

グラフデータベースによる知識表現の試み -連想するロボットをめざして-

高橋悠希^{†1} 内海哲史^{†1}

本稿では、人間の思考が連想に大きく依存しているという仮定のもと、より人間に近い情報処理を実現するための技術として、グラフデータベース上に構成した意味ネットワークによる連想処理の基礎を考える。簡単な英文から生成した構文木による意味ネットワーク構築と、文章の順序に基づいたタグ付けによる各概念の関連性の算出、また連想処理の評価実験として人間の連想結果との比較を提案する。

A Challenge for Associative Processing Using Graph Database

YUKI TAKAHASHI^{†1} and SATOSHI UTSUMI^{†1}

In this paper, we consider a basis of associative processing using semantic network to implement more humanlike information handling. Semantic network has been constructed on graph database from parse trees and associative processing is worked out on calculation of relevance rate. In addition, we propose an experiment for evaluating of the processing.

1. はじめに

「世界は高度情報化社会に移った」という声が聞かれるようになって久しい。現代において情報技術は人々の生活に完全に溶けこみ、高性能の CPU を搭載した各種情報機器は、一人の人間が複数台を所有することも決して珍しいことではなくなった。

これに伴って情報技術に触れる人間そのものが増加し、それらの内包する弱みが、機械翻訳や検索の不自然さといった形で、多くの人の目に留まるようになった。そういった箇所を改めるべく、IT 企業各社は自社の技術に改良を重ねているが、未だそれらの不自然さは払拭されていない。

本稿では、それらの問題を引き起こしている原因はロボット（ここでは情報システム全般を指す、以下同様）と人間の情報探索の差異にあると仮定し、ロボットの処理を人間により近づけるための技術について考える。

人は問題解決に際して情報探索を行うとき、問題の中で示された概念についてなんらかのつながりをもつものを無意識のうちに記憶から呼び起こし、さらに呼び出されたものに対しても同様の操作を行うことで、解決すべき問題とは直接関係の無い情報を評価対象に入れてから論理的な作業を行う。

しかし、ロボットが行う探索は、まず検索キーを設定し、データベースの中からそれに適合するデータを探し出すというものだ。

つまり、人の情報処理過程には「必要なデータ」「関係はあるが不要でないデータ」「必要でないデータ」の三種類が介在するのに対し、ロボットの情報処理には「必要なデータ」と「不要でないデータ」の二者しか関わらないのである。ここから、「関係あるが不要でないデータ」を処理の対象とすることこそが、人間の人間らしい思考を実現しているのだという仮定を導き出すことができる。

本研究ではこの「関係はあるが不要でないデータ」を呼び出すことを人の連想機能として捉え、それを模倣するための基礎的技術として、構文木からの意味ネットワークの構築と、文章の順序に基づいたタグ付けによる各概念の関連性の算出、また連想処理の評価実験として人間の連想結果との比較を提案する。

2. 技術と学術的背景

本章では、連想処理システムを考えるにあたって下敷きにした学問・技術について簡単に述べる。

^{†1} 鶴岡工業高等専門学校
Tsuruoka National College of Technology

2.1 心理学における連想と思考の関係

本研究は、「連想が人の思考の中で基本的且つ重要な役割を担っている」という仮定の上に成り立っている。このような考えを確立したのは哲学からスピントアウトしたばかりの心理学者たちであり、それは19世紀当時盛んに研究が行われていた連合心理学（連想心理学とも）の業績である。連合心理学とは、人のもつ概念が、より単純な複数の概念の連合として成立し、それらの概念の連想のみによって人の思考が説明できるというものだった。思考は連想のみで成り立つというこの考え方はその後の研究によって否定され、人間の思考は連想のみによらず、ほかの様々な精神のはたらきとの連携であるとされた¹⁾。

