

意味空間上の演算は実世界における計算論的創造性に不可欠か？ 機械学習によるコンセプト創出手法の検討

Is an Operation on the Semantic Space Essential for Computational Creativity in Real-World Tasks? Concept Creation using a Machine Learning Technique

須藤 明人[†], 藤原 直哉[‡], 徳田 慶太^{†††}, 本田 秀仁^{‡‡}, 植田 一博^{‡‡}

Akihito Sudo, Naoya Fujiwara, Keita Tokuda, Hidehito Honda, Kazuhiro Ueda

[†] 東京大学生産技術研究所, [‡] 東京大学空間情報科学研究センター, ^{†††} 東京大学医学部附属病院 薬剤部, ^{‡‡} 東京大学総合文化研究科,
the University of Tokyo
sudoa@iis.u-tokyo.ac.jp

Abstract

Automatic creation of concepts is important for various situations. Previous researches in the conceptual blending and the concept invention proposed the cognitive models which represent the process by which people combine concepts and those relationships. However, those researches do not allow one to create new concepts automatically because those models can not discover which concepts should be combined to create successful concepts. Therefore, we show the hypothesis that there is the law governing the process of concept creations which is represented by the semantic network. In the experiment, the hypothesis is verified by letting machine learning model extract the law using the training data extracted from the semantic networks built with document set written before and after the past successful concepts. After extracting such law, the machine learning model can provide new concept in accordance with its law.

Keywords — Computational Creativity, Concept Invention, Machine Learning

1. はじめに

創造的なアイデアには様々な形態があるが、なかでもコンセプトはその汎用性や結果への影響力という観点で重要である。コンセプトの創出は、「音楽や絵のコンセプト」「新商品のコンセプト」のように創造性を要するタスクの多くで共通して行われている。また、コンセプトは作品や事業等の全体を貫く基本的な概念 [1] であり、新しい作品や事業等の実現に向かう筋道を決める役割を果たすため、コンセプトの良否は結果に大きな影響を与える。

コンセプトの生成は、これまで主に Computational Creativity (以下、計算論的創造性) [2] と呼ばれる分野で、無関係な概念を合成する認知過程と合成された概念の意味に焦点を当てて研究されることが多く、これらの研究は芸術や数学といった領域を対象することがほとんどで、ビジネス等の領域でコンセプトを生成するには至っていない。Fauconnier ら [3] は、概念の合成を日常的な思考や言語運用の基本的な認知過程のひとつであると主張し、概念や概念間の関係からなる空間を結合する操作が無意識下で実行されるというモデルを提案した。Schorlemmer ら [4] は、概念合成を数理的に厳密なものとし、計算機での実装を可能にすることを目的に、[3] のモデルをベースに、多様な概念モデルとあらゆる論理推論モデルに包含することが可能な概念合成のモデルを構築するプロジェクトについて報告している。これらのモデルには一般性があると主張されてはいるものの、数学や音楽といった特殊な対象を題材にした議論が展開されている。また、これらの研究は、新たな概念を作成する際の元となる概念群が所与とされているため、無数に存在する概念が動的に変化するような実世界のタスクでは組合せ爆発が起きるため適用に限界がある。例えば、ビジネスでの商品開発や、政府による社会問題の解決といったタスクがこのような実世界のタスクの実例である。

一方、このようなビジネス等の実世界の問題に、いわゆる人工知能技術を適用する研究が近年活発に行われている。この背景には機械学習等の逆推論技術のブレイクスルーと、利用できるデータの飛躍的な増加がある。逆推論技術と膨大なデータを活用する問題解決の方法論はデータ駆動 [5] と呼ばれ、画像や音声の認識、医療分野での診療や新薬開発、対話型エージェント、自然科学における法則発見の自動化といった幅

広い領域で研究されている。データ駆動のアプローチは、記号やルールを設計を行わない点で、設計者が記号やルールを定義した世界で知能を発揮させる伝統的な記号ベースの人工知能の研究と異なっている。多くの現実のタスクでは、明示的に分節化されていない無数の構成要素が動的に相互作用するため、あらかじめ記号やルールを与えておくアプローチには限界がある[6, 7]。このようなタスクを以下では“実世界問題”と呼ぶ。データ駆動アプローチの成功は、膨大なデータに埋め込まれている対象領域の知識を機械的に抽出するアプローチの実世界問題における有効性を示しているといえる。

