

# 概念辞書と複雑ネットワークを用いた連想シミュレーション

Associational Simulation using Concept dictionary and Complex network

伍賀正典\*1      相澤洋太\*2

Masanori GOKA      Yota AIZAWA

\*1 福山大学

Fukuyama University

\*2 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

In general, human thinking process is very difficult to observe, and it is difficult to evaluate it. Therefore, we focused on constructive approach as a method of studying such unobservable phenomenon. In previous work, we clarified relation between the network structure of a concept generation process and the evaluation value of a design based on this approach. If the concept generation process acquired by the computer algorithm is coincident with it obtained in previous work, dynamic analysis of human thinking process can be conducted by the computer simulation based on constructive approach. First, in order to obtain information required for the algorithm of association, we analyze the network structure of WordNet which is used for a virtual concept generation process in previous work. Second, the simulation of association is performed using the program based on these analysis results.

## 1. はじめに

人間の思考がどのようなプロセスで実行されるかについては、いまだ明確に解明されていない。工学分野において、優れた製品が作られたとき、その製品を作成した人物がどのようにしてそれを導き出したのかという問題は、いままでほとんど議論されてこなかった。優れた製品を生み出す人間の思考はどのようなものか、創造性の高いアイデアはどのように生まれるのかという疑問を解決することは、新たな商品開発の支援を可能にするためにも有用である。しかしながら、創造性の高い業績は、説明不可能な天賦の才によってもたらされるものだという考えが根強い [Minsky 06]。特別な才能や素質、特質や天賦の才は情報量の少ない概念であり、工学的に必要なものは多くの現象を精査しパターンを発見し計算機で実行可能なように定式化することである。このような、計算機を用いて人間の心を解明しようとする、心の計算アプローチに対して、記号化に重点が置かれており、古典的な記号的計算主義に捕らわれているという批判がある [Minsky 86]。人間の思考を、計算機の処理に適する記号によって説明することは不十分であるということは、古き良き人工知能に対する批判において数多く指摘されているところである。

本研究では、人間の思考プロセスのような観測の困難な現象に対するアプローチとして、構成論的アプローチに注目している。この構成論的アプローチに基づく構成論的シミュレーションは、中島により提唱されており、対象とする現象に類似ないし相似している現象が、対象とする現象とは異なる機構（ここでは現象を生成する原理を機構と呼ぶ）から創発される場合に、創発された現象の内容から対象とする現象の機構を推測するために用いる [中島 08]。機構の観測が困難な場合や、分析の困難な現象を理解するために有効とされる。複雑系がその典型例であるが、本研究で扱う人間の思考プロセスもこれに該当する。現象の機構を直接模倣するのではなく、現象を実現する機構の共通性を探ることが構成論的シミュレーションの目的である。

連絡先: 伍賀正典, 福山大学工学部電子・ロボット工  
学科, 広島県福山市学園町一丁目, 084-936-2111,  
goka@fuee.fukuyama-u.ac.jp

先行研究では、実際に人間によって行われたデザイン生成時の思考過程を概念から概念への連想過程とみなし、仮想的にモデリングし、その概念生成プロセスの特徴をとらえることを目的とした [田浦 11]。そこでは、ある概念からデザインを生成する経路を意味ネットワーク上でモデリングしている。意味ネットワークとは、上位語と下位語の関係や連想関係など、単語間の意味の関係からなる構造のことである。先の実験には、英語で書かれた大規模な概念辞書である WordNet 3.1 を用いた。このように構築された仮想的な概念生成プロセスを複雑ネットワークの指標を用いて分析し、仮想的な概念生成プロセスの構造とデザイン成果物の創造性に関する評価値との関係を求めた。成果として、仮想的にモデリングしたネットワークの広がりや複雑さを示す指標と、デザイン成果物の独創性に関する評価値との間に有意な関係があることを明らかにした。

