

自動生成された常識的知識を表現する文の自然性判定

Recognizing Naturalness of Automatically Generated Commonsense Sentences

村本 晃一† Koichi Muramoto
 ジェプカ ラファウ† Rafal Rzepka
 荒木 健治† Kenji Araki

1. はじめに

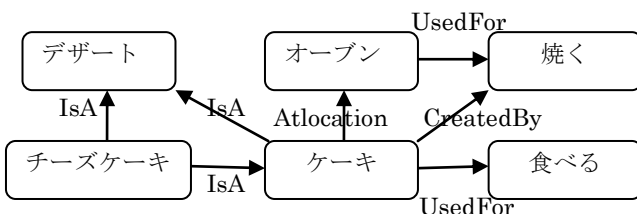
コンピュータに言葉を理解させるには、人間の持つ常識や知識といったものが不可欠である。そのため、これらの常識や知識を与えるという研究が行われている[1][2]。これらの研究は知識の対象や、質問に返答するなどの利用方法を限定することで、人手によるデータベースの作成や、知識源からの抽出を容易なものとしている。しかし、このような知識の対象や利用方法の限定を行うことは、人間の持つ知識が広い範囲にわたることに相反している。

そこで、対象を限定しない常識をデータベース化することができれば、常識を用いた様々な研究に有効に利用することが可能となると考えられる。この様な常識的知識をまとめたものとして、ConceptNet[3][4]が挙げられる。

ConceptNetとは、Open Mind Common Sense(OMCS)[5]を用いた常識的知識ベースの生成プロジェクトであり、単語がノードによって関係づけられているものである。OMCS内の自然文がConceptNetに登録されている。ConceptNetの例を図1.1に示す。図1.1における「UsedFor」や「IsA」はConceptNet中の関係を示している。これらは、MITのプロジェクトで、現在も研究が行われている。これらを用いた研究は、英語等では行われている[6]が、日本語では行われていない。これは、英語では1,035,681文もの常識的知識を表現する文が登録されているのに対し、日本語では14,546文しか登録されていないことが原因であると考えられる。この常識的知識を表現する文の量を増やすことができれば、多くの研究で利用することができると考えられる。

そこでこれまで、我々は日本語では規模の小さい常識的知識のデータベースであるConceptNetを基に、日本語の概念辞書である日本語WordNet[7]とWeb検索を用いて、常識が含まれる文を自動的に生成する手法の提案を行ってきた[8]。しかし、その精度には依然として改善の余地がある。そこで本稿では、Web検索による自然性判定の精度向上の手法の提案を行う。

図 1.1 ConceptNet の例



2. 常識的知識を表現する文の生成

2.1 常識的知識

例えば、「犬は歩く」は常識であるが、「東大寺は日

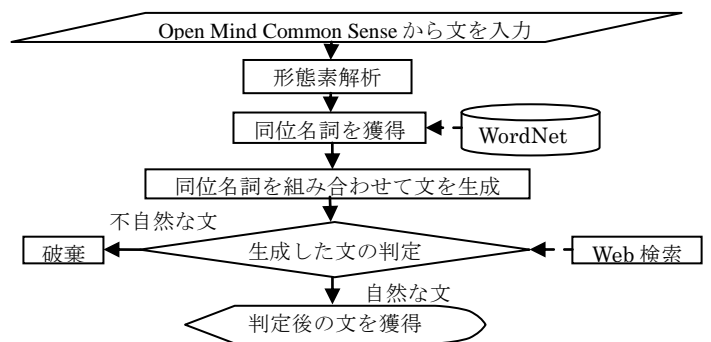
†北海道大学大学院情報科学研究科

本の寺の具体例だ」は日本についての教育を受けた者にとってのみ常識となる。本稿では、このような、人間が生活していく上で獲得する常識に加え、文化的な知識や専門的な知識も含む知識を常識的知識と呼ぶ。

