

人間の連想を模擬するシステムの開発と分析

稲垣 健吾* 山本 英子** 神崎 享子* 井佐原 均*

* 豊橋技術科学大学

** 岐阜聖徳学園大学

1. はじめに

人間とコンピュータの間のコミュニケーション支援のために、人間が、提示される語からある語を連想する過程を、Web 文書集合を知識の元としてコンピュータにシミュレートさせるようなシステムの構築を試みた。ここでいう「連想」は、個人が持つ記憶を提示されるものに基づいて関連づけ、その結果、ある考えを思いつくことに相当すると考えられる。実際、被験者に語を提示し、その語から連想する語を記述させるという実験を行うと、直接的に考えつくと思われる語のほかに、提示した語からいくつかの記憶や情景を経てたどり着いたと思われる語が含まれる。そこで本研究では、ある情報（提示する語）に関してさまざまな考え（単語）を関連付けていき、ある時点でたどり着いた考え（単語）を発言することが人間の連想であると捉えることとした。また、この連想の過程を連想過程と呼ぶこととする。

本研究では、Word2vec を用いて人間の連想をもとにした連想経路を作成し、作成した経路が人間の連想を模擬できるか検証した。

2. 関連研究

これまでに連想過程のシミュレートの研究として、大塚[1]の研究がある。また Word2vec による人間の連想体系の研究を豊嶋ら[2]が行っている。大塚の研究では、大規模コーパス中の語の出現から語同士の関係を得るために、共起に基づく補完類似度を用いた。そしてその関係データを、分類的關係が中心でありつつ、関係間の距離が定義されていなかった日本語 Wordnet に組み込むことで、人間の感覚に近い知識体系の構築を試みたが、これ

だけでは人間の思考を完全には模擬するに至らなかった。しかし、大塚の作成した新たな知識体系のもと、「腕」を刺激語として与えた場合の人間が考えた連想語の関係（図1）を見ると、連想語から連想語へのつながりがあることが見てとれる。このことが、次節で示すような連鎖的な連想が行われているという考えに思い至る要因となった。

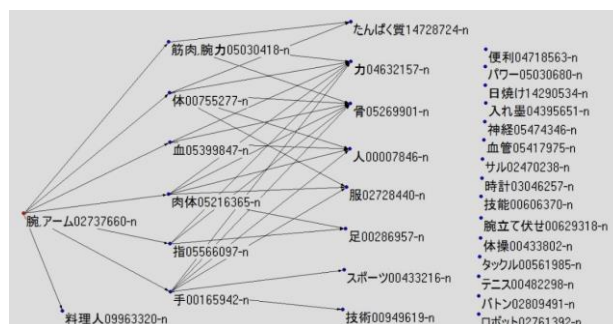


図1：大塚による連想シミュレーション

3. 人間の連想

人間の思考について支援するシステムの研究は多く存在する。その中で創造的な思考を生む方法がいくつか提唱されている ([3], [4], [5])。主な手法として、発散的思考法、収束的思考法、統合手法が存在する。発散的思考法はできるだけ幅広い視点からの情報を獲得することが重要となる方法である。平たく言うと、発散的思考法は今出ている考えをいろいろな方向にさらに広げる思考法であり、これには二つのパターンがあると考えた。一つは直接的な発散で、提示された語（刺激語）に直接何らかの関係を持つ単語を多く選び出す考え方である。もう一つは連鎖的な発散で、刺激語に何らかの関係があるいくつかの単語からさらに単語を選ぶということを続け

ていく考え方である。後者の発散方法の場合、関係が薄くならうとも単語が選り出されていく。本研究では、これらの発散的思考法が人間の脳内で行われていることが、連想過程に相当すると考える。