しかしながら、依然として連想という心的機能が人間の思考の中で主要な位置を占めていることは、現在も心理学や情報工学の多くの研究者によって連想が研究されていることや、フロイトやユングの言語連想法が今なおカウンセリングに用いられていることから推察することができる²⁾。

2.2 連想実験

連想を用いた実験では、前段で挙げたフロイトとユングの方法が有名である。今回はこの二つのうち、連想の範囲がある程度限定されることから、ユングの実験を参考した。以下にユングの実験の概要を述べる。

心理学者のカール・グスタフ・ユングは、人の潜在的なコンプレックスを描き出すことを目的として連想実験を考案した。これは被験者に対して単語（刺激語）を次々と与え、それぞれに対して最初に思いついた単語（反応語）を読み上げさせるといったものである。また、一度全単語を与えた後、一度目と同じ語を返すように指示し、はじめから同じ順序で再び単語を与える。このとき、反応を返すまでに要した時間を計測し、以下のような反応があった場合を特異なケースとして分析の対象とする²⁾。

- | | |
|-------------------|------------------------|
| (1) 反応時間の遅れ | (2) 反応語を思いつけない |
| (3) 刺激語をそのまま答える | (4) 明らかな刺激語の誤解 |
| (5) 同じ反応語がくりかえされる | (6) 明らかに奇妙な反応 |
| (7) 再検査の際の忘れ | (8) 観念の固執（前の刺激後の影響を残す） |

2.3 意味ネット

コンピュータによる意味表現の手法として、古くから意味ネットがある。意味ネットは

各概念のノードおよびISA関係やHASA関係、CAN関係といったリレーションを持ち、個々の知識の関係を記述することで知識体系を表現するものである³⁾。

本研究ではデータの格納にグラフデータベースを使用する。グラフデータベースは、ノードとリレーション及びそれぞれが持つプロパティによってデータを格納する、比較的新しいNoSQLデータベースである。以下にそのモデルを示す⁵⁾。

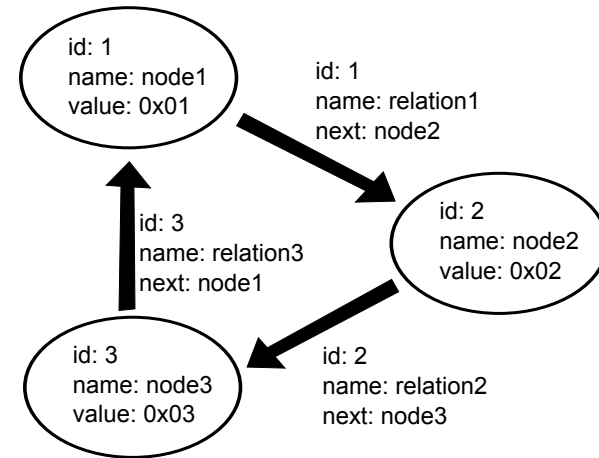


図1 グラフデータベースのデータ構造

グラフデータベースはグラフ構造をそのまま格納することのできるデータベースであり、グラフの持つ特性を損なうことなく運用できる。グラフが持つデータ構造としての利点は、複雑なネットワークを構成しやすいことや経路問題を解くのが容易な点などである⁶⁾。

2.4 構文解析

構文解析器は、文章から品詞情報や係り受け関係などの構成情報を取り出すものである⁷⁾。ここでは基本8品詞（名詞、動詞、形容詞、副詞、前置詞、接続詞、感嘆詞）のうち冠詞を除いて限定詞を加えた8品詞、及び品詞句（名詞句、動詞句、前置詞）が構文木の形で出力されるものとする^{8),9)}。

3. 意味ネットの構築

前述のように、本来の意味ネットは限定されたリレーションを用いて知識体系を表すものであるが、今回対象としているのは事実の表現であるため、リレーションの役割を変更して前置詞の表現に限定した。また、IS-A や HAS-A といった関係に加え、通常も動詞もノードを用いて表すことで、動詞に関する修飾を可能とした。