2.2 概要

入力にOMCSの文を、文の生成に日本語の概念辞書である日本語WordNetを、生成文の正誤判定にWeb検索を用いている。システムの概要を図2.1に示す。OMCSの文に含まれる全ての名詞に対して日本語WordNetを用いて同位名詞を獲得する[8]。同位名詞とは、名詞の上位概念の下位概念に当たる名詞のことである。その後基の文における名詞部分を獲得された同位名詞に置き換えることで、複数の文の生成を行う。例えば、「ライオンは吠える」という文を基にする場合には、この文に含まれる名詞「ライオン」の同位名詞である「虎」、「獅子」を得る。これらの同位名詞を用いて「虎は吠える」、「獅子は吠える」という文を生成する。その後、獲得された文からノイズの除去を行う。ノイズの除去にはWeb検索を利用する。Web検索を行った後、2.6で述べる閾値を設定し、自然な文を獲得する。Web検索にはYahoo!JAPAN¹を利用する。

図 2.1 常識的知識を表現する文の生成システム概要



2.3 日本語 WordNet

WordNetとは、プリンストン大学で開発されたPrinceton WordNetのことで、単語にsynsetと呼ばれる概念が付与されている意味辞書である。類似した単語にはそれぞれ同じsynsetが付与されている。synsetは「ID番号-品詞」という形で示されており、品詞部分には例えば名詞なら「n」、動詞なら「v」などのように、アルファベットが割り振られている。またそのsynset同士は上位概念や下位概念等数多くの関係によって関連付けられている。今回は、このPrinceton WordNetのsynsetに対して日本語を付与した日本語WordNetを用いた。日本語WordNetには、57,238のsynset、93,834語の単語、synsetと単語のペアが

¹ Yahoo! JAPAN <http://www.yahoo.co.jp>

158,058組登録されている。日本語 WordNet に含まれている単語と関係の例を表 2.1, 表 2.2 に示す。

表 2.1 WordNet に登録されている単語の例「家」

家					
synset	synset に定義されている単語				
03544360-n	家屋	住宅	住居	ハウス	館
08078020-n	家族	世帯	家庭	ファミリー	一家

表 2.2 WordNet に登録されている関係の例「家」

家					
上位概念	建築物	ビル	建造物	ビルディング	建物
下位概念	お屋敷	小屋	平屋	下宿	宿屋
一部	図書室	書斎	ロフト	屋根裏部屋	書房

2.4 同位名詞獲得と文の生成

基となる文の同位名詞を獲得することにより、常識的知識を表現する文を生成する[8]。例えば、「ボールを投げる。」を基の文として「速球を投げる。」、「フリスビーを投げる。」、「すべり台を投げる。」等を生成する。しかし、このようにして獲得した文には、常識的知識とは異なる、常識的には不自然な知識が非常に多く含まれている。この例でいえば、「すべり台を投げる」等がこれに相当する。これらの常識的には不自然な知識をノイズとする。これらのノイズの除去を行うために、Web 検索を利用する。詳しくは 2.5 で述べる。

2.5 ノイズの除去

2.4 のようにして獲得された文から、ノイズの除去を行う。ノイズ除去には、Web 検索を利用する。ここでは、ノイズ除去の手法として「スニペット共起回数」、「完全一致割合」、「活用件数割合」の3つを提案する。

2.5.1 スニペット共起回数

「スニペット共起回数」[8]の説明を以下に述べる。まず、検索クエリを生成する。生成文から名詞・形容詞・動詞を抽出し、名詞に生成文と同様の助詞を付加した上で、「名詞.助詞」「動詞(形容詞)」というような検索クエリを生成する。例えば、「ボールを投げる」という文から「ボールを」「投げる」という検索クエリを生成し Web 検索を行う。

このように生成した検索クエリを用いて Web 検索を行う。検索結果のスニペット中の一文に、生成した検索クエリが同時に同順序で出現する回数を算出する。この回数を以後、「スニペット共起回数」と呼ぶ。一文とは、句点、感嘆符、疑問符、ドット(…)で挟まれている部分のことである。スニペット共起回数について閾値を設定し、スニペット共起回数が閾値を下回った場合、ノイズと判定する。