## 2. 目的

以上で述べた研究背景を踏まえ、本研究では、思考の時間的な方向性も考慮し、始点の概念から終点へと自動的に連想していくアルゴリズムを開発することを主な目的とする。自動的に連想していくアルゴリズムの開発には思考のルールが必要であるが、これは直接観察することが困難であるために明確に解明されてはいない。しかし、本研究で使用する意味ネットワークである WordNet は、人間の言語処理の方法に関する知見を踏まえて作成された概念辞書であるため、WordNet のネットワークとしての構造を理解することは、思考のルールの推測を補助することになると考えられる。故に、本研究ではまず、WordNet の持つネットワーク構造を抽出し、複雑ネットワークの指標を用いて構造の特徴の解析を試みた。

### 2.1 WordNet について

WordNet にはトータルで約 15 万語もの語が収録されているが、それらは、名詞、動詞、形容詞、副詞の文法上の扱いが異なることから、品詞ごとに区別されて収録されている。また、収録される語が synset と呼ばれる同義語のグループに分類され、簡単な定義や他の語との関係が階層的に記述されている。単語間の関係としては以下に示すものがある。

- 上位語 (hypernym)  
すべての X が Y の種類の一であるなら Y は X の上位語である。
- 下位語 (hyponym)  
すべての Y が X の種類の一であるなら Y は X の下位語である。
- 全体を表す語とその部分を表す語 (meronym)  
Y が X の一部であるなら Y は X の meronym である。

WordNet の構造は, synset のグループ単位で意味関係の接続が記述されている。また, WordNet では, 名詞, 動詞, 形容詞のそれぞれについて階層構造があり, 異なる品詞の語とのリンクは存在しない。したがって今回は, 119,304 語の名詞のネットワークに対してのみ解析を行う。

## 2.2 WordNet からのネットワーク構築方法

本研究では, C 言語と Perl を使用して, ネットワーク構築プログラムを作成した。具体的には, 選択した単語と, その上位と下位にある単語との接続関係をネットワーク可視化ツールである Pajek で使用できるデータ形式に生成する。

WordNet は基本的に synset と呼ばれる同義語のグループで単語が収録されており, 上位語, 下位語の接続はこのシンセット単位で記述される。しかし, 思考の連想を取り扱うため, 同義語間の関係も連想の範囲に十分含まれると考えられるため, 本研究では上位下位関係に加えて, 同義語間関係も接続として含んだネットワークの抽出を行った。このイメージを図 1 に示す。

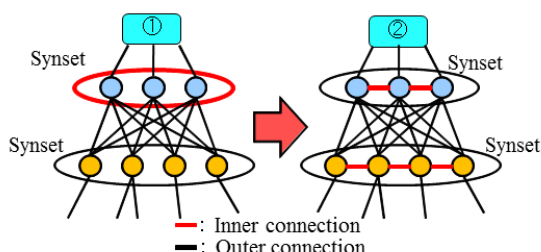


図 1: ネットワーク構築方法

## 2.3 ネットワーク評価方法

ネットワークという言葉は現在, 日常的に使われている。インターネットなどのコンピュータネットワークや, 人間関係の繋がりも一つのネットワークである。このような現実中存在するネットワークにはいくつかの共通した性質が存在する。実際に細胞の中の分子間相互作用のネットワーク構造の特徴も, インターネットや現実社会の人間関係と同様の性質を持つことが報告されている [Barabasi 04]。本研究で扱う WordNet も, 心理学的な正当性をもって製作されたデータベースであるために, これらと共通の性質が確認できるのではないかと推測した。今回はその中でも, スケールフリー性, スモールワールド性, クラスタ性, の三つの性質の観点から評価を行う。

## 3. WordNet の構造評価

図 2 に WordNet のネットワーク構造を Pajek を用いて可視化したものを示す。同様に Pajek を用いてネットワークの解析を行い, ネットワークの度数分布, ネットワーク全体の平均距離, クラスタ係数を得た。

