

表 3 解熱鎮痛薬の組み合わせ候補単語上位 10 位

組み合わせ候補	共起単語	一致度
鎮痛	鎮痛	1.00
鎮痙	鎮痛	0.64
鎮静	鎮痛	0.64
体温効果	体温	0.62
ムスカリン	アスピリン	0.61
体温上昇薬	体温	0.55
中枢興奮	中枢神経系	0.51
中枢抑制	中枢神経系	0.51
抗鬱	抗菌薬	0.48
解熱	発熱	0.47

文抄録を解析し、MeSH¹⁾に準拠して類義語を整理したソースである。本研究では、LSD RDF Data Portal²⁾を利用し、機能発揮主体(機能を持つ概念、薬物や蛋白質等)と共起する MeSH 統制語について、共起頻度上位 25 件を取得した。これにより、機能発揮主体ごとに共起する単語が得られる。表 2 に解熱鎮痛薬との共起語の上位 25 件を示す。これらの共起語には“鎮痛薬”、“発熱”、等の“解熱鎮痛薬”が持つべき“鎮痛”、“解熱”などの作用と部分一致する語が存在する。

(2) 組み合わせ候補の順位付け

2.3.(1)で得られた共起単語を用いて、2.2.の組み合わせ候補を順位付けする。組み合わせ候補のうち、共起単語と近い単語に高スコアを付ける。なぜなら、共起単語には、“解熱鎮痛薬”の作用として適切な概念のラベルと部分一致する語が含まれ、その一致度が高い概念はより、共起元になっている薬物や蛋白質との関連が大きいと考えられるからである。この単語間距離の計算には、java のライブラリである Apache Lucene²⁾を利用し、レーベンシュタイン距離、ジャロウィンクラー距離、Ngram 距離の平均値を単語間距離とした。表 3 に解熱鎮痛薬の組み合わせ候補単語の順位付け結果の上位 10 位を示す。組み合わせ候補の最上位に既存の関係である“鎮痛”が来ており、順位付け

することによって、関係を結ぶべき概念を上位に推定することができたことが分かる。

図 2 に推定手法の全体の流れを示す。例では、蛋白質・ペプチドの一つである“オピオイドペプチド”と function-of 関係を結ぶべき概念を推定している。まず、バイオサイエンスオントロジーから概念階層比較によって蛋白質の機能候補を取得する。次にライフサイエンス辞書の LOD から共起単語を検索・取得する。機能候補の概念群と共起単語群を比較し、単語間距離の近い順に並べることで、各機能発揮主体に対する作用の組み合わせ候補を尤もらしい順に提示することができる。

3. バイオサイエンスオントロジーへの適用例

本章では、具体性を用いて、これまで述べてきた手法を薬物、蛋白質の概念階層に適用した結果について述べる。ただし、これらの例は、専門家による評価を受けたものではなく、筆者がウェブ上の情報を参考に判断したものである。

3.1 組み合わせ候補を適切に順位付けできた例

既存の“Function-of”関係を利用して、概念階層間を比較したとき、組み合わせ候補を抽出することができた概念は、薬物では 35 個、蛋白質では 11 個である。

(1) 薬物: アンギオテンシン変換酵素阻害剤

一つ目の例は、先ほどの“解熱鎮痛薬”と同じく薬物の一つである“アンギオテンシン変換酵素阻害剤”と“function-of”の関係を結ぶべき薬理作用の候補を順位付けした例を紹介する。“アンギオテンシン変換酵素阻害剤”(ACE 阻害剤)は、“解熱鎮痛薬”と同じ段数に存在する兄弟概念であるので、概念階層

表 4 ACE 阻害薬と共起する単語

"血压"@ja
"心筋梗塞"@ja
"アンジオテンシンII"@ja
"死亡率"@ja
"カルシウムチャネル遮断薬"@ja
"アンジオテンシン"@ja
"降圧薬"@ja
"降圧薬"@ja
"プラセボ"@ja
"レニン・アンジオテンシン系"@ja
"エナラプリル"@ja
"腎疾患"@ja
"血管拡張薬"@ja

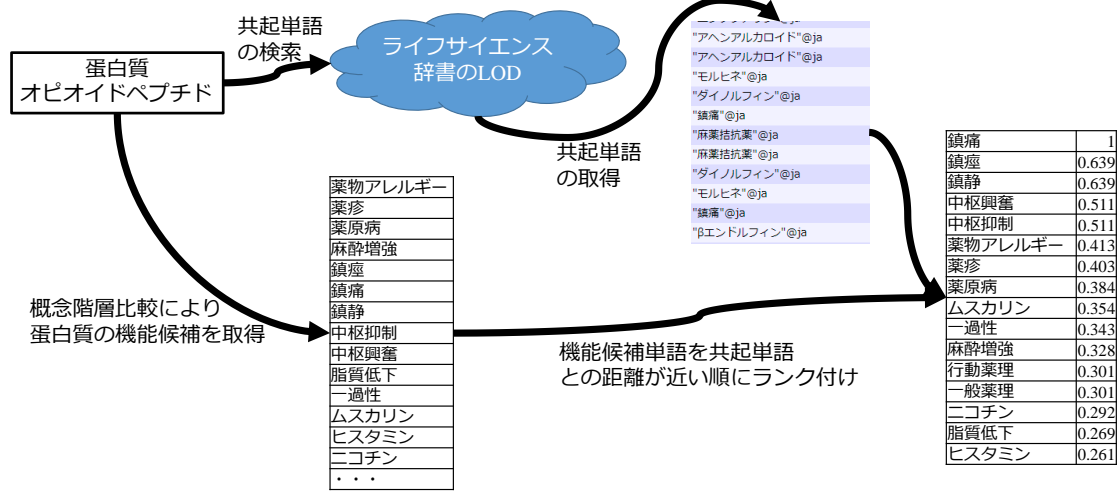


図 2 概念間関係推定の全体の流れ

1 <http://www.nlm.nih.gov/mesh/meshhome.html>
 2 <https://lucene.apache.org/core/>

