

否定表現「ない」を考慮した文間関連度計算方式

谷裕一朗^{†1} 土屋誠司^{†2} 渡部広一^{†2}

概要: 情報検索など自然言語を扱う様々な場面において否定表現が用いられている。情報検索手法として表記のみではなく、文の意味を捉えた文間の関連性の定量化手法がある。この手法は単語間の関連性に基づいて計算を行っており自立語を用いるため、否定表現「ない」が含まれる文に対して考慮されない問題がある。本稿では否定表現「ない」が含まれた文に対し、言い換え変換や各単語の重み変更を行い、文間の関連性の正確さを評価した。

キーワード: 情報検索, Earth Mover's Distance, 否定表現

Degree of Association between Japanese Sentences in Consideration of one of Negative Expressions.

YUICHIRO TANI^{†1} SEIJI TSUCHIYA^{†2} HIROKAZU WATABE^{†2}

Abstract: Negative expressions are treated in many situations handling with natural languages such as information retrieval. There is not only quantification technique of the relevance between sentences which match with words matching but also grasp the meaning of sentences as an information retrieval method. Because this approach uses substantive terms and calculations based on the relevance between expressions, there is a problem that is not considered for sentences including one of negative expressions. This paper evaluated the accuracy of relevance to sentences with the one of negative expressions by converting words in the sentence or modifying the computations method of it.

Keywords: Information Retrieval, Earth Mover's Distance, Negative expressions

1. はじめに

近年人間の日常生活において、コンピュータが良きパートナーとして機能することが求められている。コンピュータが良きパートナーとなるためには、人間の意図を汲み取る必要があると考え、その中で会話は有効な手段の一つである。人間同士が日常的に行っている会話を人間とコンピュータでも実現することが必要とされる。新聞記事や日常会話などに用いられる自然言語を扱うあらゆる表現にも適宜対応することが必要だと考える。

コンピュータが人間の意図を理解するには、人間が発話された文に対してどのような内容文と関連性があるか数値化することで実現できると考えられる。文同士の関連性を求める手法として EMD (Earth Mover's Distance) を用いた文間関連度計算方式[1]が提案されている。この手法では文中の各単語間の関連性を表現するために構築された概念ベース[2]を用いることによって、文間の意味的な関連度を求めることができる。

EMD を用いた文間関連度計算方式では、各文中にある自立語を用いて文同士の関連性が求められるが、付属語が考慮されずに計算される問題がある。否定表現の例として、「試合に勝たない」など文の末尾に付属語「ない」が付く

と否定文となり、その文の事象を打ち消す意味合いを持つ文の文間関連度の値が考慮されていない。

そこで否定表現を文同士の関連度に考慮することを目的とした文間関連度計算方式を構築する。本稿では、人間とコンピュータの雑談型の会話システムへの応用を想定している。否定表現「ない」を考慮した文間関連度計算方式を構築することにより、ユーザとの会話から得られるユーザの嗜好を理解し、興味のある情報の提供が可能になる。ユーザに適した情報を提供することで、コンピュータがより人間らしいコミュニケーションが取れるのではないかと考えられる。

2. 語概念連想システム

語概念連想システムでは概念ベースやシソーラスなどの知識を利用し、想起語処理と未知語処理[3]を提供している。想起語処理とは与えられた語からその語に対して関連の強い語を想起する処理である。また、未知語処理とは常識判断システムが知識として所持していない語（未知語）を知識として所持している語（既知語）に置き換える処理であり、この未知語処理によってシステムが保持しておく知識を最小限に抑えることができる。我々は、これらの処理を概念ベースと関連度計算方式[4]によって実現している。以

^{†1} 同志社大学大学院理工学研究科
Graduate School of Engineering, Doshisha University

^{†2} 同志社大学理工学部
Faculty of Science and Engineering, Doshisha University

下、概念ベースと関連度計算方式について述べる。

2.1 概念ベース

概念ベースとは、電子化された国語辞書や新聞記事などから機械的に構築された知識ベースである。ある語を概念と定義し、概念の意味特徴を表す語（属性）とその重要さを表す数値（重み）の対の集合によって定義しているある概念Aはm個の属性 a_i と重み w_i (>0) の対によって次のように表現される。

$$\text{概念}A = \{(a_1, w_1), (a_2, w_2), \dots, (a_m, w_m)\} \quad (1)$$

概念ベースの特徴として、属性を成す単語群も概念ベースの中で概念として定義されている点がある。概念Aの意味定義を行う属性 a_i を、概念Aの一次属性と呼ぶ。この属性 a_i を概念とみなして更に属性を導くことができ、概念 a_i から導かれた属性 a_{ij} を、元の概念Aの二次属性と呼ぶ。概念ベースの具体的な例を表1に示す。