本研究では、人間の連想方法において、以下の三つの仮説を立てた。

仮説 1：刺激語から直接、連想を行う場合、人間は刺激語に強く関連している単語を多く連想する。

仮説 2：人間は連想した語から、さらに次の語を連想するといった連鎖的な連想を行う場合もある。この場合も人間は関連性の強い単語を連想しやすい。

仮説 3：連鎖的な連想を行う際に、関連性が比較的弱い単語の連鎖が行われる場合、人間はさまざまなジャンルの連想を発散的に行う。

図 2 はそれぞれの仮説について、刺激語を「ミカン」とした場合、どのような連想が行われるかを示した例である。

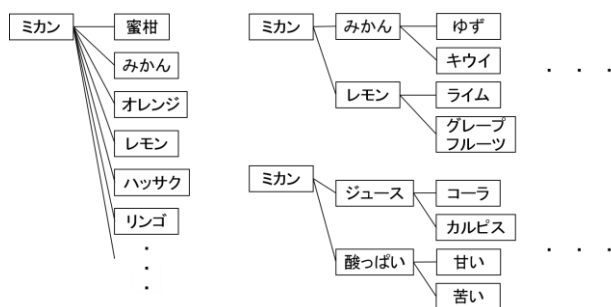


図 2 刺激語「ミカン」に対する連想

左：仮説 1 の例，右上：仮説 2 の例，右下：仮説 3 の例

以上の仮説に基づき、Word2vec で得られる共起関係を利用し、このような人間の思考状況を模倣するプログラムを作成した。

4. 連想シミュレーション

前節で立てた人間の連想方法における仮説をもとにしたプログラムを構築するために、仮説を実現するための以下のような手続きを考えた。

A. 抽出する連想語の選択

ある単語からいくつかの連想語を選択する場合、仮説では関係性のある単語としている。Word2Vec の作者ミコロフらは、意味的に関連の強い単語はベクトルが近くなると主張している[6]。ならば、Word2Vec によって構築されるベクトルデータの単語は分布が近くなるにつれて、関連性が大きくなるので、ある単語から関連性の高い単語を抽出する際は、コサイン類似度の大きい単語を連想語として順に選ばばよいと考えた。

B. 刺激語から直接連想される単語の選択

直接的な連想は記憶する必要がないため、刺激語に対するコサイン類似度が高い単語を連想語として抽出する。しかし、人間は関連性の非常に低い単語を直接連想することはできないと思われるため、どこかで抽出する単語数の上限を決めておく必要がある。

C. 連想の連鎖

刺激語に対するコサイン類似度の高い連想語を抽出し、それぞれを次の連想語を抽出するための刺激語とする。これを任意の回数を繰り返すことにより、連鎖的な連想を行うことになると考えた。

D. 関連性の比較的低い語の選択（発散的連想の実現）

ある単語から連想語を抽出する際、抽出される単語は A に示したように上位の語が順に選ばれる。そのなかでも、下位に位置する単語は上位の単語より関連性が低く、この単語から連鎖的な連想を行うと、その連想語は刺激語からの関連性がさらに低くなる。これを繰り返すことで広い連想による単語が得られると考えた。また、連想語を抽出する際、直前までに連想語として抽出されたことのある単語を除外することで、抽出する連想語の範囲をさらに広めることが可能になる(図 3)。

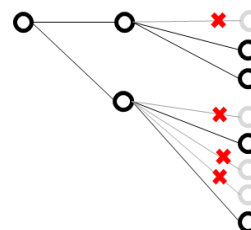


図 3 発散的連想の実現イメージ

Aはすべての仮説における単語の選び方について示し、Bは仮説1における、刺激語からの直接的発想を行う方法を示し、CとDは仮説2、3における連鎖的発想を行う方法を示している。今回のシステムでは入力時に連鎖を行う回数の上限を設定できるようにした。これを最大深度と呼ぶ。たとえば、この上限を1にすれば直接的な連想のみ行うことができ、1より高くすることで連鎖的な連想を行えるように構築した。また、1つの刺激語から抽出する連想語の数を設定できるようにし、これを連想経路の幅と呼ぶ。

構築する連想経路の幅、最大深度、元となる刺激語を入力とし、WebコーパスをWord2Vecで解析したベクトルデータを用いてシステムを実行することで、刺激語に対応する可能な連想経路を構築し、出力されるようにした。例えば図2や図3に示したような構造が出力される。また、元となる刺激語と対応する連想語を入力することで、構築した連想経路群における、刺激語から連想語までの経路を出力できるようにした。

構築する連想経路の幅は、連鎖的な連想を行う際には4、7、10とし、直接的な連想を行う際には抽出できる単語の最大数を500と設定した。最大深度はどれだけでも深く設定できるが、今回は深さ5までとした。

5. 被験者による連想実験

仮説およびシステムを検証するため、被験者による連想実験の結果と、システムの出力とを比較した。被験者は高校生20名と大学生31名に10個の刺激語(表1)を提示し、刺激語から連想できる語を可能な限り書き出してもらった。例を表2に示す。

