

# コモンセンス知識と推論を用いた幼児教材「仲間外れ概念探し」問題への 取組みと評価

An evaluation of commonsense knowledge-based reasoning approach to the problem of recognizing the dissimilar concept from the concepts sets taken from the educational materials for human child

中原 和洋\*<sup>1</sup>  
Kazuhiro Nakahara

内田 咲\*<sup>1</sup>  
Saki Uchida

小林 実央\*<sup>1</sup>  
Mio Kobayashi

山田 茂雄\*<sup>1</sup>  
Shigeo Yamada

\*<sup>1</sup> 日本ユニシス株式会社 総合技術研究所  
Technology Research & Innovation, Nihon Unisys, Ltd.

With the development of the knowledge base, ConceptNet, and its reasoning algorithm AnalogySpace, the basis of the commonsense AI is now ready for the further research to be carried out. The authors had been involving in the acquisitions of Japanese commonsense knowledge and the development of the knowledge base. This paper proposes a method to assess the commonsense reasoning techniques that use our knowledge base and inference algorithms, and reports the results of the assessment. The assessed task was to make inference in solving the problem of identifying the most dissimilar concept from the sets of concepts taken from the educational materials for human child aged 3-4. In the evaluation, a comparison was made with the results brought by the method of Latent Semantic Analysis using Wikipedia and the method of similarity calculation using WordNet, demonstrating the superiority of our commonsense AI approach over those existing methods.

## 1. はじめに

人工知能研究において、コモンセンスを対象とした知識ベースや推論は古くから重要な課題の1つとして認識されてきた。2000年代に入り、コモンセンス知識ベース ConceptNet[Havasi 07]や推論手法 AnalogySpace[Speer 08]が開発されるなど、コモンセンス AI 研究の基盤が整備されつつある。筆者らも日本におけるコモンセンス知識の獲得を進めてきた。一方で、ある時点におけるコモンセンス知識ベースを使った推論の性能を客観的、定量的に評価することは、研究の方向性や課題抽出、応用先の検討等の様々な観点から重要である。本論は、筆者らが日本で収集したコモンセンス知識と AnalogySpace を利用した概念間類似度推定の客観的、定量的な性能評価および評価手法の確立を目的とする。市販されている幼児向け教材の「仲間外れ概念探し」問題を利用し、正答率などを評価した。

本論の構成は以下の通りである。2章で ConceptNet および AnalogySpace コモンセンス推論、日本における知識獲得研究について、3章で評価方法について、4章で評価結果、5章で考察、6章でまとめを述べる。

## 2. コモンセンス知識ベースと推論

本章では、評価実験で利用した ConceptNet および AnalogySpace、筆者らが行ってきた日本におけるコモンセンス知識収集について記述する。

### 2.1 ConceptNet

ConceptNet は、マサチューセッツ工科大学メディアラボ(MITメディアラボ)が開発中のコモンセンス知識ベースである。ConceptNet では、概念(Concept)をノード、概念間の関係(Relation)をアークとした表明(Assertion)の集合(意味ネットワーク)でコモンセンス知識を表現する。概念はそれを表す単語

や短いフレーズで表現し、関係は IsA, HasProperty, PartOf, Desire などあらかじめ規定されたものを用いる。ConceptNet では、意味表現(Assertion)と表層表現(Sentence)を対応づけてデータを保持している(図1)。