コンセプトの創出についても、成功するコンセプトに共通したルール（つまり法則性）が過去のデータに埋め込まれている可能性がある。データ駆動型のアプローチを用いて、成功したコンセプトの創出の法則性をデータから抽出することができれば、その法則性にしたがって生成することにより質の良いコンセプトが得られると考えられる。そこで本論文では、意味ネットワークの経時変化で表現されたコンセプトの創出過程の計算論的なモデルを提案し、そのモデルの枠組みのもとで機械学習で過去のデータの抽出を試みることにより、“成功した新商品のコンセプトの生成過程に共通した法則性が存在し、少なくともその法則性の一部は意味ネットワーク上の演算で表現される”という仮説の検証を行う。この仮説が検証されれば、過去に成功したコンセプトに共通する法則性を抽出した学習機を用いて、まったく新しいコンセプトを機械的に生成できるようにもなる。

コンセプトの生成パターンを機械学習で抽出するためには、コンセプトとその生成過程を計算論的なモデルで表現する必要がある。人の知識表現としての妥当性の研究[8, 9]や、意味ネットワークの工学的な柔軟性をふまえた研究がなされてきたこと[10]をかんがみ、意味ネットワークとその接続関係の変化を用いて、コンセプトとその生成過程を表現する。

意味ネットワークで表現されたコンセプト生成の法則性を機械学習で抽出する問題は、一般にはコンセプト登場前後の意味ネットワークの経時変化をモデルの入出力とすることで、ネットワークをネットワークに写像する関数の獲得を機械学習で行うことになる。しかし、入出力がともにネットワークであるようなデータを学習することが難しいので、ネットワークの接続を推定するリンク予測問題として定式化することで、機械学習に解ける問題とした。

なお本論文では、単にコンセプトを「作品や商品の

全体につらぬかれた、骨格となる発想や観点」[1]という意味に限定して用いる。例えば「新しい商品のコンセプトを議論する」といった用法である。また、概念と述べた場合は「物事の概括的な意味内容」という意味で用いる。この場合、概念としての犬は、4本足であり、尾を持ち、ペットとして人気があり、という共通の意味を持った、個々の犬の集合である。

2. 本研究の仮説

本研究は、成功するコンセプトが生成されてきた背後に何らかの共通した法則性が存在していることを前提としている。さらに、その法則性の一部は意味ネットワークで表現できる（つまり意味ネットワークから抽出可能な説明変数で記述される）ことが仮定されている。より具体的には、コンセプトの登場前後の意味ネットワークの経時変化で記述される説明変数で、法則性が説明されるという仮定である。これは、必ずしもコンセプトが生成される法則性の全てを意味ネットワークで表現できるということではない。しかし、コンセプトの生成の法則性の一部は、意味ネットワークの経時変化を表す何らかの演算であることになる。これは、実世界問題における創造的なアイデアの重要な形態であるコンセプトの生成の自動化に、意味ネットワーク上での演算が必要であることを示唆する。その法則性を計算機に実装することで、成功するコンセプトに共通した法則性にしがったコンセプトの自動生成が可能になる。

この仮説を検証するためには、過去に成功したコンセプトに関するデータを用いて、実際にそのコンセプトの生成パターンに法則性があることを確認すればよい。5章の実験では、過去のコンセプトのデータを学習することで法則性の獲得をこころみた学習モデルと、ランダムな識別しか行わないモデルを比較することで、法則性の有無の検証を行う。

3. コンセプト創出過程のモデル

本章では、概念とコンセプトを計算論的な表現で定義し、その後、コンセプトの創出過程を計算論的にモデル化する。まず、概念とその意味を表現するモデルとして、意味ネットワークのノードとその接続関係を用いる。意味ネットワークではノードが概念に対応しており、意味的に関連のある概念同士がエッジで接続されている。本研究では、概念の意味を直接エッジで接続された概念の集合とする。このとき、 V を概念の集合、 E を概念間の繋がりであるエッジの集合として、意味ネットワークは $S = (V, E)$ と書け、ある概念