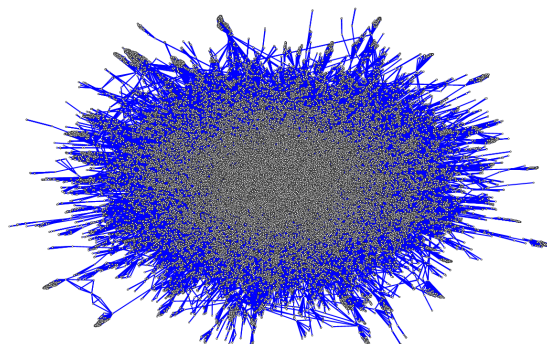


図 2: WordNet ネットワーク構造の概観

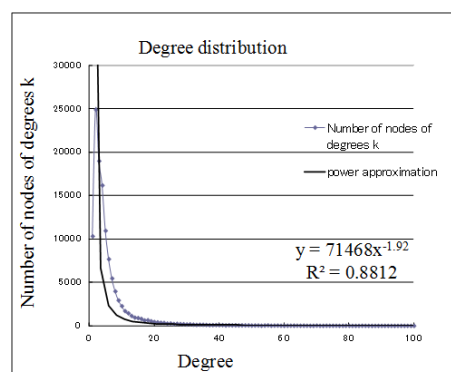


図 3: WordNet ネットワーク構造の度数分布

図 3 に度数の分布とその近似曲線を示す。この近似曲線は,  $y = 49672 \times x^{-1.863}$  であり, 近似曲線の予想値が実際のデータにどの程度近いかを示す値は  $R^2 = 0.8812$  であり, このデータの近似曲線として適していることが分かる。図 2 の外観からはやや分かりにくいだが, 度数分布がべき則に従っていると言えることから, ネットワークはスケールフリー性があり, ハブが存在する。今回のネットワーク内の度数の大きいノード (単語) の上位には, writer (作家, 物書き) や herb (ハーブ) のように下位語に固有名詞 (作家名, 植物の名前) を多く抱えるノードが多数存在した。これらの固有名詞は, そこで行き止まりである場合が多く, 連想先に選んでしまった場合構築される連想のネットワークがあまり広がらないという問題が発生することが考えられる。この問題を回避するために, ネットワークが広がりを持つ可能性の高い単語が選択されるように, 新たな指標として二次の和を提案する。これは, あるノードのリンクの先に存在するノードが持っている度数の総和である。今回, 度数の高いノード上位 50 個の二次の和を計算した。この結果から, 単純に度数の高いノードのほうが必ずしも大きな広がりを持っているわけではないことが分かった。しかし, WordNet 全体で見れば, やはり度数が高くなるにつれて, 二次の和も大きくなっていくので, 下位語に固有名詞を多く抱えるノードも, 行き止まりの確率が高いとはいえ, 全体的に見れば十分なネットワークの広がりを持っている。二次の和が大きいノードはリンク先の行き止まりの確率が低いと考えられるので, 連想アルゴリズムの行先選びの判断材料として, この指標は有効である。また, クラスタ係数は  $C = 0.718$ , ネットワークの平均距離は  $L = 6.048$  であった。

この解析によって、WordNet のネットワークとしての構造についての知見を得た。WordNet の基本的構造はツリー構造であると考えられ、これは WordNet の単語収録が階層構造を持っていることが原因であると考えられる。また、基本はツリー構造でありながら、多くの意味を持つ単語が多数の synset に含まれることで、クラスターが発生していると考えられる。これらのデータを参考にして、連想シミュレーションを行うためのアルゴリズムを展開する。

#### 4. 連想シミュレーション

本研究では、思考の時間的な方向性を考慮し、始点の概念から終点へ自動的に連想していくアルゴリズムを開発し、思考過程の「動的」な構造解析を行うことを目的としている。これまでも、知的活動のシミュレーションモデルを構築することで人間の思考についての知見を得る研究は行われてきた。思考の中でも特に問題解決に絞ってみても、問題解決の過程を状態空間分析という考え方によって説明したニューウェルとサイモンの一般問題解決機 (General Problem Solver: GPS) がある。GPS は、目標状態と初期状態の差を考えて、その差を減らし問題解決に近づくということが基本的なアイデアである。また、事例に基づく推論 (Case-Based reasoning: CBR) と呼ばれるモデルも存在する。