2.5.2 完全一致割合

「完全一致割合」の説明を以下に述べる。まず、検索クエリを生成する。生成文から名詞・形容詞・動詞を抽

出し、名詞に生成文と同様の助詞を付加した上で、「名詞.助詞.動詞(形容詞)」というような検索クエリを生成する。例えば、「ボールを投げる」という文から「ボールを投げる」という検索クエリを生成し Web 検索を行う。同様に、「名詞.助詞」「動詞(形容詞)」, 「名詞.助詞」, 「動詞(形容詞)」というクエリも生成し、Web 検索を行う。それぞれの検索結果と、式(1)を用いて完全一致割合を定義する。これは、生成文の完全一致件数が、「名詞.助詞」及び「動詞(形容詞)」の一致件数の中にどれほどの割合で存在するのかを算出したものである。完全一致割合について閾値を設定し、完全一致割合が閾値を下回った場合、ノイズと判定する。

$$Pp = \frac{Np}{Nn + Nv - Nc} \quad \dots (1)$$

Pp: 完全一致割合

Np: 「名詞.助詞.動詞(形容詞)」の検索件数

Nn: 「名詞.助詞」の検索件数

Nv: 「動詞(形容詞)」の検索件数

Nc: 「名詞.助詞」「動詞(形容詞)」の検索件数

2.5.3 活用件数割合

「活用件数割合」の説明を以下に述べる。日本語の検索においては、活用のある語句に対してステミングが行われていない。ステミングとは活用された語句でも同一のものとして扱い、検索を行う手法である。これを日本語に対して行うことで、正しい文の検索ではより検索件数が増えることが予想される。例えば、「本を読む」という文から「本を読ま」「本を読み」「本を読む」「本を読め」「本を読も」「本を読ん」という6つの検索クエリを生成する。これらの検索クエリについて Web 検索を行う。完全一致割合と同様に式(2)で活用件数割合を定義する。活用件数割合について閾値を設定し、活用件数割合が閾値を下回った場合、ノイズと判定する。

$$Pc = \sum \frac{Np}{Nn + Nv - Nc} \quad \dots (2)$$

Pc: 活用件数割合

Np: 「名詞.助詞.動詞(形容詞)」の検索件数

Nn: 「名詞.助詞」の検索件数

Nv: 「動詞(形容詞)」の検索件数

Nc: 「名詞.助詞」「動詞(形容詞)」の検索件数

2.6 ノイズの除去の閾値設定のための予備実験

2.5 で述べたように、ノイズの除去を行うためには、閾値を設定する必要がある。OMCS の文から名詞一つを含む文をランダムに 50 文選択する。それらに対して同位名詞を獲得し、文を生成する。これらの文から 1,000 文をランダムに選び、第一著者が評価を行った。評価基準は「0点: 不自然」、「1点: 有り得るが普通ではない・場合によっては有り得る」、「2点: 自然」の3種類である。スニペット共起回数についての予備実験[8]の結果を表 2.3 に示す。第一著者が不自然と判断した文のスニペット共起回数の平均は 7.3 回、場合によっては有り得ると判断

² 動詞と形容詞はどちらか一方しか使用しない。

した文は 11.6 回, 自然と判断した文は 16.4 回となった. そこで, スニペット共起回数が 11 回以下と算出された文を 0 点, 11 回以上 16 回以下と算出された文を 1 点, 17 回以上と算出された文を 2 点とするように, 閾値を設定した.

同様の実験を完全一致割合, 活用件数割合についても行った. 結果を表 2.4, 表 2.5 に示す. 完全一致割合では各点数の完全一致割合の平均は, 0.00185, 0.00390, 0.00414 であり, これらを閾値とした. また, 活用件数割合では各点数の活用件数割合の平均は, 775,849 件, 860,909 件, 1,349,698 件であり, これらを閾値とした. 生成した文のうち, 2 点とされた文は各データセットにおいて 22.1%~24.4%となった.

表 2.3 閾値設定のための実験結果(スニペット共起回数)

著者の評価	文数	スニペット共起回数の平均
0点	545文	7.3回
1点	211文	11.6回
2点	244文	16.4回

表 2.4 閾値設定のための実験結果(完全一致割合)

著者の評価	文数	完全一致割合の平均
0点	723文	0.00185
1点	56文	0.00390
2点	221文	0.00414

表 2.5 閾値設定のための実験結果(活用件数割合)

著者の評価	文数	活用件数割合の平均
0点	709文	0.000106
1点	60文	0.000587
2点	231文	0.00131

2.7 評価実験

2.6 で設定した閾値を用いてノイズを除去した文について, その文の自然性を被験者が評価する実験を行った. 2.5 で用いた文のうち, 閾値の設定実験に使用した文を除く文に, 設定した閾値によって 0 点, 1 点, 2 点の評価を与えた. それぞれの点数毎にランダムに 50 文ずつを抽出し, 計 150 文を用いて本実験を行う. それぞれについて被験者が 0 点, 1 点, 2 点の評価を行った被験者は全員 20 代大学生で, 理系男性 2 名, 文系男性 2 名, 文系女性 2 名の計 6 名であった.