$\hat{v} \in V$ の意味 $V^{\hat{v}}$ をエッジで直接つながった概念の集合 $V^{\hat{v}} = \{v | \bar{e}(v, \hat{v}) \in E\} = \{v | v \sim \hat{v}\}$ と書ける. ここで, $\bar{e}(v', v'')$ は v' と v'' のエッジを表し, $v' \sim v''$ は v' と v'' がエッジで接続されていることを示す.

次に, コンセプトを意味ネットワークを用いて定義する. 本研究で扱うコンセプトは「高級感がある日本車」といった, 新しい意味付け(“高級感”)がなされた概念(“日本車”)である. よって, 新しい意味付けがなされた概念であるコンセプトの意味ネットワークでの表現は, 従来は直接つながっていなかった概念との新しい接続が付与されたノードとなる. 例えばレクサスであれば, 図1のように, 発売前には“信頼性”, “耐久性”といった概念が“日本車”の意味として意味ネットワーク上で接続されていたのに対し, 新しいコンセプトは“高級感”, “デザイン性”, “重厚感”, “富裕層”といった新たな概念が“日本車”に接続された意味ネットワークとして表現できる.

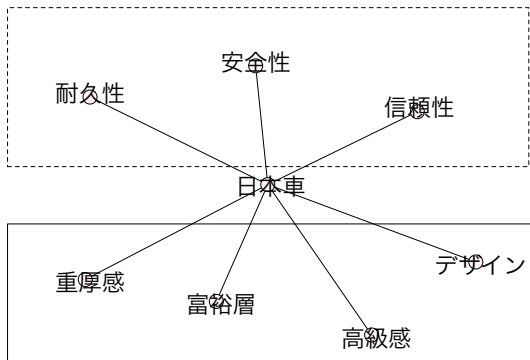


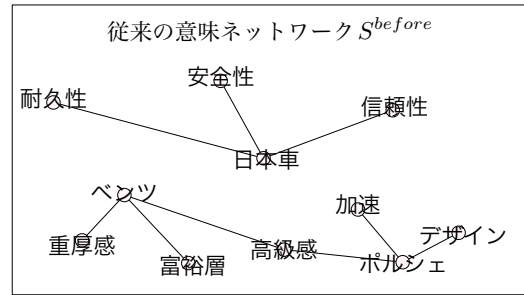
図1 “レクサス”のコンセプトの例. 実線が新たに付与された意味である.

以上から, コンセプトは意味ネットワークを用いて次のように定義できる.

Definition 1. コンセプトの計算論的表現

意味ネットワーク (V, E) と概念 v^{target} が与えられたとき, $v^{target} \in V$ の新しいコンセプトとは, 概念 v^{target} のもともとの意味 $V^{v^{target}} = \{v | v \sim v^{target}\}$ には含まれない新たな概念を含む概念集合 V^{new} と, V^{new} と v^{target} の接続からなる意味ネットワークである.

このコンセプトの定義のもとで, コンセプトを創出する過程は既存の概念に対して適切な新しい接続を発見することに他ならない. 図1の“レクサス”の例でいえば, 従来の“日本車”とは接続されていなかった“高級感”や“重厚感”といった概念を見いだすことが, “レクサス”のコンセプトの創出過程である. 従って, コンセプトの創出過程は次のように意味ネットワーク



$$\Psi(S, v^{日本車}) = \{\bar{e}(v^{日本車}, v^{高級感}), \bar{e}(v^{日本車}, v^{信頼性}), \dots\}$$

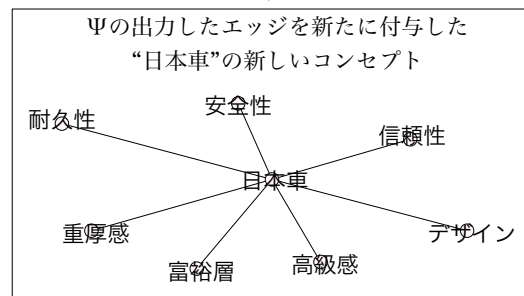


図2 “レクサス”のコンセプトの創出過程の例

の新しい接続を与える関数 Ψ で与えられる.

