

Wikipediaの概念に基づく連想関係テストコレクション 「WikiSimi3000」

“WikiSimi3000”: A Test Collection based on Wikipedia Concepts

伊藤 雅弘*¹ 中山 浩太郎*² 原 隆浩*³ 西尾 章治郎*⁴
Masahiro Ito Kotaro Nakayama Takahiro Hara Shojiro Nishio

*¹*³*⁴大阪大学大学院情報科学研究科

Dept. of Multimedia Eng., Graduate School of Information Science and Technology, Osaka University

*²東京大学知の構造化センター

Center for Knowledge Structuring, The University of Tokyo

Recently, researches on semantic relatedness measurements using Wikipedia have been conducted. However, test collections which are used for experimental evaluations are likely low coverage compared with that of Wikipedia concepts. Moreover, terms of the test collections are not disambiguated to corresponding concepts. In this paper, we introduce the test collection “WikiSimi3000,” which is based on Wikipedia concepts.

1. はじめに

近年, Wikipedia を用いた概念間の関連度計算手法に関する研究が行われてきた。関連度計算手法研究における評価方法は, 主に主観評価, アプリケーションによる評価, テストコレクションによる評価の3つがある。主観評価は [Ollivier 07] のように, 提案する関連度計算手法を用いて抽出した関連性のある項目ペアのリストを複数の被験者に提示し, 関連度を3段階や10段階で主観評価してもらい, 精度を測る方法である。アプリケーションによる評価は [Banerjee 03] のように, 手法を語義曖昧性解消や情報検索などのアプリケーションに適用して, その精度や再現率を評価している。テストコレクションによる評価は [Gabrilovich 07, Ito 08, Strube 06] などの研究のように, “M&C”, “R&G”, 近年では語彙数の多さから “WordSimilarity-353 Test Collection” を用いた評価が行われている。とりわけ, テストコレクションによる評価は, 複数の手法を同じ基準で評価できるため, 手法間の比較を行う上で重要な評価手法である。しかし, これらのテストコレクションは, Wikipedia がカバーする概念に比べて語彙の網羅性が低く, また曖昧性が解消されていないなどの問題がある。本稿では, 筆者らの構築しているテストコレクション「WikiSimi3000」について解説する。

2. 従来のテストコレクションの問題点

WordSimilarity-353 Test Collection は Wikipedia を用いた概念間の関連度計算に関する研究 [Ito 08, Strube 06] において, もっともよく用いられているテストコレクションである。WordSimilarity-353 Test Collection は, 353組の単語を13人~16人の被験者によって関連性を主観で10段階評価し, その平均を関連度としている。評価する際は, 各手法によって353組の単語の関連度を計算し, 各手法によって計算された353個の関連度とテストコレクションの353個の関連度と

の相関を, スピアマンの順位相関係数やピアソンの相関係数などで求めることによって, 手法の精度を測る。

しかしながら, このテストコレクションは, Wikipedia 研究における概念間の関連度計算手法を評価する際のいくつかの問題を抱えている。まず一つ目に, 単語ペアの数が非常に限られているということである。また, その単語も一般語に偏っており, Wikipedia のような専門語や固有名詞も含めた様々な概念を有するコーパスに対しては, 網羅性のある評価という観点から不十分である。そして, さらに大きな問題として, 定義されている単語の曖昧性が解消されておらず, 評価する際に Wikipedia の概念(記事)にマッピングするという作業を行わなければならないことが挙げられる。しかし, このマッピングが正確であるという保証はない上, 研究者によってマッピング方法も異なっているため, 評価精度に影響を与える。また, そもそもテストコレクション作成時に単一概念として被験者に提示されているわけではなく, 意味の曖昧性も含んだ単なる単語として提示され, 関連度を判定している点も問題である。

3. WikiSimi3000

前章で述べた問題を解決するために, 筆者らは, 英語版 Wikipedia の概念を基にしたテストコレクション「WikiSimi3000」を構築している。Wikipedia の概念に基づいたテストコレクションを構築することによって, 概念間の関連度計算手法を評価する際に, 概念をテストコレクションの単語にマッピングする作業を必要とせず, より精度の高い評価が可能となると考えられる。本章では, 「WikiSimi3000」の構築方法と構築結果について述べる。

3.1 構築方法

「WikiSimi3000」の構築は, 主に以下の流れで行った。

1. Wikipedia 記事群のクリーニング

- Forward Link もしくは Backward Link の数が5以下の記事をノイズ記事として除外
- リストページやカテゴリページなどの通常記事以外を除外

2. 残った4000万リンクを含む82万記事から, 人物・地理・文化などの様々な分野の概念から, 各被験者が30程度の

連絡先: 伊藤雅弘, 大阪大学大学院情報科学研究科, 〒565-0871 大阪府吹田市山田丘 1-5 大阪大学大学院情報科学研究科 マルチメディア工学専攻 マルチメディアデータ工学講座(西尾研究室), 06-6879-4513, 06-6879-4514, ito.masahiro@ist.osaka-u.ac.jp