本研究では、WordNet のネットワーク構造から得られたデータを基に自動連想プログラムを作成した。このプログラムを用いたシミュレーション実験で、先行研究で用いた実際の人間の連想から得られたネットワークと同様のものが得られれば、人間の思考過程の動的な構造解析を行うことの可能性を示すことができる。本研究の実験では、先行研究で得られたネットワークの再現を目的としている。

##### 4.1 アルゴリズムの仕様

先行研究では、平均次数と創造性 (Originality) との間、多義性と創造性との間に相関があることが示された。WordNet において、多義性が大きい単語ほど多くの synset に収録されており、結果的に多義性の大きい単語は次数が大きくなる。創造的な思考を行う場合、WordNet における次数の大きさは、次の連想先を選択するための指標として有力な候補と成り得る可能性が高い。今回のプログラムでは、「人間は連想をするときに、連想の行き先として他よりも連想しやすい単語 (次数の大きい単語) を選びやすい」と仮定してアルゴリズムを構築する。人間の思考には確率的な揺らぎが存在すると考えられるため、行き先候補の持つ次数に一樣乱数を掛けた値を比較して、行き先を決定する。この行き先の選択方法のイメージを図 4 に示す。

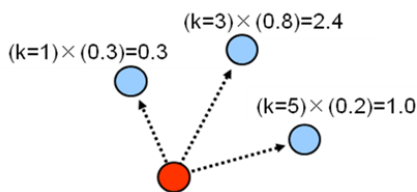
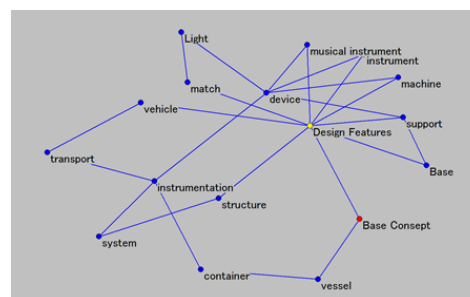
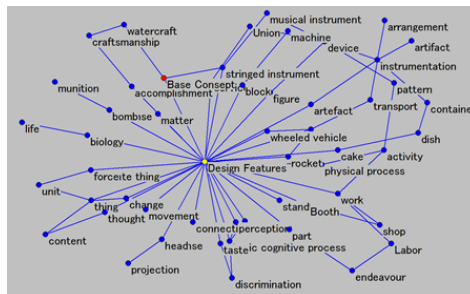


図 4: 連想決定の方法

また、先行研究で得られたネットワークで特に出現数の多い単語には重みを付けた。先行研究のネットワーク内で次数が 3 以上の単語は頻出することが見て取れるので、本実験ではこの頻出単語に連想が到達した場合、思考が分岐しネットワーク



(a) 連想回数 6 回



(b) 連想回数 10 回

図 5: 連想によるネットワーク構造

が複雑さを得ていると推測した。これを踏まえ本実験でも、次数が 80 を超えている単語に連想が到達した場合、連想の枝を 3 つとする。また、WordNet には上位語下位語の概念が存在するが、このアルゴリズムでも連想に上下位語の方向性を持たせ、具体的には、総連想数の半数は上位語、残りは下位語方向に連想する。これは、例えば、度数に依存した連想方向は上位語方向に遷移する傾向があり、抽象度が高い単語になるからである。今回のシミュレーションでは、基底概念を開始点とし、デザイン特徴を目標点としている。先行研究の基底概念である「船・ギター」の 2 単語を開始点とし連想を開始し、既定の連想数を終えた時点で到達している単語の集団をデザイン特徴としている。シミュレーションは、総連想数 6, 8, 10, 12, 14, 16 回の計 7 パターンで、各々の設定で 100 回行った。得られたネットワークデータは Pajek を使用して解析した。