表1 概念ベースの例

概念	(属性,重み)
医者	(医師, 0.34), (患者, 0.11), (病院, 0.08), ...
病院	(医院, 0.25), (手術, 0.11), (施設, 0.04), ...
治す	(治療, 0.43), (医療, 0.21), (病気, 0.13), ...

2.2 関連度計算方式

関連度計算方式とは概念ベースにある2つの概念の関連の強さを定量的に表現する手法である。算出された数値を関連度と呼ぶ。関連度は0.0から1.0までの実数値で表現され、関連度が大きいほど概念間の関連が強いといえる。このように概念と概念の関連を定量化することで、コンピュータに語の関連を判断させることが可能となる。関連度計算方式にはお互いの概念が持つ属性の一致度と重みを利用する重み比率付き関連度計算方式を使用する。

2.2.1 一致度

ある概念A, Bにおいて、その属性を a_i, b_j 対応する重みを u_i, v_j それぞれ属性がL個, M個 ($L \leq M$) とすると、概念A, Bはそれぞれ

$$A = \{(a_1, u_1), (a_2, u_2), \dots, (a_L, u_L)\} \quad (2)$$

$$B = \{(b_1, v_1), (b_2, v_2), \dots, (b_M, v_M)\} \quad (3)$$

となる。このとき、概念Aと概念Bの属性一致度 $DoM(A, B)$ を以下のように定義する。

$$DoM(A, B) = \sum_{a_i=b_j} \min(u_i, v_j) \quad (4)$$

$$\min(u_i, v_j) = \begin{cases} u_i & (u_i \leq v_j) \\ v_j & (u_i > v_j) \end{cases} \quad (5)$$

ここで、 $a_i = b_j$ は属性同士が一致した場合を示している。つまり、一致度とは概念Aと概念Bそれぞれの属性を比較して一致した場合に、小さい方の重みを選択して足し合わせた合計値ということになる。これは、小さい方の重みは互いの属性の重みの共通部分となっているので、概念Aと概念Bどちらにも有効な重みだと考えられるためである。こ

のとき各概念の重みの総和が1になるように正規化する。よって、一致度は0.0から1.0の実数値をとる。

2.2.2 重み比率付き関連度計算方式

重み比率付き関連度は次の手順で求める。2.2.1項で述べた概念A, Bにおいて、まず属性数の少ない方の概念Aを基準とし、その属性の並びを固定する。その上で概念Bの属性を概念Aの各属性との一致度の和が最大になるように並び替える。このときの概念Bの属性と重みを (b_{xi}, v_{xi}) として次のように定義する。

$$B = \{(b_{x1}, v_{x1}), (b_{x2}, v_{x2}), \dots, (b_{xM}, v_{xM})\} \quad (6)$$

これらの概念についての重み比率付き関連度 $DoA(A, B)$ を次の式で定義する。

$$DoA(A, B) = \sum_i DoM(a_i, b_{xi}) \times \frac{(u_i + v_{xi})}{2} \times \frac{\min(u_i, v_{xi})}{\max(u_i, v_{xi})} \quad (7)$$

関連度 $DoA(A, B)$ は、属性の一致度に、属性間の重みの比率と平均値を乗じた値となる。

3. EMD(Earth Mover's Distance)

EMD[5]とは分布間の距離を表すもので、最適な輸送コストを用いて定義される。EMDを求める際、2つの分布P, Qは特徴量ベクトル p_i, q_j 、その特徴量に対する重みを w_{pi}, w_{qi} とし特徴量をm個, n個とすると以下の式で表現される。

$$P = \{(p_1, w_{p1}), \dots, (p_m, w_{pm})\} \quad (8)$$

$$Q = \{(q_1, w_{q1}), \dots, (q_n, w_{qn})\} \quad (9)$$

今、 p_i と q_j の距離を d_{ij} とし、全特徴間の距離を $D = [d_{ij}]$ とする。ここで、 p_i から q_j への輸送量を f_{ij} とすると、全輸送量は $F = [f_{ij}]$ となる。ここで、式(10)に示すコスト関数を最小とする輸送量Fを求め、EMDを計算する。