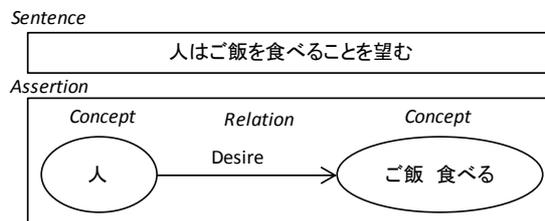


図1: ConceptNet の知識表現

### 2.2 AnalogySpace と概念間類似度計算

AnalogySpace は、MIT メディアラボが開発した ConceptNet に基づいたコモンセンス推論手法である。AnalogySpace は、ConceptNet のコモンセンス知識ベースに対して主成分分析(SVD と次元削減)を適用した手法である。ConceptNet の Assertion を Concept-Feature マトリクス  $A$  に変換する。Feature は Concept と Relation の組である。このマトリクス  $A$  に対して SVD と次元数  $k$  による次元削減を行うことで得た  $A$  の近似  $A_k = U_k S_k V_k^T$  によって、 $A$  の 0 成分であった ConceptNet には含まれない Concept-Feature の組(未知の Assertion)の類推値を得ることができる。また、Concept 間の類似度計算を  $U_k S_k \cdot (U_k S_k)^T$  により算出する。AnalogySpace を提案した論文[Speer 08]では限定的なサンプリングによる推論の正確性の評価にとどまっておき、人間の持つコモンセンスに比べた推論性能という観点では、客観的、定量的な評価はなされていない。

## 2.3 日本におけるコモンセンス知識収集

コモンセンス知識は当たり前すぎて Web 文書等には表明されにくい傾向にあるため、これまで人手による知識ベース化方法が取られてきた。Cyc[Lenat 95]では知識専門家の手によって、OMCS[Singh 02]ではインターネット上のボランティアによって、また近年では Game With A Purpose(GWAP)によりインターネット上のゲームによって知識ベース化をスケールさせる方法も行われている[Lieberman 07][Kuo 09]。筆者らも、日本におけるコモンセンス知識獲得を目的とした2つのインターネットゲームサイトを立ち上げて知識ベース化を進めてきた。

### (1) ナージャとなぞなぞ

2010年に開始したインターネットゲームサイト「ナージャとなぞなぞ」[中原 11]は、ナージャというキャラクターとインターネット上のプレイヤーの間で行われる連想ゲームで、ナージャが出す5つのヒントをもとに彼女が頭に思い浮かべることをプレイヤーが当てるゲームである。プレイヤーの回答がコモンセンス知識として収集される。2013年7月時点で、24万件のコモンセンス知識(Sentence数)を獲得している。

### (2) 日本人検定

2012年に開始した Facebook アプリの「日本人検定」[中原 13]は、自分の「日本人レベル」を調べることができる一回20問のクイズ形式のゲームである。回答者全員の回答から「日本人っぽさ」を導き出して、その得点を Facebook の友人たちと互いにシェアして競い合う。ユーザの回答内容がコモンセンス知識として収集される。2013年7月時点で、約61万件のコモンセンス知識(Sentence数)を獲得している。

## 3. 評価実験方法

本論の目的はコモンセンス推論の一つである概念間の類似度推定の性能についての客観的、定量的な評価および評価手法の確立である。市販の3~4歳児向けの幼児教材[ピグリ編集室 12] [こぐま会]における仲間外れ概念探し問題に取組み正答率等を評価することとした。

### 3.1 評価実験手順

#### (1) 教材問題のテキスト化

幼児教材の仲間外れ概念探し問題は、4ないし5つの概念を示すイラストが提示され、その中からもっとも仲間外れの概念を1つ回答する問題である。例えば、出題概念集合  $C = \{\text{犬, 猫, うさぎ, 鳥}\}$  の各イラストが与えられ、最も仲間外れの概念である鳥を回答する問題である。様々な理由付けにより異なる概念を仲間外れと見なすことも可能であるが、幼児教材では1つの正解(常識的に判断し最も典型的で自然な仲間外れ)が割当てられており、これを正解と見なすことで客観的で定量的な評価が可能になると考えた。なお、本評価では画像認識は対象外とし、著者らが人手でイラストから出題概念と正解概念のテキスト化を行い、問題の入力、評価システムの回答、正解はすべてテキストとした。人手によるテキスト化は、教材に書かれている「題意」と「正解」に従うことで客観性は担保し、イラストをテキスト表現する際の多様性に対しては以下の方針を定めて人手で対応した。