意味ネットの構築は、各品詞節ごとにグラフ構造への変換法を定め、入力文を構文解析器にかけて得られた構文木に対して根から順に適用してゆくことで行う。表 1 に各品詞節ごとのグラフ表現を挙げ、図 2 に例文 1 から得られる構文木を、図 3 と図 4 に処理過程と結果として出力される意味ネットを示す。なお、構文木左の数字は階層番号であり、処理過程の Step 数と対応している。図上では名詞ノードを丸、動詞、形容詞、副詞ノードを四角で表す。

表 1 文節・品詞ごとの表現・展開法

文節	表現もしくは展開法
S (文)	NP と VP に展開し、NP から VP へリレーションを作成する。
NP (名詞句)	NP と PP, Adv に展開し、PP, Adv それぞれから NP にリレーションを作成、既存のリレーションを NP に連結。 それらを含まない場合は Det と Adj, N に展開し、Adj から N にリレーションを作成、既存のリレーションを N に連結。
VP (動詞句)	V と NP, PP に展開。PP から V, V から NP にリレーションを作成し、既存のリレーションを V に連結。
PP (前置詞句)	Prep と NP に展開。既存のリレーションに Prep をプロパティとして設定し、NP に連結。
N (名詞)	ノードとして表現。
V (動詞)	ノードとして表現。
Prep (前置詞)	リレーションとして表現。
Adj (形容詞)	ノードとして表現。
Adv (副詞)	ノードとして表現。
Det(限定詞)	今回は評価しない。

例文 1 Two men look out through the same bars.

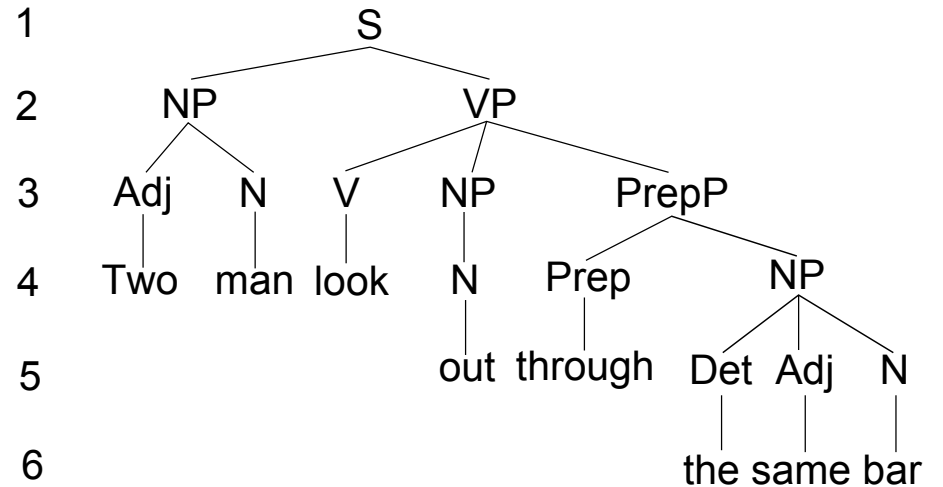


図 2 例文 1 から得られる構文木

4. 関連度の計算

ユングの実験が成立していることから、人の連想には順序があり、ある概念異なる概念の関係には重みづけがされていることがわかる。そのための指標として関連度という値を導入するとともに、複数の文章から作成した意味ネットをまとめて扱うことを目的として、各文章から構築した意味ネットを文章番号によって三次元的にマッピングすることを提案する。以下の例文 2,3,4 から生成した意味ネットの三次元マップを図 5 に示す。

例文 2 Humpty sat on a wall.

例文 3 Humpty had a great fall.