$$WORK(P, Q, F) = \sum_{i=1}^m \sum_{j=1}^n d_{ij} f_{ij} \quad (10)$$

ただし、上記のコスト関数を最小化する際、以下の制約条件式を満たす必要がある。

$$f_{ij} \geq 0, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \quad (11)$$

$$\sum_{j=1}^n f_{ij} \leq w_{pi}, \quad 1 \leq i \leq m \quad (12)$$

$$\sum_{i=1}^m f_{ij} \geq w_{qj}, \quad 1 \leq j \leq n \quad (13)$$

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left[\sum_{j=1}^n w_{qj}, \sum_{i=1}^m w_{pi} \right] \quad (14)$$

ここで、式(11)は輸送量が正であることを表し、 p_i から q_j に送られる一方通行であることを表している。式(12)は輸送元である p_i の重み以上に輸送できないことを表す。式(13)は輸送先である q_j の重み以上に受け入れることができないことを表す。最後に式(14)は総輸送量の上限を表し、それは輸送先または輸送元の総和の小さい方に制限されることを表す。

以上の制約条件の下で求められた最適な輸送量Fを用いて、分布P, Q間のEMDを以下のように表現される。

$$EMD(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n d_{ij} f_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (15)$$

ここで、最適なコスト関数 $WORK(P, Q, F)$ をEMDとしてそのまま用いないのは、コスト関数は輸送元もしくは輸送先の重みの総和に依存するので、正規化することによってその影響を取り除くためである。

3.1 EMDを用いた文間関連度計算方式

文間関連度計算方式とは文と文との関連の強さを定量的に表現する手法である。文間関連度は0.0から1.0の間までの実数で表され、文と文との関連が強いほど大きな値を示す。また、本稿で用いる文間関連度計算方式に利用されるEMDは自立語間の関連性と重みを考慮したM対Nでの柔軟な対応が可能である。EMDの特徴として、需要地と供給地との距離の値が0.0から1.0であるなら、EMDも0.0から1.0の値になる。そして、EMDは二つの分布間の関連性が強いと値が低くなり、弱いと値が高くなる。よって1からEMD値を引いた値を文間関連度と見ることができる。EMDを用いて文間関連度を算出するにはEMDにおける需要地と供給地、需要量と供給量、各需要地と供給地間の距離を定義する必要がある。需要地と供給地は文間関連度を算出したい二つの文それぞれの文中の自立語を割り当てる。需要量と供給量はそれぞれの自立語の $tf \cdot idf$ [6]重み付けを用いる。この重みを用いることによって、文中の自立語がその文にとってどれだけ重要かを表すことが出来る。 $tf \cdot idf$ についての詳細は3.1.2節で述べる。そして、需要地と供給地間の距離は自立語間の関連性と見立てることができ、本稿では概念ベースを用いた一致度計算により求める。ただし一致度は関連性が高いと値も大きくなるため、1から一致度の値を引いた値に変換する。

図1に文1「梅が咲く」と文2「うめ祭りが始まる」の2文のEMDを文に適応した例を示す。なお、括弧内の数値が各自立語の重み、太線上にある数値が使用する一致度を示している。

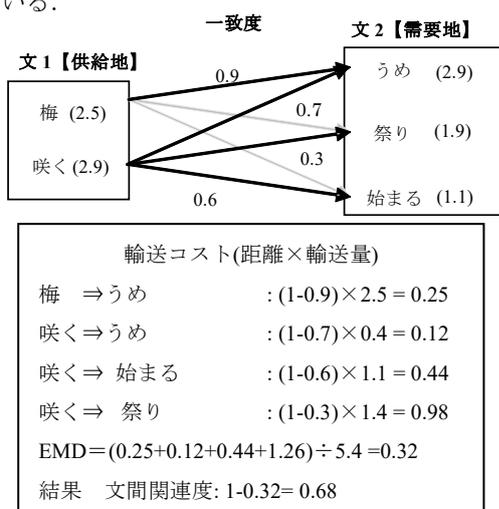


図1 EMDを用いた文間関連度算出例

3.1.1 日本語形態素解析器

日本語形態素解析器[7]とは、入力された文に対して形態素解析を行うシステムである。形態素解析とは、自然言

語で書かれた文を意味持つ最小の単位である形態素の列に分割し、それらの品詞を判別することである。本稿では、代表的な形態素解析器である茶釜を用いた。茶釜の機能例として、文の一部を形態素解析した結果を表2に示す。