##### 4.2 シミュレーション結果

連想によって生成されたネットワークデータは先行研究と同様、基底概念とデザイン特徴をそれぞれ一つのノードとして出力した。図 5 に連想数 6, 10 回の結果を示す。図中の Base Concept が基底概念、Design Features がデザイン特徴を表している。先行研究が提示したネットワークと、今回のシミュレーションの結果得られたネットワークがどれだけ一致するかの判断基準として、ネットワーク指標値を利用することが考えられる。この値が近ければ、ネットワークの構造が似ているということができると推測される。先行研究では、デザイン実験の結果得られた 20 人分のデザイン成果物から構築された 20 個のネットワークが存在する。これらのネットワーク指標の最大値と最小値の間に、今回のシミュレーション結果のネットワークの指標が収まっていれば、先行研究で示されたネットワークと近いと判断する。この最大最小値の範囲から、連想シミュレーションの結果の値がどれほど外れているか (外れ率と呼ぶ) を計算し、その値を比較することで先行研究で得られたネットワークに近いものを探す。先行研究で得られたネットワーク指標の最大最小値を表 1 に示す。この表の値を用い、各連想シミュレーションの外れ率を以下の式で求める。

$$\text{外れ率} = \frac{|\text{実験の値} - \text{指標の最大値 (最小値)}|}{\text{指標の最大値} - \text{指標の最小値}}$$

表 1: 先行研究のネットワーク指標の最大最小値

|     | n       | < k > | L     | C     | Density |
|-----|---------|-------|-------|-------|---------|
| 最大値 | 168.000 | 3.00  | 6.259 | 0.161 | 0.082   |
| 最小値 | 28.000  | 2.286 | 3.552 | 0.000 | 0.018   |

まず、今回のシミュレーション結果から、連想回数がどれくらいの場合がもっとも先行研究で得られたネットワークに近づくかを調べるために、連想回数と外れ率の比較を図 6 で行った。ここでは縦軸が外れ率、横軸が連想回数を示している。

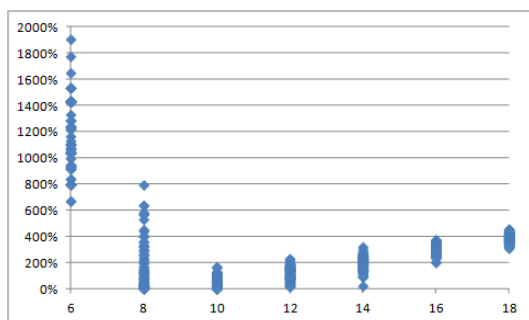


図 6: 連想回数と外れ率

各連想回数において、各々のネットワーク指標がどれほど外れているかを見るために、連想回数毎に、各指標の外れ率を合計した。表 2 に連想回数毎のネットワーク指標の外れ率合計を示す。表中の外れ率計は、ネットワーク指標毎の外れ率合計をさらに合計したもので、その連想回数の持つネットワーク指標のトータルの外れ率である。また、連想回数の変化による各指標の変化を明らかにするために、各設定での 100 回のシミュレーション結果のネットワーク指標の値の平均であるネットワーク指標平均を表 3 に示す。

表 2: 連想回数毎のネットワーク指標の外れ率

| 連想回数 | n      | < k > | L     | C     | Density | 外れ率計    |
|------|--------|-------|-------|-------|---------|---------|
| 6    | 6.99   | 0.00  | 29.89 | 0.08  | 1101.91 | 1139.67 |
| 8    | 0.31   | 0.24  | 23.58 | 6.03  | 112.20  | 142.38  |
| 10   | 0.00   | 2.39  | 25.05 | 24.62 | 0.00    | 52.06   |
| 12   | 0.07   | 3.93  | 35.21 | 77.05 | 0.30    | 116.56  |
| 14   | 27.13  | 6.68  | 35.71 | 82.67 | 47.58   | 199.77  |
| 16   | 85.71  | 10.25 | 31.15 | 63.05 | 112.75  | 302.91  |
| 18   | 150.43 | 1.16  | 24.88 | 45.45 | 153.14  | 385.06  |