関係のある事柄が続けて述べられると仮定すると、二つの概念の関係度はそれらの文章中の距離に近いほど高くなる。文章中の距離はグラフ中の距離に受け継がれるため、その性質は意味ネット上でも同様である。ここで「距離」とはグラフ上で一方から他方に移動する際のリレーションの経由数、またその間に文章を行き来した回数の合計とする。関連度の上限は 1 とし、距離 0 のノード (同じノード) がこれにあたるとすれば、経由したリレーション

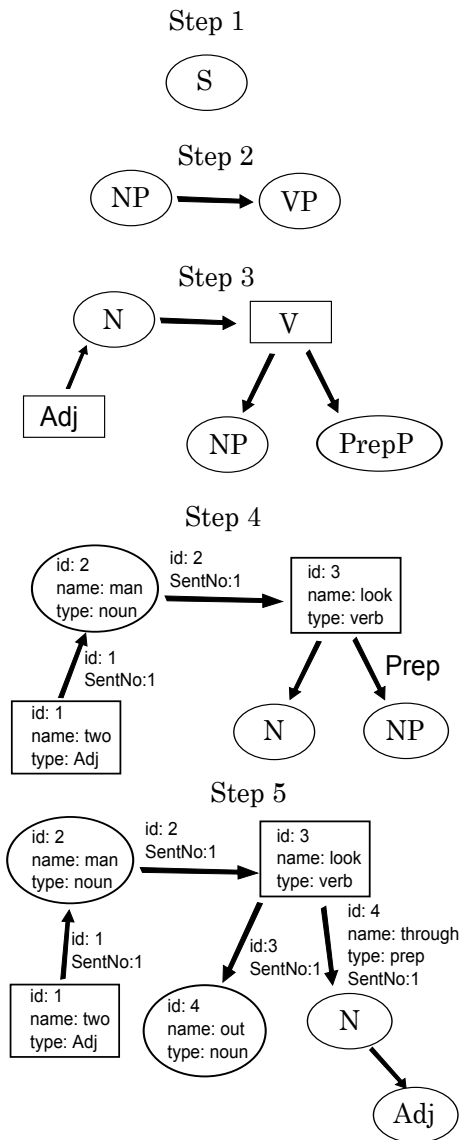


図 3 グラフ生成手順 階層 1-5

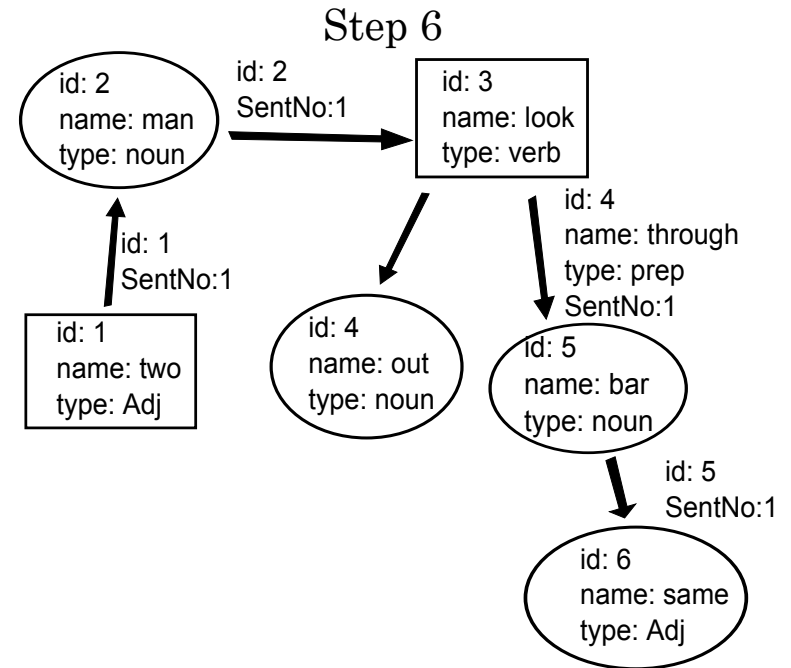


図 4 グラフ生成手順 階層 6 生成された構文木

数を S 、文章間の移動回数を J とすると、関連度 R_r を以下の式 (1) で表すことができる。

$$R_r = \frac{1}{1 + S + J} \quad (1)$$

式 1 に基づく、名詞 "wall" と他の名詞の関連度を表 2 に示す。表中では経由するノードの id で最短ルートを表す。

表 2 名詞 "fall" と他の名詞の関連度

対象語句	最短ルート	リレーション数 (S)	文章間の移動回数 (J)	関連度 (R)
humpty	12,11,10	2	0	0.33
fall	12,11,10,13,14	4	1	0.17

5. 実験方法

連想処理が人間と同様に行われているか否かの検証実験として、第二章で挙げたユングの連想試験をもとに以下のような方法を提案する。

まず、人間とロボット（ここでは連想システムを指す）の双方にある程度長さのある同じ文章を読ませる。その後、文中から取り出したいいくつかの単語を刺激後としてそれぞれに与え、最初に連想したものを返させる。なお、反応時間の計測は行わない。

その後、人間とロボットの反応語を比較し、その類似の度合いによってロボットの連想処理に対する評価を行う。

ここで連想処理とは「与えられた単語の周辺概念を想起する」ことであるとし、ロボットは与えられた単語の周辺概念のうち、最も関連度の高いものを返す。関連度が同じであった場合は、記憶の時間劣化の観点から、より遅く述べられたものを返す。

6. 課題

現在判明している本提案の問題及び課題として、以下が挙げられる。

まず、連想処理のアルゴリズムが不完全であることが一点。今回の方法では、与えられた単語に対し、常に隣のノードが連想されてしまうが、人間の連想機能ももっと複雑なもののように思われる。

さらに、人間の連想機能についても検証が必要である。本提案は「周辺概念を想起するこ