表2 形態素解析結果

単語	読み仮名	活用前の単語	品詞
インターネット	インターネット	インターネット	名詞-一般
カフェ	カフェ	カフェ	名詞-一般
の	ノ	の	助詞-連体化
略称	リヤクショウ	略称	名詞-サ変接続
.	.	.	記号-句点

3.1.2 $tf \cdot idf$

自立語に対する重み付けは、情報検索の分野で広く用いられている $tf \cdot idf$ による重み付けを用いる。本稿では、新聞記事一か月分の記事数を N とし、重み付けに使用する記事 d における自立語 t の重み w_t^d は、以下の式(16)で定義する。

$$w_t^d = tf(t, d) \times \log\left(\frac{N}{df(t)} + 1\right) \quad (16)$$

N : 文書数 d : 文書

tf : 文書 d 内における自立語 t の出現頻度

df : 自立語 t の文書頻度

t : 自立語 w_t^d : 重み

4. 文間関連度計算方式の問題点

EMDを用いた文間関連度計算方式では、文中に出現する自立語間の関連性を基に文間関連度を求めている。そのため、否定表現「ない」などが文中にあった場合でも考慮されない。以下例を図2に示す。

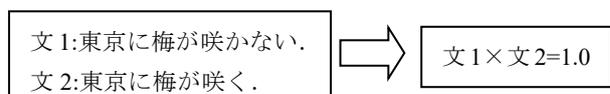


図2 EMDを用いた文間関連度計算方式の例

計算で用いられる自立語がすべて同じであるため、文間関連度の値は1.0が算出される。このように、否定表現「ない」など付属語は文間関連度には考慮されず、文同士の正確な関連度を求めることができない。そこで、否定表現を考慮する文間関連度計算手法が必要であると考えた。

5. 提案手法

5.1 言い換え変換手法

文中に否定表現「ない」を含む文の場合、その文の意味を保持した言い換え表現を反対語辞書から取得、または述語を新たに加える手法を提案する。意味を保持した言い換え変換ならば、否定表現「ない」が文間関連度に考慮でき

るのではないかと考えた。以下、言い換え変換手法を次の手順で行う。

- ①「ない」が含まれる文を入力文に対し、否定表現か否か判断するため知識ベース検索を行う。この知識ベースに存在する文は、非否定表現と判断する。
- ②知識ベース検索に存在しなかった文に対して、日本語係り受け解析器の南瓜[8]を用いて否定表現「ない」が含まれた文節を抽出する。文節には否定表現「ない」が係り受けする述語が含まれており、この単語が言い換え変換対象語となる。
- ③変換対象語に対し反対語辞書を参照し、反対語が存在する単語があれば変換を行う。
- ④変換対象語が反対語辞書に存在しない場合、変換対象語を変換せずに変換対象語に述語の追加を行う。
- ⑤反対語を用いて変換または述語追加を行った後、変換対象語前後の文とを連結させ、文間関連度計算方式に適用する。以下に、提案手法全体の流れを図3に示す。

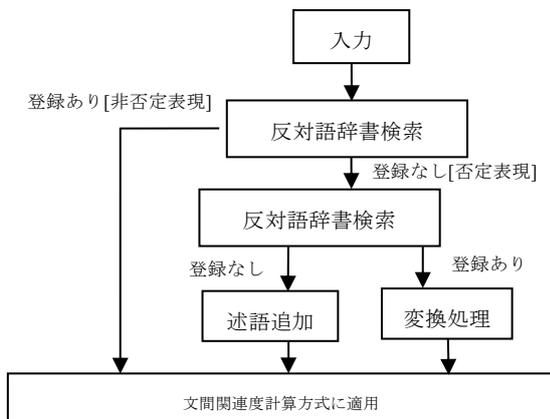


図3 提案手法の流れ

5.1.1 日本語係り受け解析器

日本語係り受け解析器とは係り受け解析を行うシステムである。係り受けとは語の間にある修飾-被修飾の関係である。文の係り受けの関係を句・文節を単位として解析する。本稿では、南瓜を用いた。