Definition 2. コンセプトの創出過程

意味ネットワーク $S = (V, E)$ と概念 v^{target} が与えられたとき, v^{target} のコンセプトの創出過程は, 意味ネットワークと概念のペアから新しい概念との接続を与える関数

$$\Psi : (S, v^{target}) \mapsto E^{new} \tag{1}$$

で表現される. ここで, E^{new} は対象の概念 v^{target} の新しい接続 $\{e | \bar{e}(v, v^{target}), v \in V^{new} \setminus V^{v^{target}}\}$ である.

図2は, “レクサス”のコンセプト創出過程の例である. 従来の意味ネットワークにおいて, “日本車”は“信頼性”や“耐久性”を有している一方, “ベンツ”や“ポルシェ”の持つ“高級感”は有していない. この意味ネットワークを Ψ に入力して得られる新たな接続 $\bar{e}(v^{日本車}, v^{高級感})$ や $\bar{e}(v^{日本車}, v^{信頼性})$ 等を, “日本車”のノードに新たに接続することで, 高級感を兼ね備えた新しい日本車であるレクサスのコンセプトが得られる.

4. データからのコンセプトの創出

4.1 意味ネットワークの自動構築

本研究では学習機に与える学習データの作成のために、成功したコンセプトが登場した前後の意味ネットワークが必要である。意味ネットワークの構築は手作業では大変手間のかかる作業のため、自動的に構築できることが望ましい。

エッジに属性が付与されず、ノードに概念が紐付いたような意味ネットワークを構築するためには、すべての概念の集合と概念のあいだの接続関係がわかればよい。すべての概念が与えられ、その概念の任意のペアの意味的なつながりの有無（“日本車”と“燃費”は接続され、“日本車”と“辛い”は接続されない等）がわかれば、概念に対応するノードの集合を V 、つながりのある概念のペアのあいだのエッジの集合を E とすることで、意味ネットワーク (V, E) を構築できる。

概念の集合と、概念のあいだの接続関係は、形態素解析と統計的意味論 [11, 12] と呼ばれる手法を用いることで、十分な量の文書データから得ることができる。まず、文書に形態素解析で分かち書き処理（膠着語である日本語の文章を、文の最小単位である形態素に分割する処理）をほどこし、形態素を概念として扱うことにすれば、その文書データに含まれる全ての概念を得ることができる。十分に多くの文書データを収集すれば、その十分な規模を持った文書データには対象とする概念やその概念に新たに付与されるであろう概念が全て含まれているので、得られた概念に対応するノード群を意味ネットワークのノード集合 V として用いる。次に、統計的意味論の手法を用いると、それぞれの概念の実数値ベクトルでの表現を得ることができる。多くの統計的意味論の手法は、概念の意味の近さがベクトルの間の距離となるようにベクトル表現が与えられる。そこで、ふたつの概念 v_1 と v_2 のベクトル表現 v_1, v_2 の距離 $d(v_1, v_2)$ がしきい値 λ より小さい場合、すなわち $d(v_1, v_2) < \lambda$ である場合に、 v_1 と v_2 に対応するノードをエッジでつなげば、意味の近い概念が接続された意味ネットワークを得ることができる。 λ はハイパーパラメーターであり、距離 $d(\cdot, \cdot)$ には例えばコサイン類似度を用いれば良い。

4.2 リンク予測によるコンセプト創出

定義 2 によれば、コンセプトの創出は意味ネットワーク上で、概念の新しい接続を見いだす処理である。人間の専門家によって新しいコンセプトを創出す

る際には、“日本車”といった対象となる概念に新たに付与するのに適切な概念を探索的に見出していると考えられる。つまり専門家は式 (1) の関数 Ψ の入出力関係を与える何らかの認知プロセスにしたがって、既存の意味関係から新しいコンセプトを生み出していると解釈できる。