- 題意が変わらない範囲で ConceptNet に含まれる Concept のテキスト表現を用いる
- ConceptNet の複数の Concept に対応する場合は、Concept の Feature 数が最も多い Concept のテキスト表現を用いる

実際に作成した問題総数は4択問題が130問、5択問題が38問の合計168問である。

#### (2) 前処理

ConceptNet4.0には複数表現を1つの Concept(見出し語)に統一するための lemmatizer が組み込まれているが日本語の lemmatizer は貧弱であり、例えば漢字とひらがなは統合されない。筆者らは別途これらの統合処理を行った。

#### (3) 回答処理

$n$  択の仲間外れ探し問題における概念集合  $C = \{c_1, c_2, \dots, c_n\}$  から仲間外れ概念  $c_{out}$  を以下の式で導出する。

$$c_{out} = \operatorname{argmin}_{c_i \in C} \left( \sum_{\substack{c_j \in C \\ c_j \neq c_i}} \operatorname{sim}(c_i, c_j) \right) \quad (1)$$

$C$  に含まれる Concept のうち3つ以上の Concept が既知である(Concept-Feature マトリクスの行成分に存在する)場合、その問題に対して回答可能とし、既知の Concept 集合内のみで  $c_{out}$  を計算しシステムの回答とした。既知の Concept が3未満であった場合は、その問題に対して回答不能とした。

#### (4) 回答の採点と評価指標の導出

全問題総数を  $N=168$ 、 $N$  のうち回答可能であった問題数を有効回答数  $N_a$ 、正答した問題数を  $N_c$  とした時に、性能評価指標として、有効回答率  $RR=N_a/N$ 、有効回答正答率  $RCR=N_c/N_a$ 、正答率  $CR=N_c/N$  を導出した。

## 3.2 評価実験システム

評価実験システムは、コモンセンス知識ベースに ConceptNet4.0<sup>1</sup>を用い、概念間の類似度計算に AnalogySpace の実装である Divisi2<sup>1</sup>を用いた。仲間外れ概念の回答には式(1)を用いた。以降本手法をコモンセンス手法と呼ぶ。

#### (1) コモンセンス知識セット

知識量や、知識収集手法毎に異なる知識セットを作成し、評価を行った。All は2013年7月時点のすべての知識セット、Mid は All から50%を、Small は All から10%の Sentence をランダムサンプリングして生成した知識セットである。Default は ConceptNet4.0 の配布版で提供されている日本語知識セット、Nadya はナージャとなぞなぞで、Kentei は日本人検定で収集した知識のみで生成した知識セットである。それぞれの知識セットの Sentence 数、Assertion 数、Concept 数を表に示す。

表 1: 評価実験で利用したコモンセンス知識セット

知識セット名	Sentence 数	Assertion 数	Concept 数
All	868,228	181,820	65,295
Mid	434,114	110,558	43,847
Small	86,823	32,160	17,590
Default	14,368	12,825	11,100
Nadya	243,010	102,424	15,462
Kentei	610,850	73,083	52,126

#### (2) テストパラメータ

結果に影響を与えるテストパラメータとして、SVD の圧縮次元数  $K = \{100, 200, 300, 400\}$ 、Concept-Feature マトリクス  $A$  の有効データとする行および列方向の最小非0成分数  $Cutoff = \{1, 3, 5\}$ 、 $A$  の正規化(各成分を行および列ベクトルのノルムで割る)の有

<sup>1</sup> <http://csc.media.mit.edu/docs/index.html>

無  $Prenorm=\{1,0\}$ , 類似度計算における  $U_k S_k$  の正規化(各成分を行ベクトルのノルムで割る)の有無  $Postnorm=\{1,0\}$  の 4 つのパラメータのすべての組み合わせについて実施した。