5.1.2 知識ベース検索

本稿では、否定表現「ない」を考慮した文を対象としている。しかし、入力文のなかには、否定表現「ない」ではない表現が存在すると考えられる。そのため、この知識ベース検索では、知識ベースには、非否定表現として用いられる慣用句および副詞 41 語[9]を格納されている。入力文が知識ベース内の表現に表記一致した場合、非否定表現と識別する。以下表3に知識ベースの一例を示す。

表3 知識ベース

推量・慣用句	ではないの、ではないかしら、 ではなからうか、かもしれない...
副詞	非常に、とても、いくらか...

5.1.3 反対語辞書検索

反対語辞書は反対語大辞典[10]を参考に、改定常用漢字表[11]の例文に記載されている述語を基に反対語を格納し

ている。この反対語辞書は一対一対応となっており、反対語として合計 362 語が登録されている。反対語辞書の例を図4に示す。

勝つ	⇔	負ける
複雑だ	⇔	簡素だ

図4 反対語辞書の例

この辞書では、「魚を食べない」という入力文があった場合、「食べる」の反対語である「吐く」が取得され、言い換えとしては不適切であると考えられる。このように変換に不適切だと考えられる反対語を目視で削除を行っている。

5.1.4 述語追加

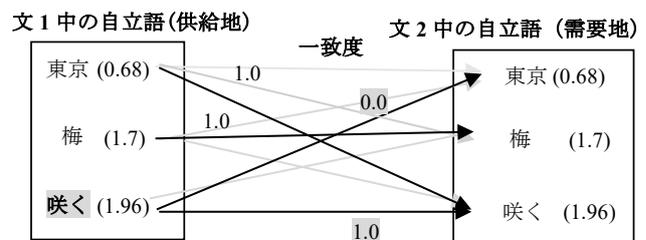
反対語辞書を参照し該当しなかった場合、変換対象語を変換せず、変換対象語に述語を追加を行う。変換対象語が動詞の場合、変換対象語の原型に「ことをやめる」を追加する。変換対象語が形容詞や形容動詞の場合、変態対象語の原型に「とは異なる」を追加する。これらの語句を追加する理由は、否定表現「ない」が用いられている文の内容と比較して意味保持性が保たれていると考えられるからである。図5に述語追加例を示す。

・動詞の場合 動詞の原型+「ことをやめる」 (ex:「魚を食べない」⇒「魚を食べることをやめる」)
・形容詞、形容動詞の場合 形容詞、形容動詞の原型+「とは異なる」 (ex:「その壁は白くない」⇒「その壁は白いとは異なる」)

図5 述語追加の例

5.1.5 距離計算変換手法

EMDを用いた文間関連度計算方式において、輸送コストである距離は1から一致度を引くことで求められる。ここで否定表現「ない」に係る述語と需要地の各自立語の距離に一致度を使用する。一致度を用いることにより、自立語間の距離が逆になり、同じ単語は距離を離すことができる。図4に文1「東京に梅が咲かない」と文2「東京に梅が咲く」の一致度変換後の算出例を図6示す。



輸送コスト (距離×輸送量)		
咲く⇒東京	:(0.0)×0.68	= 0.0
梅 ⇒梅	:(1-0.0)×1.7	= 0.0
咲く⇒咲く	:(1.0)×1.28	= 1.28
東京⇒咲く	:(1-1.0)×0.68	= 0.0
EMD	(0.0+0.0+1.28+0.0)÷4.34 = 0.29	
結果	文間関連度: 1-0.29 = 0.71	

図6 EMDを用いた文間関連度算出例

図6のように文1の「咲く」の距離を一致度とすることで、文間関連度の値が0.71となった。なお、一致度が同じ値の場合、供給地の重みが大い自立語から計算される。

6. 評価手法

アンケートを用いて、評価セットの作成を行う。被験者5名に主語および述語を最低一つ含めた肯定文 A 及び、A中の述語に対する反対の意味を持つ肯定文 B を一組として、1人につき20文自由作成してもらった。以下アンケート合計100組の一例を表4に示す。