表 1: 評価基準

スコア	基準
9 - 10	強く連想される。 (例: Microsoft -> Windows)
7 - 9	連想される。
5 - 7	ある程度連想される。
3 - 5	連想されるかもしれない。
2 - 3	自信がないが、連想されないと思う。 (例: Microsoft -> Mac OS X)
1 - 2	基本的には連想されない。
0 - 1	絶対連想されない(例: Cabbage -> Microsoft)

表 2: WikiSimi3000 の一例

Concept1	Concept2	Relatedness
Apple Inc.	Steve Jobs	10
Apple Inc.	NeXT	6
Microsoft	Steve Ballmer	10
Microsoft	Solar panels	0
Personal computer	Laptop	9
Personal computer	Steel	0
Television	Broadcasting	10
Television	Greek language	0
Dog	Collie	8
Dog	Houston	1
Book	Magazine	9
Book	Scriptorium	2
Pasta	Noodles	9
Pasta	Food dye	5
Police	Sheriff	9
Police	United States gov.	4

概念 (記事) を選定

(例: Television, Book, Tea, United States, Coca-Cola, Silk Road, Macintosh, Cheese, Bill Clinton, Japan, Olympic Games, Halloween, Bicycle, Elvis Presley)

3. 各被験者が各概念の記事内から、手動で他の概念 (記事) へのリンクを選択
4. 各被験者が各概念ペアの関連度を 0 から 10 で評価

0 から 10 の評価の基準は、被験者に対して表 1 のように提示した。

3.2 構築結果

前節の方法で構築された本テストコレクションは、合計で 3,604 の概念ペアを含む大規模なものとなっている。構築したテストコレクションの例を表 2 に示す。特筆すべき点は、各ペアの関連度評価がそれぞれの概念の記事を確認した上で行われるので、確実に単一概念どうしの関連度評価が行われているということである。また、WikiSimi3000 の利点として、Wikipedia のリンクを利用した概念間の関連度を計算する研究との親和性が高いことが挙げられる。例えば、[Nakayama 07] や [Ito 08] では、取り扱っているすべての概念が Wikipedia に存在する概念 (記事) であるため、評価のためのテストコレクションとして WikiSimi3000 をそのまま用いることができる。

また、現在英語版だけでなく、日本語版 Wikipedia の概念を基にしたテストコレクションも構築中である。日本語版を構築する理由は、Wikipedia を用いた関連度計算手法は、基本的に言語非依存な手法であるという利点があり、そのことを示す

有効な手段は、多言語での評価を行うことと考えたためである。一方、「WordSimilarity-353 Test Collection」は英語のみの対応であるため、多言語テストコレクションを構築することが必要である。現在までで、677 の概念ペアを、それぞれ 3 人の被験者が評価し、延べ 2031 の評価が集まっている。

4. まとめ

本論文では、Wikipedia の概念に基づく連想関係テストコレクション「WikiSimi3000」について述べた。「WikiSimi3000」は、これまでの単語を基にしたテストコレクションと違い、語彙の曖昧性を解消した Wikipedia の概念 (記事) を基にしたテストコレクションである。テストコレクションを構築する際の被験者は、評価対象の語彙がどの概念を表しているかを明確に認識しながら関連度を評価している。このテストコレクションを用いることによって、Wikipedia の概念にマッピングされた語彙同士の関連度をより正確に評価することができる。

今後は、より多くの被験者に評価してもらうことによって、テストコレクションの品質を高めるとともに、Wikipedia の概念を基にした照応関係や上位下位関係を定義したテストコレクションの構築も行っていく予定である。

謝辞

本研究の一部は、特定領域研究 (18049050)、科学研究費補助金基盤研究 (C)(20500093) およびマイクロソフト産学連携研究機構 CORE 連携研究プロジェクトの助成によるものである。ここに記して謝意を表す。

参考文献

- [Banerjee 03] Banerjee, S. and Pedersen, T.: Extended Gloss Overlaps as a Measure of Semantic Relatedness, in *IJCAI*, pp. 805–810 (2003)
- [Gabrilovich 07] Gabrilovich, E. and Markovitch, S.: Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis., in *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 1606–1611 (2007)
- [Ito 08] Ito, M., Nakayama, K., Hara, T., and Nishio, S.: Association Thesaurus Construction Methods based on Link Co-occurrence Analysis for Wikipedia., in *Proceedings of Conference on Information and Knowledge Management*, pp. 817–826 (2008)
- [Nakayama 07] Nakayama, K., Hara, T., and Nishio, S.: Wikipedia Mining for An Association Web Thesaurus Construction., in *Proceedings of IEEE International Conference on Web Information Systems Engineering*, pp. 322–334 (2007)
- [Ollivier 07] Ollivier, Y. and Senellart, P.: Finding Related Pages Using Green Measures: An Illustration with Wikipedia, in *AAAI*, pp. 1427–1433 (2007)
- [Strube 06] Strube, M. and Ponzetto, S. P.: WikiRelate! Computing Semantic Relatedness Using Wikipedia, in *Proceedings of National Conference on Artificial Intelligence and Innovative Applications of Artificial Intelligence Conference*, pp. 1419–1424 (2006)