被験者が連想した単語群から重複している単語を除外し、Webコーパスに含まれていない単語を削除した結果、刺激語に対応する連想語の数は、高校生の実験では529語、大学生の実験では587語となった。

このとき、記述された文字の状態そのものがその人物の経験、柔軟さが現れていると考えたため、ひらがな、カタカナ、漢字の違いは重複とは考えなかった。

表1 刺激語

カレー	サッカー	はさみ	パスタ
愛	学校	通信	音楽
林檎	腕		

表2 連想語の例

ID	年齢	性別	刺激語	連想語
1	17	男	音楽	歌手
1	17	男	カレー	からい
1	17	男	林檎	赤い
2	16	女	音楽	楽しい
2	16	女	音楽	歌
2	16	女	カレー	からい

6. 実験結果および考察

構築したシステムが人間の連想を模擬できているかどうかを確認するため、被験者に提示した刺激語から、被験者が実際に連想した連想語への経路が、システムの出力に含まれているかどうかを検証した。結果を数値化したものを表3に示す。表中の見出しは以下のものを示す。

W_D: それぞれ連想経路を構築する際に指定した、幅(Width)と最大深度(Depth)。

M: 被験者の考えた連想語と、構築した連想経路内の単語が一致した数(Match)。

M/N: 一致した数(Match)と、連想語データの全連想語の数(Number)の割合。

表3 一致結果

連想語データ A			連想語データ B		
W _D	M	M/N	W _D	M	M/N
4_5	133	25.14%	4_5	143	24.36%
7_5	305	57.65%	7_5	334	56.89%
10_5	423	79.96%	10_5	476	81.09%
500_1	148	27.97%	500_1	149	25.38%

W_Dが500_1のデータは直接的な連想を模擬した場合の結果を表している。この場合、M/Nが3割を超える

ことができなかった。これは、直接的な連想だけでは人間の連想を表すことができないことを暗示しているように思われる。連鎖的な連想を模擬した、W が 4, 7, 10 のデータは、それぞれ W が増えるたびに、連想語の一致数も増えている。10_5 の場合は 80%前後の一致数が得られ、この条件においてはある程度の人間の連想を模擬できうることを示唆している。しかし、刺激語「腕」に対して連想語「時計」、刺激語「林檎」に対して連想語「赤い」など、重複を除外する前は複数の被験者が連想した連想語が、連想経路の連想語と一致しなかった。これは、これらの関連性を Word2vec による連想体系のみではカバーしきれていないことが関係しているのではないかと推察する。

7. おわりに

本研究では、人間とコンピュータの間のコミュニケーション支援のために、人間が、提示される語からある語を連想する過程を、Web 文書集合を知識の元としてコンピュータにシミュレートさせるようなシステムの構築を試みた。そのために、人間の連想方法を考察し、その考察をもとに仮説を立て、その仮説を基づいた、人間の連想を模擬するシステムの構築を行った。結果として新しい連想体系の構築を行うことができたが、人間の行う連想を模擬しきることはできなかった。

ただし、Word2vec で作成するベクトルデータによって連想経路が大きく変わる可能性があるため、様々な条件で連想経路の構築を試みるのが今後の課題である。

謝辞

本研究は JSPS 科研費 23700251 の助成を受けたものである。

参考文献

- [1] 大塚道長：人間の発想を模倣するシステムの開発に向けた複数の概念体系の融合に関する研究 豊橋技術科学大学修士論文 (2012)
- [2] 豊嶋章宏, 奥村紀之：語の連想体系としての概念ベースの評価 信学技報 NLC2015-1 1-5

- [3] 伊藤淳子, 東孝行, 宗森 純：単語共起度の低い単語を提示する発想支援システムの提案と適用, 情報処理学会論文誌, vol.56, no.6, pp.1528-1540, 2015-06-15
- [4] 平石徳己：収束的思考と態度特製の関連性, 久留米信愛女学院短期大学研究紀要, 第 30 号, 7 ~12, (2007)
- [5] Guilford J.P. : Three faces of intellect, American Psychologist Vol.14 No.8 pp.469-479(1959)
- [6] 西尾泰和：Word2vec による自然言語処理, O'REILLY (2014)