	A	B
1	上司に従う	上司に逆らう
2	経済と法律に詳しい	経済と法律に疎い
3	業績が好調だ	業績が不調だ
:	:	:

表4 アンケートの一例

評価セットとして、Aの述語に対して否定表現「ない」を加えた否定文を X を加える。また肯定文 B において、反対の意味を持つ表現がないと被験者が判断した文には、反対語はないものとして述語を削除した文を B に格納する。以上の X をテストセットとした。アンケートの中で述語の重複を除き、合計95組の評価セットを用いて評価を行った。表5に95組中の品詞の割合、表6に評価セットの一例、表7に反対の意味をもたない評価セットの一例を示す。

表5 評価セット中の述語の品詞の割合

動詞	形容詞	形容動詞
31%(29組)	36%(34組)	33%(32組)

表6 評価セットの一例

X	A	B
その日の試合に負けない	その日の試合に負ける	その日の試合に勝つ

表7 反対の意味をもたない評価セットの一例

X	A	B
晩御飯を食べない	晩御飯を食べる	晩御飯を

以上の評価セットより式(17)を評価の正解とする。

$$DoA(X,A) < DoA(X,B) \quad (17)$$

7. 評価結果

以下精度結果を図7に示す。

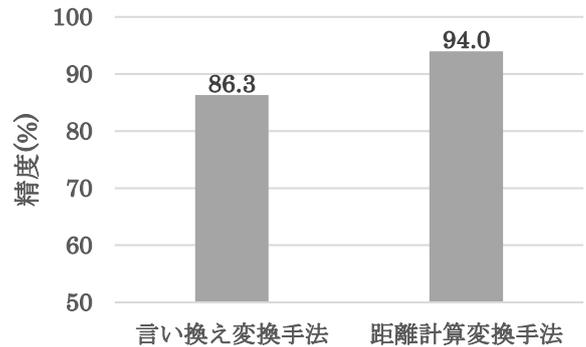


図7 精度結果

8. 考察

8.1 言い換え変換手法の考察

精度結果より、言い換え変換手法の改善点は2つあると考える。一つは、複数の反対語に対応である。単語によって反対語は複数あると考えられ、単文の入力のみでどの単語が最適か判断することは難しい。そのため複文の入力で候補語との関連性を求めることにより、最適な単語を選択できるのではないかと考える。

二つめとして、一致度計算において異なる品詞では、同じ品詞よりも属性が一致しにくいことが挙げられる。表8に不正解の評価セットを示す。

表8 不正解の評価セットの一例

X	A	B
午前中に起きない	午前中に起きる	午前中に眠る

表8中の一致度計算の中で、「眠る」と「午前」はなんらかの関係性があると考えられる。しかしこれらの自立語の共通した属性がないため、一致度は0.0が算出される。このように関係性があると考えられる異なる品詞の自立語間の一致度は、同じ品詞の自立語の一致度と比べ、差が大きくなる場合がある。適切な文間関連度が求められない可能性がある。そこで反対語辞書を用いず、述語追加のみを行った精度を求めた。以下表9に精度結果を示す。

表9 述語追加のみの精度結果

	追加する述語	精度(%)
1	やめる	62.1
2	異なる	62.1
3	違う	61.0
4	終わらず	72.6
5	相違なる	60.0
6	変わる	60.0
7	控える	65.2
8	抑える	67.3
9	止まらず	76.8
10	停める	74.7

表9の結果を用いて、追加する述語によって有意な差があるか一元配置分散分析を行った。10パターンを行った精度を標本とし、帰無仮説を「追加する述語に関わらず精度の平均は等しい」として有意水準0.05で検定を行った。以下表10に結果を示す。

表10 一元配列分散分析の結果

F 値	F 境界値	P 値
1.740215	1.889824	0.075939

表10より、F値<F境界値となるので、統計的優位差は認めなかった。以上の結果より、打消しの1単語による述語追加のみの手法には、単語の選択によって精度の差はないと考えられる。

そのため、各自立語の属性の類義語を取得し、属性の品詞の比率を変えることで精度向上が見込めると考えられる。

8.2 距離計算変換手法の考察

距離計算変換手法では、否定表現「ない」が含まれた供給地の自立語と需要地の各自立語の一致度を距離として用いた。これにより自立語が同じ場合でも表記一致することが無く、距離を離すことができた。DoA(X,B)を求めた結果87組が0.9以上という高い文間関連度の値が算出された。正解した組の計算結果を見ると、距離計算変換手法に当てはめた自立語間において0.0に低い一致度が存在している。そのため、DoA(X,B)において高い関連度の値が算出されたと考えられる。以下、正解した組における一致度例を図8に示す。