図 6 から、連想回数が 10 回までは外れ率は大きく低下するが、10~18 回ではなだらかに上昇しているといえる。また、表 3 から、ノード数 n と Density は連想回数の増加に相関して増加し続けるが、他の指標はほぼ一定の値域に収まることがわかる。Density は n に依存するため、連想回数の増加によってノード数の増大が起こっていると推測される。今回のアルゴリズムでは、連想回数 10 回前後の場合、先行研究によって得られたネットワークに近い結果が得ると期待できる。しかし、

表 3: 連想回数毎のネットワーク指標平均

| 連想回数 | n      | < k > | L    | C    | Density |
|------|--------|-------|------|------|---------|
| 6    | 18.22  | 2.61  | 2.74 | 0.02 | 0.14    |
| 8    | 35.891 | 2.76  | 2.93 | 0.09 | 0.08    |
| 10   | 72.95  | 2.91  | 2.89 | 0.17 | 0.04    |
| 12   | 134.25 | 2.97  | 2.60 | 0.28 | 0.02    |
| 14   | 205.35 | 3.03  | 2.59 | 0.29 | 0.02    |
| 16   | 288.00 | 3.07  | 2.71 | 0.26 | 0.01    |
| 18   | 378.60 | 3.08  | 2.88 | 0.23 | 0.01    |

表 2 を見ると、比較的優れている連想回数 10 回の場合でも、大きな外れ率であるのは L, C である。表 1 と表 3 を比較すると、C は基準値より大きく、L は小さい。これは、今回のアルゴリズムによって生成されたネットワークは、基準と比較して複雑であり、広がりを持っていないことを意味している。今回のアルゴリズムでは、次数の高いノードを選びやすいために、結果的に狭い領域に連想が留まってしまうことが推測される。これらのことから、アルゴリズムに連想先が次数の高いノード近傍に留まらないようなルールを組み込むことが必要であると考えられる。

## 5. おわりに

本研究では、WordNet の解析結果を参考にして作成した連想シミュレーションを実行し性能を評価した。そして、シミュレーションの連想回数を調整することで先行研究で得られたネットワークとの一致率が向上するという知見を得た。今回の実験では、連想の行き先がある程度決まってしまうが、行き先の選択にばらつきを持たせることで、実際の人間の連想に近い意味ネットワーク上での連想シミュレーションを実現させることが可能であると考えられる。また、今後は日本語 WordNet や Wikipedia からの大規模な意味ネットワークを利用し、連想シミュレーションの精度を上げ、アプリケーションとしてブラッシュアップすることを予定している。

## 参考文献

- [Minsky 86] ミンスキー, M : 認知科学の基底, 産業図書 (1986)
- [Minsky 06] Minsky, M : The Emotion Machine : Common-sense Thinking, Artificial Intelligence, and the Feature of the Human Mind, Simon & Schuster (2006) (邦訳 : ミンスキー博士の脳の探検, 竹林洋一 訳, 共立出版 (2009))
- [Barabasi 04] Barabasi, A. -L., Oltvai, Z. N., "Network Biology: Understanding the Cell's Functional Organization", Nature Reviews Genetics 5, pp. 101-113 (2004)
- [中島 08] 中島秀之 : 構成的研究の方法と学問体系 シンセシオロジーとはどういう学問か?, Synthesiology, Vol.1, No.4, pp.305-313 (2008)
- [田浦 11] 田浦俊春, 山本英子, Nor Fashiha Mohd Yusof, 伍賀正典, 永井由佳里, 中島秀之 : デザインにおける創造的思考の構成論的研究の試み 概念生成プロセスの構成論的シミュレーション, 認知科学, Vol. 18, No.2, pp.329-341 (2011)