### 3.3 比較手法

コモンセンス手法の特長は、Web 上等の文書には表明されづらい当たり前すぎるコモンセンス知識を人間から直接獲得し利用するところにある。そこで、本評価における比較手法の1つとして、Web 上の大規模文書である Wikipedia を用いた潜在意味解析(LSA)手法を選定した。もう1つの比較手法としては、概念間の類似度計算で広く用いられている WordNet を利用した手法を選定した。それぞれの比較手法で、概念間類似度  $\text{sim}(c_1, c_2)$  を計算し、式(1)を用いて仲間外れ概念を回答する。

#### (1) 日本語 Wikipedia 記事を利用した LSA による概念間類似度計算

日本語 Wikipedia の記事を Concept, 記事内に登場する単語を Feature と捉え、Concept-Feature マトリクスを作成し、コモンセンス手法と同様のアルゴリズムを利用して、Concept (記事)間の類似度を計算した。成分値は tf-idf とした。出題概念テキストと日本語 Wikipedia 記事の対応付けは筆者らが人手で行った。形態素解析には Mecab を使い、辞書は新語などに対応した独自辞書を利用した。対象単語は名詞、動詞、形容詞とし、“する”、“ある”などの一般語は除外した。また、マトリクスの行ベクトルおよび列ベクトルの非 0 成分数が 5 未満となるデータは除外した。さらに、Wikipedia 記事には、常識的な概念とは言い難い記事が大量に存在するため、対象記事の絞りこみを行った。絞り込み方法は、Wikipedia の記事カテゴリのリンクデータ(カテゴリグラフ)を利用し、全出題概念に対応する記事カテゴリからカテゴリグラフ上の距離  $L$  以下のカテゴリに属する記事のみを対象記事とした。表 2 に示すカテゴリ距離  $L=\{0,2,4, \text{全カテゴリ}\}$  の 4 種類の記事セットを作成し、それぞれで評価した。この記事セットと、コモンセンス手法と同様のテストパラメータ, SVD 次元数  $K=\{100,200,300,400\}$ ,  $Prenorm=\{1,0\}$ ,  $Postnorm=\{1,0\}$  のすべての組み合わせで実験を行った。LSA の実装には Divisi2 を用いた。上述の手法を以降 Wikipedia 手法と呼ぶ。

表 2: 日本語 Wikipedia の LSA による比較手法

記事セット名	カテゴリ距離	カテゴリ数	記事数	単語数
WLSA_A	全カテゴリ	116,162	884,583	390,761
WLSA_4	4	4,902	150,823	224,932
WLSA_2	2	2,416	84,029	157,190
WLSA_0	0	566	34,607	78,764

#### (2) WordNet を用いた概念間類似度計算

WordNet では概念間の類似度計算手法がいくつか提案されており、nlk<sup>1</sup>に関数として実装されている 6 種類の概念間類似度計算手法(Path Distance, Leacock Chodorow[Leacock 98], Wu-Palmer[Wu 94], Resnik[Resnik 95], Jiang-Conrath[Jiang 97], Lin[Lin 98])を比較手法として用いた。日本語 WordNet1.1 [Isahara 08]の英語版 WordNet へのリンクファイルを用いて英語版 WordNet の synset へのマッピングを行った後に、nlk を用いて英語版 WordNet 上で synset 間の類似度の計算をおこなった。1つの出題概念に対して複数の synset へのリンクが存在する場合は、筆者らが人手で題意として適切な synset への対応付

けを行った。概念の情報量を必要とする類似度手法については、Brown コーパスを用いて算出した情報量を利用した。上述の手法を以降 WordNet 手法と呼ぶ。

## 4. 評価実験結果

本章では評価実験の結果について記述する。

### 4.1 最大性能の比較

コモンセンス手法と比較手法の実験結果について、それぞれ最大の正答率を出したテストケースの結果を表 3 に示す。コモンセンス手法は、知識ソース=All, Cutoff=3, K=400, Prenorm=0, Postnorm=1 である。Wikipedia 手法は記事セット=WLSA\_0, K=300, Prenorm=1, Postnorm=0 である。WordNet 手法は Resnik Similarity による結果である。表 3 の p 値は、コモンセンス手法との片側検定による p 値である。