実験結果を表 2.6 に示す. システムが 2 点と評価した文の正解率が平均 55.0%, システムが 1 点と評価した文の正解率が平均 28.0%, システムが 0 点と評価した文の正解率が平均 48.6%となった. また, 被験者の 3 人以上または 4 人以上の評価が一致した文についての結果を, 表 2.9 に示す.

同様の実験を完全一致割合, 活用件数割合についても行った. ただし, 完全一致割合と活用件数割合については, システムが 1 点と評価した文が非常に少なかったため, 今回は評価実験に加えていない. 被験者は全員 20 代男性で, 理系大学院生 3 名, 文系大学院生 2 名, 文系大学生 1 名の計 6 名であった. 実験結果を表 2.7, 表 2.8 に示す. また, 被験者の 3 人以上または 4 人以上の評価が一致した文についての結果を, 表 2.10, 表 2.11 に示す.

完全一致割合では, システムが 2 点と評価した文の正解率が平均 60.6%, システムが 0 点と評価した文の正解率が平均 21.4%となった. 活用件数割合では, システムが 2 点と評価した文の正解率が平均 58.8%, システムが 0 点と評価した文の正解率が平均 18.4%となった.

表 2.6 人手による評価(スニペット共起回数)

システムの評価	被験者の評価	被験者平均
2点	2点	27.5文
	1点	12.5文
	0点	10.0文
	正解率	55.0%
1点	2点	21.0文
	1点	14.0文
	0点	15.0文
	正解率	28.0%
0点	2点	14.7文
	1点	11.0文
	0点	24.3文
	正解率	48.6%

表 2.7 人手による評価(完全一致割合)

システムの評価	被験者の評価	被験者平均
2点	2点	31.3文
	1点	10.9文
	0点	7.8文
	正解率	62.6%
0点	2点	25.5文
	1点	13.8文
	0点	10.7文
	正解率	21.4%

表 2.8 人手による評価(活用件数割合)

システムの評価	被験者の評価	被験者平均
2点	2点	31.0文
	1点	9.5文
	0点	9.5文
	正解率	62.0%
0点	2点	28.3文
	1点	10.2文
	0点	11.5文
	正解率	23.0%

表 2.9 被験者の 3 人または 4 人以上の評価が一致した文 (スニペット共起回数)

システムの評価	被験者の評価	3人以上	4人以上
2点	2点	25文	25文
	1点	5文	5文
	0点	10文	5文
	正解率	62.5%	71.4%
1点	2点	22文	20文
	1点	7文	5文
	0点	15文	13文
	正解率	15.9%	13.2%
0点	2点	12文	8文
	1点	8文	2文
	0点	25文	21文
	正解率	55.6%	67.7%

表 2.10 被験者の3人または4人以上の評価が一致した文
(完全一致割合)

システムの評価	被験者の評価	3人以上	4人以上
2点	2点	36文	31文
	1点	6文	4文
	0点	7文	6文
	正解率	73.5%	75.6%
0点	2点	28文	21文
	1点	8文	6文
	0点	7文	6文
	正解率	16.3%	18.2%

表 2.11 被験者の3人または4人以上の評価が一致した文
(活用件数割合)

システムの評価	被験者の評価	3人以上	4人以上
2点	2点	35文	33文
	1点	4文	2文
	0点	7文	7文
	正解率	76.1%	78.6%
0点	2点	29文	24文
	1点	8文	7文
	0点	9文	7文
	正解率	19.6%	22.6%

表 2.3 より、ノイズの除去を行わなかった場合、2点とされた文は 24.4%であった。スニペット共起回数をを用いた場合はシステムが2点と評価した文の正解率は 55.0%、完全一致割合では 62.6%、活用件数割合では 62.0%であった。完全一致割合を用いた場合の正解率が最も高く、ノイズの除去を行わない場合に比べ、38.2 ポイントの精度の向上が見られ、予備実験[8]で行ったスニペット共起回数をを用いた場合よりも、7.6 ポイントの精度の向上が見られた。