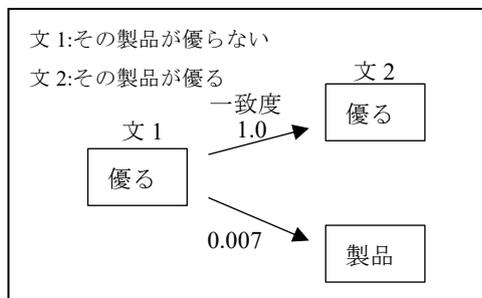


図8 一致度計算例

図8より、8.1節で述べたように一致度の低さに問題があると考えられる。「優る」と「製品」では品詞が異なるため、共通する属性が少なく、一致度が0.007となった。従って、否定表現「ない」が含まれた自立語間の一致度は距離として用いて計算を行うため、表記一致しない自立語間の一致度が低いものが多かったため0.9以上の文間関連度が求められ、正解率が94.0%となった可能性がある。

9. おわりに

本稿では否定表現「ない」を含む単文に対して、否定表現「ない」を考慮した文間関連度計算方式を構築した。言い換え変換手法では、最大86.3%、距離計算変換手法では94.0%の精度を得られ、否定表現「ない」が存在する文とし

ない文との識別ができ、文間関連度として適切な値が求められたと考える。

今後の課題として本稿では否定表現「ない」に限定して研究を行ったため、その他の否定表現も理解できる手法が必要になると考えられる。

特に、否定表現には「食べたいわけではない」、「寒くはない」、「朝に起きるわけではない」など数値化することが困難な曖昧な表現がある。本研究より、すべての言語が言い換え可能とは限らず、否定表現がある場合の文同士の関連度を求めるには、自立語のみならず必要に応じて付属語も用いることでより人間が意図する文の意味を文間関連度として算出でき、文検索など自然言語を処理する様々な場面において有効活用できるのではないかと考えられる。

謝辞 本研究の一部は、JSPS 科研費 16K00311 の助成を受けて行ったものです。

参考文献

- [1]藤江悠五, 渡部広一, 河岡司, “概念ベースと Earth Mover’s Distance を用いた文書検索”, 自然言語処理学会, Vol16, No.3, pp.25-49, 2009.
- [2]井筒大志, 渡部広一, 河岡司, “概念ベースを用いた連想機能実現のための関連度計算方式”, 情報科学技術フォーラム FIT2002, pp159-160, 2002.
- [3]土屋誠司, 小島一秀, 渡部広一, 河岡司, “常識的判断システムにおける未知語処理方式”, 人工知能学会論文誌, Vol. 17, No. 6, pp. 667-675, 2002.
- [4]渡部広一, 奥村紀之, 河岡司, “概念の意味属性と共起情報を用いた関連度計算方式”, 自然言語処理, Vol.13, No.1, pp.53-74, 2006.
- [5]Rubner, Y, Tomasi, C, and Guibas, I: The Earth Mover’s Distance as a Metric for Image Retrieval, International Journal of Computer Vision, 14(3), pp.130-137, 2000.
- [6]G.Salton and C.Buckley. Term-weighting approaches in automatic text retrieval. Information Processing and Management, Vol.41, No.4, pp.513-523, 1998.
- [7]ChaSen -- 形態素解析器, <http://chasen-legacy.sourceforge.jp/>, 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座(松本研究室), 2011.
- [8]松本裕治, “形態素解析システム「南瓜」”, <https://taku910.github.io/cabocha/>, 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座(松本研究室), 2015.
- [9]吉江誠, 目良和也, 市村匠, 山下利之, 相澤輝昭, 吉田勝美, “真偽疑問文における返答発話の肯定/否定意図の解析と Web インターフェースへの応用”, 日本ファジィ学会誌 Vol.14, No.4, pp393-403, 2002.
- [10]中村一男, “反対語大辞典”, 東京堂出版, 1965.
- [11]文化審議会, “改定常用漢字表”, 文化庁, 2010.