表 3 手法毎の結果

手法名	コモンセンス	Wikipedia	WordNet
有効回答数	159	160	153
正答数	103	91	83
有効回答率	0.95	0.95	0.91
有効回答正答率	0.65	0.57	0.54
正答率	0.61	0.54	0.49
有効回答正答率 p 値		0.09153	0.03755
正答率 p 値		0.1122	0.01853

### 4.2 知識量別の結果

コモンセンス手法における知識量別の結果を表 4 に示す。また図 2 に知識量(Assertion 数)と有効回答正答率の関係を示す。

表 4 知識量別の結果

知識ソース名	Small	Mid	All
有効回答数	111	154	159
正答数	60	90	103
有効回答率	0.66	0.92	0.95
有効回答正答率	0.54	0.58	0.65
正答率	0.36	0.54	0.61

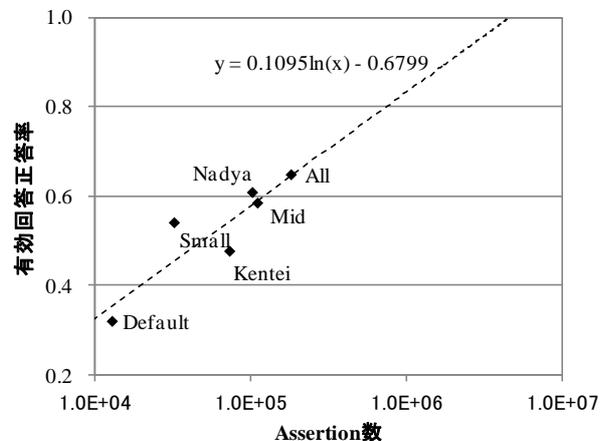


図 2 Assertion 数と有効回答正答率の関係

<sup>1</sup> <http://www.nltk.org/>

### 4.3 知識収集手法別の結果

コモンセンス手法における知識収集手法別の結果を表 5 に示す。

表 5 知識収集手法別の結果

知識ソース名	Default	Nadya	Kentei
有効回答数	25	148	86
正答数	8	90	41
有効回答率	0.15	0.88	0.51
有効回答正答率	0.32	0.61	0.48
正答率	0.05	0.54	0.24

## 5. 考察

本章では、4章の評価実験結果に基づく考察を述べる。

### 5.1 性能評価

本評価により 3~4 歳児が身につけるべき概念間の類似性を評価する能力に対して、現時点のコモンセンス手法がどの程度の性能であるかを客観的、定量的に把握可能となった。表 3 に示すように、コモンセンス手法は、Wikipedia 手法や WordNet 手法に比べ有効回答正答率や正答率について良好な結果が得られた。有効回答率についての差異は少ない。コモンセンス手法の有効回答率は 0.95 であり、3~4 歳児に問われる概念の大部分をカバーできていると言える。一方で回答不能であった 5%については今後知識ベース化を進める必要がある。5%に含まれる概念の例としては、“お年寄りに席を譲る”、“お年玉をもらう”など複数の文節で表現が必要な概念である。このような概念を扱える知識収集手法や知識表現を検討する必要がある。

各手法の回答傾向について考察する。出題問題には、種類の違いを問う問題と用途の違いを問う問題が多く含まれる。WordNet 手法や Wikipedia 手法は、種類の違いを問う問題にはコモンセンス手法に近いレベルの正答率を示す傾向が見られたが、用途の違いを問う問題にはコモンセンス手法に比べて低い正答率を示す傾向が見られた。また、Wikipedia 手法については対象記事数を増やすほど性能が悪化する傾向にあるため、今回のような常識的な判定問題に対してはノイズとなる記事が多く存在していると推測できる。従って、良い性能を出すためにはノイズ記事の除去が必要となり、コモンセンス手法に比べると安定した性能を出すためのチューニングが難しいと言える。