文書には、書かれた際に概念が持っている意味が埋め込まれていると考えられる。例えばブラックベリーが最も売れているスマートフォンであったときに書かれた文書には、スマートフォンとタッチパネルのつながりを示唆するような文書は極めて少なく、一方、iPhone が発売された後に書かれた文章には、スマートフォンとタッチパネルは密接なつながりを持っていることが読み取れるはずである。従って、成功したコンセプトが登場した前と後に書かれた文書をそれぞれ集めた文書データから意味ネットワークを構築すれば、過去のコンセプトの意味ネットワークでの表現を構築でき、そこから意味ネットワークの経時変化に共通する法則性を抽出できると考えられる。これは、成功したコンセプトの登場前と登場後の意味ネットワークから、人間の専門家がコンセプト創出の過程で用いた Ψ の入出力関係を近似的に取り出せることを意味する。以下では、成功したコンセプトを生み出した卓越した専門家の持つ Ψ を Ψ^* と書く。

次に、 Ψ の入力と出力を定める。 Ψ の入力はコンセプトを生み出した時点で所与な情報なので、そのコンセプトが登場する以前の意味ネットワーク $S^{before} = (V^{before}, E^{before})$ と、新たなコンセプトを生み出した対象の概念 v^{target} とする。計算機がコンセプトを創出したり法則性を獲得する際には、 S^{before} は 4.1 章で述べた通りコンセプトが登場する前に書かれた文章から自動で構築することが可能である。コンセプトを考案した専門家や組織が保持していた概念や、概念の意味の完全な再現ではないが、web 等から収集された十分な量の文章から作成された意味ネットワークは、コンセプト発案の際に暗黙的に使われた概念や意味関係の良い近似になると考えられる。もうひとつの入力である v^{target} はコンセプトの対象そのものであるから所与とすることは自然であり、また過去のデータから学習データを作成する際に v^{target} を特定することは容易におこなえる。

Ψ の出力は、対象の概念 v^{target} へ付与する新しい接続 E^{new} である。学習データを作成する際には、コンセプトが登場した後の文書データから構築した意味ネットワークから新しい接続が抽出できる。まず、コンセプト登場後の文書データから意味ネットワーク

$S^{after} = (V^{after}, E^{after})$ を構築する。その意味ネットワーク S^{after} の v^{target} に対応するノードの接続のうち、従来のエッジ集合 E^{before} には存在しなかったエッジの集合が、対象とする概念の新しい接続である。

以上で Ψ の入力と出力が定まり過去のデータからこの入出力の組を抽出する方法が得られたので、入出力の組のデータ $D = \{(d_i^{in}, d_i^{out})\}_{i=1}^{N_D}$ を過去の成功したコンセプトの登場前後に書かれた文書データの組から作成することができる。この D は過去に成功したコンセプトが生み出された際の入出力関係を表したデータなので、十分な性能を持つ学習機 ψ があり、 N_D が十分に大きいという仮定のもとで、 D を用いて ψ を訓練して得られるモデルが成功したコンセプトを生み出した専門家の持つ Ψ^* の良い近似になる。

一般に、新たな接続の組み合わせ V^{new} の全てを出力するモデルの学習よりも、ノードのペアの接続を個別に予測するモデルの学習のほうが機械学習の問題としては容易である。本節でここまで述べてきた問題を2値分類問題とするには、 ψ のかわりに2値分類のモデル ψ^{bin} を使い、学習データには入力の組 (S, v^{target}, v^i) と出力 $v^i \overset{new}{\sim} v^{target}$ のペアからなるデータを用いればよい。ここで、 v^i は v^{target} との接続を判定したいノードであり、 $\overset{new}{\sim}$ は、 v^i, v^{target} の間のエッジが V_D^{new} に含まれるかのラベルを2値で返す2項演算子である。例えば、接続されている場合のラベルを1として与えることにし、 v^i は v^{target} と接続されているとすれば、 $v^i \overset{new}{\sim} v^{target}$ は1である。この $v^i \overset{new}{\sim} v^{target}$ を定めるためには、 S^{new} において v^i と v^{target} が接続されているかを確認すればよい。このように、2値問題としてエッジの接続関係を予測する問題はリンク予測問題と呼ばれ、以上の議論から、コンセプト創出はリンク予測問題として定式化できることがわかった。

