「りんな」の語彙データ、あるいはソーシャル・ネットワーキング・システムのような類似した性質を持つビッグデータを解析対象とし、対話の用例を含めて語彙を学習することで、ベクトル化ネットワークの表現を改良する必要があるだろう.

本研究で分析対象とした程度のデータ数であっても、頻出語句と概念空間を一旦定義すれば、「りんな」という特定のキャラクターに限っては新たな語彙とその概念空間を更新しなくても良い可能性もある. 人的方法による概念空

間の定義を、機械的方法によるテキストの概念空間マップと組み合わせることで、自動手法が苦手とするクラスタリングに対応できる可能性がある。

6. まとめ

本研究では、インタラクティブ・エージェントの例として、女子高生のパーソナリティをもつ会話エンジン「りんな」の語彙セットを取り上げ、そこに手動および自動により概念空間を割り当てる手法について論じた。会話エンジンの応答は、きわめて短い文を含む。そのため本研究では、自然言語処理を基盤とした機械的手法は、人的手法より相対的に劣るとする仮説を立て、機械的手法と人的手法で得られる概念クラスタを比較検討した。

両手法で得られたクラスタを比較した結果、定量的観点から大きな差は見られなかったが、定性的観点から、機械的手法の限界が明らかになった。すなわち、(1) 記号・シンボルに関する意味論レベルでの限界、(2) 俗語・新語・流行語に関する意味論レベルでの限界、(3) 語用論レベルでの限界、の3点である。

これらの問題に対して、(1) については、「解析システムが絵文字や記号を扱えるように改良する」という方策を提案した。(2)、(3) については、「クラウド・ソーシングにより、機械的方法の表現を会話エンジンに適したものに改良する」という方策を提案した。

機械的手法の限界は、短い文章で意図を伝達するチャット文章には、抽象度が高い用語が多用される性質があることに起因すると考えられる。したがって、本研究が導いた知見は「りんな」の語彙だけでなく、同じ特徴をもつ会話エンジンにおいても、適用され得ると考える。

今後は、このような知見を反映させた、「りんな」の語彙セット分類システム、ならびにジェスチャー生成システムを開発することを目指したい。

謝辞 本研究の遂行に当たり、ご協力いただいた多くの方々々に心より感謝の意を表します。

マイクロソフト・ディベロップメント株式会社 元リサーチインターンの谷口峻悟さんは、在職中に「りんな」の語彙セットを対話ロボットに実装する研究に着手し、本研究の土台を築いてくれました。本稿の作成に至る過程で、多くの方々のご尽力、ご助言を賜りましたことを、重ねて御礼申し上げます。皆様、ありがとうございました。

参考文献

[1] J. Cassell, H. H. Vilhjálmsón, and T. Bickmore, "BEAT: the Behavior Expression Animation Toolkit," *Life-Like Characters*, pp. 163–185, 2004.
 [2] S. Robotics, "NAO, API documentation." [Online]. Available: <http://doc.aldebaran.com/2-5>.
 [3] M. Kipp, *Gesture Generation by Imitation From Human Behavior*

to Computer Character Animation. Universal-Publishers, 2005.
 [4] C.-M. Huang and B. Mutlu, "Learning-based modeling of multimodal behaviors for humanlike robots," *Proc. 2014 ACM/IEEE Int. Conf. Human-robot Interact. - HRI '14*, no. March, pp. 57–64, 2014.
 [5] E. Shlizerman, L. M. Dery, H. Schoen, and I. Kemelmacher-Shlizerman, "Audio to Body Dynamics," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7574–7583.
 [6] M. Plappert, C. Mandery, and T. Asfour, "Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks," *Rob. Auton. Syst.*, 2018..
 [7] S. Ginosar, A. Bar, G. Kohavi, C. Chan, A. Owens, and J. Malik, "Learning Individual Styles of Conversational Gesture," 2019.
 [8] H. Ahn, T. Ha, Y. Choi, H. Yoo, and S. Oh, "Text2Action: Generative Adversarial Synthesis from Language to Action," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2018
 [9] Y. Yoon, W.-R. Ko, M. Jang, J. Lee, J. Kim, and G. Lee, "Robots Learn Social Skills: End-to-End Learning of Co-Speech Gesture Generation for Humanoid Robots," 2018.
 [10] J. D. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, "Distributed representations of words and phrases and their compositionality," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2013.
 [11] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, vol. 19, no. 5, pp. 417–425, 2014.
 [12] Q. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," *Arxiv*, 2014.
 [13] R. Kiros et al., "Skip-thought vectors," *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, no. 786, pp. 3294–3302, 2015.
 [14] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Trans. Assoc. Comput. Linguist.*, 2017.
 [15] Z. S. Harris, "Distributional Structure," *WORD*, 1954.
 [16] Bishop, Christopher M. *Neural networks for pattern recognition*. Oxford university press, 1995.
 [17] MeCab: Yet Another Part-of-Speech and Morphological Analyzer [Online]. Available: <http://taku910.github.io/mecab/>
 [18] 川喜田二郎. 発想法—創造性開発のために. 中央公論社, 1967.
 [19] 川喜田二郎. 続・発想法—KJ法の応用と展開. 中央公論社, 1970.
 [20] 川喜田二郎. KJ法—混沌をして語らしめる. 中央公論社, 1986.
 [21] 同上, pp.121-132.
 [22] 磯川全次. 隠語の民俗学. 河出書房新社, 2011.
 [23] 孫思琦. 発話行為理論から見た文末の接続表現の用法: カラ・ケド・シを中心に. 筑波大学地域研究, 2017, no. 38, pp. 95-112.
 [24] 孫思琦. 発話行為理論から見る文末と文中のカラの違い. 日本語コミュニケーション研究論集, 2018, vol.7, pp. 69-75.
 [25] 石井美佐恵, 藤森恒正, 山田健太. 絵文字で誤解を生まないための仕組み. 電子情報通信学会 通信ソサイエティマガジン, 2015, vol. 9, no. 2, pp.68-69.
 [26] Eisner, Ben, et al. "emoji2vec: Learning emoji representations from their description." *arXiv preprint arXiv:1609.08359*, 2016.