### 5.2 知識量と性能の関係

本評価により、知識量と推論性能の関係を把握できるようになった。図 2 に示す通り、知識量の対数オーダーの増加に合わせて有効回答正答率が線形的に向上していく傾向が見られた。また本評価により、知識の収集手法の妥当性や有効性の判断も可能となった。表 5 の結果から、筆者らが行ってきたナージャとなぞなぞや日本人検定を利用して収集した知識は、Default の有効回答正答率を上回っており、有効な知識収集手法であると判断できる。ナージャとなぞなぞと日本人検定を比べると、ナージャとなぞなぞの方がすべての指標において上回っており、より良い知識収集手法であると判断できる。

## 6. おわりに

幼児教材の仲間外れ概念探し問題を利用することで、コモンセンス推論の定量的、客観的な評価が可能となった。Wikipedia

や WordNet を利用した手法に比べてコモンセンス推論を利用した手法は良好な結果が得られた。コモンセンス知識量の増加に伴う性能の向上が見られた。

## 参考文献

- [Havasi 07] Havasi, C., Speer, R., Alonso, J.: ConceptNet 3:a flexible, multilingual semantic network for common sense knowledge, In Recent Advances in Natural Language Processing, 2007.
- [Isahara 08] Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., Kanzaki, K.: Development of the Japanese WordNet, in LREC-2008, 2008.
- [Jiang 97] Jiang, J. J., Conrath, D. W.: Semantic similarity based on corpus statistics and lexical taxonomy, in 10<sup>th</sup> Intl. Conf. Research on Computational Linguistics (ROCLING), 1997.
- [Kuo 09] Kuo, Y. L., Lee, J.C., Chiang, K., Wang, R., Shen, E., Chan C., Hsu, J.: Community-based game design: experiments on social games for commonsense data collection, Proceeding KDD-HCOMP'09, ACM, Inc., 2009.
- [Leacock 98] Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense disambiguation, WordNet: An Electronic Lexical Database, MIT Press, 1998.
- [Lenat 95] Lenat, D.: CYC: a large-scale investment in knowledge infrastructure, Communications of the ACM, ACM, Inc., 1995.
- [Lieberman 07] Lieberman, H., Smith, D., Teeters, A.: Common Consensus: A Web-based Game for Collecting Commonsense Goals, Intelligent User Interfaces, ACM, Inc., 2007.
- [Lin 98] Lin, D.: An Information-Theoretic Definition of Similarity, In Proc. of Conf. on Machine Learning, 1998.
- [Resnik 95] Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy, in IJCAI'95, 1995.
- [Singh 02] Singh, P., Lin, T., Mueller, E., Lim, G. Perkins, T., Zhu, W.: Open Mind Common Sense: Knowledge acquisition from the general public, Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems, Springer Verlag, 2002.
- [Speer 08] Speer, R., Havasi, C., Lieberman, H.: AnalogySpace: Reducing the dimensionality of common sense knowledge, AAAI, AAAI Press, 2008.
- [Wu 94] Wu, Z., Palmer, M.: Verb Semantics and Lexical Selection, in ACL'94, 1994.
- [こぐま会] こぐま会教材開発室: ひとりごとつく 29 仲間はずれ, 株式会社 幼児教育実践研究所 こぐま会.
- [中原 11] 中原和洋, 山田茂雄: 日本でのコモンセンス知識獲得を目的とした Web ゲームの開発と評価, ユニシス技報通巻 107 号, 日本ユニシス, 2011.
- [中原 13] 中原和洋: コモンセンス知識獲得を目的としたソーシャルゲーム“日本人検定”, ユニシス技報通巻 115 号, 日本ユニシス, 2013.
- [ピグリ編集室 12] ピグリ編集室, 伊藤恭監修: 能力育成問題集 28 仲間はずれ, 株式会社ピグマリオン, 2012.