さらに、グラフである S を含んだ入力をそのまま与えることができる学習モデルは我々の知る限り存在しないので、 (S, v^{target}, v^i) を何らかの特徴量に変換する必要がある。通常、特徴量は実数値ベクトルで与える必要がある。良く用いられる特徴量として、 v^i, v^{target} の S における隔たりや、共通の隣接ノードの数がある。本アプローチではこの特徴量は [13] 等で報告されている指標の任意の組合せを用いることができる。以下では、 (S, v^{target}, v^i) の特徴量を与える関数を $f(v^i, v^{target}, S)$ と書く。

以上の議論から、データから得られる特徴量 $f(v^i, v^{target}, S)$ と接続の有無 $(v^i \overset{new}{\sim} v^{target})$ を学習

データの入出力の組 $\mathbf{x} \in \mathbb{R}^{N^f}, y \in \{0, 1\}$ を

$$\mathbf{x} = f(v^i, v^{target}, S) \quad (2)$$

$$y = (v^i \overset{new}{\sim} v^{target}) \quad (3)$$

とし、2値分類の識別モデルを ψ^{bin} として、 ψ^{bin} を学習データ $\mathbf{x} \in \mathbb{R}^{N^f}, y \in \{0, 1\}$ で訓練して得られる ψ^{bin} を、コンセプトの新たな概念として採用するか否かの識別に用いればよいことがわかった。

5. 検証実験

2.章で述べた仮説を検証するため、過去に成功したコンセプトが意味ネットワークの接続という観点で何らかの法則性を持って生み出されてきたことを確認する。そのためには、過去に成功したコンセプトの前後の意味ネットワークから式(2)と式(3)でデータ D を作成し、 D を学習データとテストデータに分割して評価するクロスバリデーションを行って学習モデルとランダムに予測するモデルの予測精度を比較すれば良い。すなわち、データ D を k 分割し、分割したうちの $k-1$ 個で学習モデルを訓練する。そして学習に用いていない残りのデータの \mathbf{x} をモデルに与えたときの出力が D で与えられている正解の y に一致しているかを、学習モデルとランダムモデルの両方について評価し、それらの精度を比較する。もし、学習モデルの予測精度がランダムに予測を行うモデルよりも十分に高ければ、モデルはランダムではない何らかの法則性にしがってコンセプトに付与すべき概念を推定していることになり仮説が支持される。

5.1 データと実験の設定

文書がパブリッシュされた時期を絞り込めるインターネット検索エンジンを用いて、2008年以前と2009年以降のそれぞれで“ラー油”というキーワードで検索した文書を取得した。これは“食べるラー油”の発売前と後を想定した文書セットである。それぞれの文書長は、30,326,404 単語、27,271,611 単語であった。取得した文書は Mecab [14] で形態素解析を行い、分かち書きされた個々の形態素を意味ネットワークの概念として用いた。ただし、記号等を除外する目的で出現回数が200回未満の概念は無視した。分析対象とする概念は、“食べるラー油”登場の前と後の文書に共通して現れる概念に限定した。概念の種類は最終的に7,392種類になった。概念の間の距離を得るため、word2vec [12] を用いて単語のベクトル表現を得た。概

表 1 接続予測の精度. 値は Cross Validation の平均であり, 括弧内はその標準偏差.

	Precision	Recall	F-score
RF	0.51 (\pm 0.014)	0.55 (\pm 0.050)	0.53 (\pm 0.030)
Random	0.048 (\pm 0.0037)	0.46 (\pm 0.034)	0.087 (\pm 0.0067)

念の距離をこのベクトルのコサイン類似度で測り, 小さい値を 0.2 として意味ネットワークを構築した.

リンクの接続有無を予測するモデルの学習データの特徴量には, [13] で報告されている指標を全て用いた. 2 値の識別モデルとして Random Forests (以下, **RF**) [15] を用いた. RF の適切なハイパーパラメーターは, Random Search 法 [16] で見出して用いた.