と」が人間の連想機能であるという仮定のもと為されているため、人間の連想機能に対し、その基本的性質を明らかにするような試験、すなわち完全にロボットと同じ位置に立った状態での連想実験を行わなくてはならない。

加えて、連想試験の方式も万全とは言い難い。実験の目的は、背景知識の影響を排して人間とロボットの連想を比較することであるが、そのためには、取り上げる英長文は被験者が未読であるほか、被験者にとって全単語が完全に未知のものであることが理想である。しかし、実際にそのような文章を作成するのは困難である。

7. まとめ

本稿では、連想処理を実現するにあたって必要となる、意味ネットの構築法とそれを用いた概念間の関連度の計算法、連想処理の評価実験について提案した。今後は上に挙げた問題を解決に向かわせるとともに、これら技術の実装・評価を行い、より人間に近い連想処理の実現に向けて取り組む。

参考文献

- 1) 波多野完治, 藤永保 ほか: 思考心理学, 大日本図書 (1976)
- 2) 河合隼雄: ユング心理学入門, 培風館 (1967)
- 3) Allan M. Collins, M. Ross Quillian: Retrieval time for Semantic Memory, Journal of Verbal Learning and Verbal Behavior 8 (1969)
- 4) 服部元信, 荻原将文: Multidirectional Associative Memory(MAM) による知識処理, 信学技報 NC94-57, 電子情報通信学会 (1994)
- 5) Neo Technology: neo4j open source nosql database, <http://neo4j.org>, Neo Technology (2007 - 2011)
- 6) Oystein Ore, Robin J. Wilson: Graphs and Their, The Mathematical Association of America (1990). (大石康彦訳: やさしくくわしいグラフ理論入門, 郵政研究所研究叢書 (1990))
- 7) Eric Brill: A Simple Rule-Based Part of Speech Tagger, Department of Computer Science University of Pennsylvania (1992)
- 8) 藤原博: 英語の構造, 大修館書店 (1984)
- 9) Robert Lowth: A Short Introduction to English Grammar, R. Aitken (1799)