完全一致割合を用いた場合にシステムが2点と評価し、被験者4人以上が0点とした文は7文存在した。内訳を以下に述べる。まず、検索クエリとなった単語の前後の単語によって、自然な表現になったり不自然な表現になったりする場合がある。例えば、「前提で書く」という文は、「前の話を読んでいる前提で書くことにする」のような文が Web 中に存在するため、システムは正しいと判断する。このような文が5文存在した。「夏は寒い」は、何か比較対象がある場合に Web 上に存在する可能性のある文である。また、完全一致割合、活用件数割合ともにシステムが0点と評価した文の正解率が極端に低くなっている。これは、どちらも完全一致件数を利用しているため、「ウェブ上には極端に少ないが常識的である知識」をノイズとして判断してしまったことによる。しかし、正しい常識的知識を破棄することよりも、誤った常識的知識を獲得することの方が問題であると考えられる。よって、2点の正解率を重視している。

3 おわりに

本稿では、日本語 WordNet を用いて Open Mind Common Sense に含まれる文の名詞について同位名詞を獲得することで、常識的知識を表現する文を生成し、Web 検索を用いて常識的ではない文の除去を行う方法について述べた。評価実験の結果、完全一致割合を用いた場合が最も正解

率が高く 62.6%で、予備実験と比べると 7.6 ポイントの精度向上が見られた。しかし、これはまだ十分な精度ではなく、今後精度を向上させていく必要がある。

今後の課題をいくつか挙げる。まず、今回は名詞を一つ含む文のみに着目したが、名詞が複数個含まれる文についても同じ処理が有効かを検証する必要がある。動詞や形容詞などの、名詞以外の品詞についても同様である。

また、今回は常識的ではない文をノイズとして除去する方法を述べたが、これも常識的知識を表現する文ととらえることができる。例えば、「夏は寒い」はノイズであるが、システムがこの文を0点と評価することができれば、否定文「夏は寒くない」に変換するだけで常識的知識を含む文となる。同様に、1点の文に対して「主に」や「こともある。」等を付加することで、1点に相当する常識的知識を含む文を生成することも可能である。

また、ノイズ除去の精度を向上させる方法をいくつか挙げる。まず、検索クエリを設定する際に、「名詞+助詞」を検索クエリとしたが、助詞は基となる文の助詞をそのまま使用している。しかし、助詞を変更することにより精度を向上させることが考えられる。例えば「鳥は飛ぶ」という文において、「鳥は」だけでなく、「鳥が」を検索クエリとすることで、より多くの文を収集することができると考えられる。「は」を「に」に変える等、意味が大きく変わる助詞の変更もあるので、それぞれの助詞の関係を考慮する必要がある。さらに、閾値を決定する実験を第一著者のみでなく、多くの被験者で行うことも精度向上には重要である。

参考文献

- [1]大江奈緒子, 渡部広一, 河岡司, "地理に関する常識知識を用いたシステムの構築", 電子情報通信学会技術研究報告, Vol. 104, No. 728, pp. 19-24, 2005.
- [2]秋葉友良, 藤井敦, 伊藤克且, "質問応答における常識的な解の選択と期待効用に基づく回答群の決定", 情報処理学会, NL-163, pp. 131-138, 2004.
- [3]ConceptNet4
<http://conceptnet.media.mit.edu/>
- [4]Havasi, C., Speer, R. & Alonso, J. "ConceptNet3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge." Proceedings of Recent Advances in Natural Languages Processing ,2007.
- [5]Open Mind Common Sense
<http://openmind.media.mit.edu/>
- [6]Liu, H., Lieberman, H. and Selker, T. "A Model of Textual Affect Sensing using Real-World Knowledge", International Conference on Intelligent User Interfaces, pp.125-132,2003.
- [7]Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi and Kyoko Kanzaki "Enhancing the Japanese WordNet" in The 7th Workshop on Asian Language Resources, in conjunction with ACL-IJCNLP2009,2009.
- [8]村本晃一, ジェブカ ラファウ, 荒木健治, "WordNet 及び Web 検索による常識的知識ベースのための文生成手法", ことば工学研究会 35,1-7,2010.