また比較のため, 接続の有無をランダムに予測するモデル (以下, ランダムモデルまたは **Random**) の精度も評価した. ランダムモデルは, ノードが接続されるかを確率 0.5 でランダムに判断するだけのモデルである.

5.2 接続予測の精度

以上の設定のもとで行った実験の精度を表 1 に示す. 問題の正負のラベルの数が大きく偏っていることをふまえて, 評価指標には, Precision, Recall, F 値を用いた. 3 fold のクロスバリデーションの平均と分散をそれぞれの評価指標について算出した. F 値は Precision と Recall の調和平均である.

表 1 によると, ベースラインであるランダムモデルは, 0.5 の確率で接続されると予測するため Recall は 0.46 とおおよそ 0.5 になっている一方, 本来は接続されないノード同士も同じく確率 0.5 で接続すると予測するため Precision は 0.048 と低い値となった. 結果, Precision と Recall の調和平均である F 値も 0.087 と低い精度となった. 一方, 学習データからヒット商品のコンセプトで接続されたノードを学習した RF では, Precision がそれぞれ 0.51 とランダムモデルより明らかに良い精度となり, Recall も 0.55 と 0.5 をこえる結果となった. F 値は 0.53 となり, ランダムモデルの 0.087 を明らかにこえる精度となった. 学習モデルがランダムモデルよりも十分に高い精度であったことは, コンセプト創出におけるリンクの接続のされかたに, [13] の特徴量を説明変数とする法則性が存在していることを示唆する結果であり, 2. で述べた仮説を支持する結果である. RF は過去のデータの学習によりその法則性を獲得したことになる. ただし, 精度そのものは 50%前後であり, 完全な自動化には精度の向上が必要な水準である.

6. まとめ

成功したコンセプトの生成に共通した法則性が, コンセプトの登場前後の意味ネットワークの経時変化で記述される説明変数で記述されるという仮説を提示し, 機械学習で過去に成功したコンセプトに関する意味ネットワークの変化を学習することを通じてその仮説を支持する結果を得た. 学習を通じてこの法則性を獲得したモデルを用いて, 過去のコンセプトの法則性にしたがった新しいコンセプトを生成することができる.

今後の課題として, 接続関係の予測精度の向上, よりリッチなコンセプト表現の活用, 収集した文書データの質と量と精度の関係の分析があげられる.

参考文献

- [1] 松村 明 (2016). “デジタル大辞泉”.
- [2] Jon McCormack, Mark dInverno (2012). Computers and creativity: The road ahead. Computers and Creativity, 421424. Springer.
- [3] Fauconnier, G. & Turner, M. (2008). The way we think: Conceptual blending and the mind’s hidden complexities. Basic Books.
- [4] Schorlemmer, M., Smaill, A., Kuhnberger, K. U., Kutz, O., Colton, S., Cambouropoulos, E., & Pease, A. (2014). Coinvent: Towards a computational concept invention theory.
- [5] 樋口 知之 (2016). データ駆動科学技術を担う人材の育成 確率的思考と逆推論. “情報管理”, 59 (1), 5356.
- [6] Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. Trends in cognitive sciences, 7 (2), 8491.
- [7] 谷口 忠大 (2014). “記号創発ロボティクス 知能のメカニズム入門”. 講談社.
- [8] Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. Psychological review, 82 (6), 407.
- [9] Anderson, J. R. (2000). Learning and memory.
- [10] Oh, J., Kim, T., Park, S., & Yu, H. (2012). PubMed search and exploration with real-time semantic network construction. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 15721575. ACM.
- [11] Dumais, S. T. (2004). Latent semantic analysis. Annual review of information science and technology, 38 (1), 188230.
- [12] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 31113119.

- [13] Lu, L. & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390 (6), 11501170.
- [14] Kudo, T., Yamamoto, K., & Matsumoto, Y. (2004). Applying Conditional Random Fields to Japanese Morphological Analysis.. *EMNLP*, Vol. 4, 230237.
- [15] Breiman, L. (2001). Random forests. *Machine learning*, 45 (1), 532.
- [16] Bergstra, J. & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb), 281305.