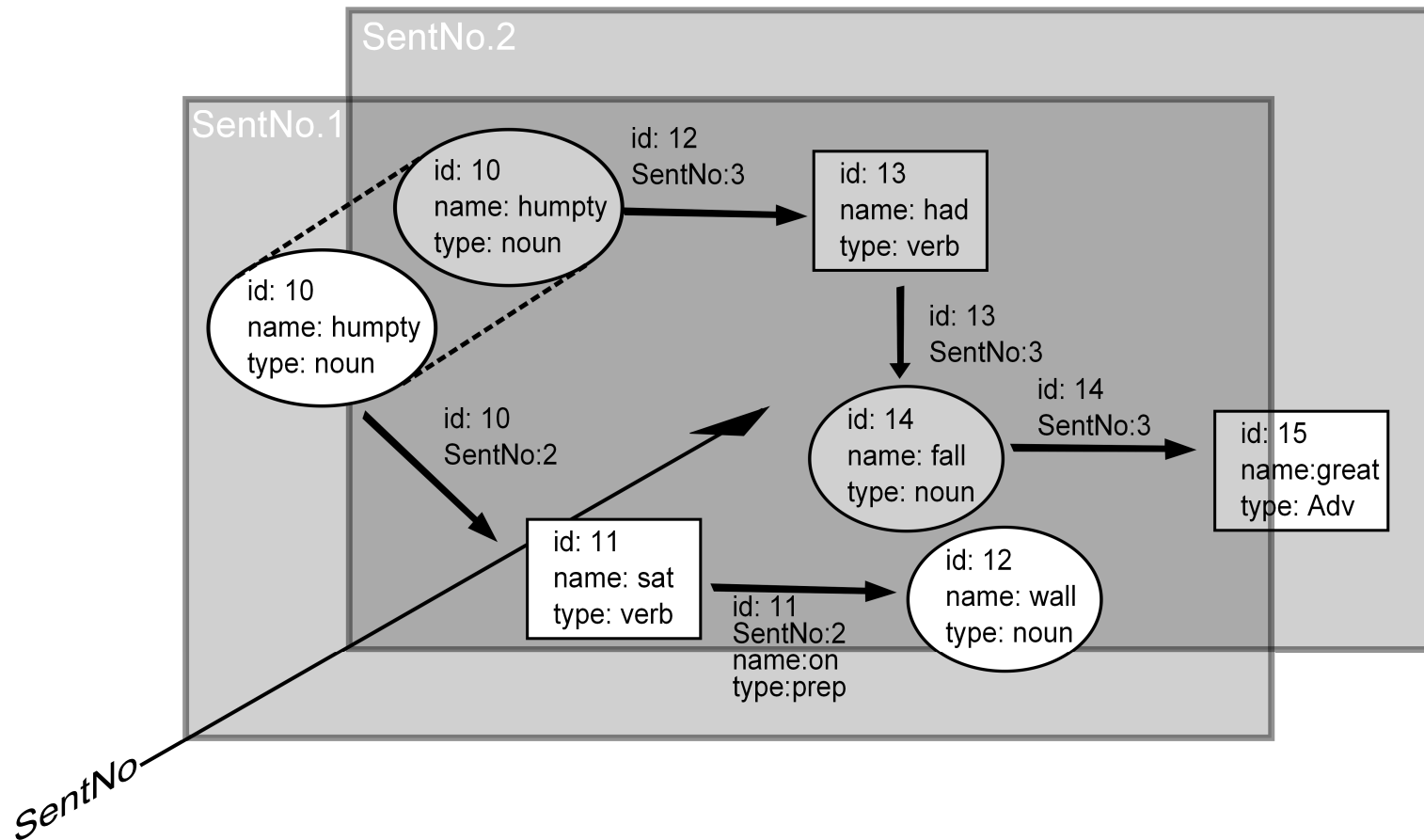


図 5 文章番号による意味ネットの三次元マップ