表 5 ACE 阻害薬と関係を結ぶべきと推定された組み合わせ候補単語

組み合わせ候補	共起単語	一致度
血管拡張	血管拡張薬	0.85
降圧	降圧薬	0.75
利尿	利尿薬	0.75
カルシウム拮抗	カルシウムチャネル遮断薬	0.57
血管収縮	血管拡張薬	0.51
昇圧	血圧	0.47
気管支拡張	血管拡張薬	0.44
抗不安	心不全	0.41
血糖低下	血圧	0.40
血糖上昇	血圧	0.40

比較によって抽出される、組み合わせ候補の薬理作用の集合は、先ほどと同じ 62 の概念である。一方、LOD から得られる共起単語群は、“ACE 阻害剤”との共起となるので異なる集合が得られる。表 4 に共起単語の一部を示す。この薬物は、末梢血管を拡張し、血圧を下げる作用を持つが、共起語の中にも“血圧”、“降圧薬”、“血管拡張”などの語が見られる。

次に、表 5 に組み合わせ候補の単語を共起単語との一致度で順位付けを行ったものを示す。組み合わせ候補の中で、“血管拡張”がトップに来ており、次いで“降圧”となっている。一方“血管収縮”という逆の作用を示す語も上位であるが、この順位付けは、ドメインの専門家による概念間の関係の追加作業時の支援を目的にしているため、この例のように明らかに関係を結ぶべきではない単語であれば、容易に判断可能であると考えられる。

(2) 蛋白質:オピオイドペプチド

次に蛋白質の一つである“オピオイドペプチド”と“Function-of”の関係性を結ぶべき作用の候補の順位付けをした例を紹介する。“オピオイドペプチド”は、先ほど紹介した薬物とは違う階層を参照しているので、72 の作用からなる異なる概念の集合が、概念階層比較によって組み合わせ候補として抽出される。“アンギオテンシン変換酵素阻害剤”の時と同様に、共起単語を取得すると、表 6 のような単語が得られる。“オピオイドペプチド”は、体内麻薬の一種であるため、“アルカロイド”や“モルヒネ”のような薬物、“鎮痛”などの機能が共起単語として得られる。

この共起単語との一致度で“オピオイドペプチド”の組み合わせ候補の 72 の概念を順位付けすると表 7 のようになる。上位に

“鎮痛”、“鎮痙”、“鎮静”、“中枢興奮”、“中枢抑制”が来ており、“オピオイドペプチド”の作用として適切なものが来ていることが確認できた。

表 6 オピオイドペプチド

と共起する単語

"アヘンアルカロイド"@ja

"アヘンアルカロイド"@ja

"モルヒネ"@ja

"ダイノルフィン"@ja

"鎮痛"@ja

"麻薬拮抗薬"@ja

"麻薬拮抗薬"@ja

"ダイノルフィン"@ja

"モルヒネ"@ja

"鎮痛"@ja

"βエンドルフィン"@ja

3.2 LSD にエントリが無く検索できない例

今回推定に利用した蛋白質 11 種と薬物 37 種のうち、薬物の 16 種では、LSD にエントリが存在しないため検索することができず、順位付けができなかった。具体例としては、“ペプチド系抗生物

表 7 オピオイドペプチドと関係性を結ぶべきと推定された組み合わせ候補単語

組み合わせ候補	共起単語	一致度
鎮痛	鎮痛	1.00
鎮痙	鎮痛	0.64
鎮静	鎮痛	0.64
中枢興奮	中枢神経系	0.51
中枢抑制	中枢神経系	0.51
薬物アレルギー	薬物治療	0.41
薬疹	薬物治療	0.40
薬原病	薬物治療	0.38
ムスカリン	エンケファリン	0.35
一過性	痛覚過敏	0.34

質”、“抗腫瘍抗生物質”などの抗生物質の下位概念と“硝酸化阻害剤”、“土壌殺菌剤”などの農業殺菌剤の下位概念である。農業殺菌剤は、ライフサイエンス論文を集めた PubMed では扱っていないと考えられる。今後、分野によって検索する LOD を適宜変更する必要があると考えられる。

4. まとめ

本論文では、構築途中段階のオントロジーに対して、追加すべき概念間関係の候補をオントロジー構築者に提示する手法を開発した。まず、兄弟概念の比較による組み合わせ候補の抽出手法について述べた。その結果、薬物・蛋白質の両方で取りうる組み合わせの候補を限定することができた。この概念階層同士の比較による組み合わせ候補の抽出は、オントロジー化の作業を実際に進めている JST のドメインの専門家からも有用であると評価された。さらに、これらの組み合わせ候補からより結ぶべき関係をわかりやすく提示するために順位付けを行った。ライフサイエンス関係の論文での共起情報を利用して組み合わせ候補に順位付けを行った結果、蛋白質・薬物ともに関係を結ぶべき機能を上位に推定することが可能となった。ただし、これらの結果は、専門家による評価を受けたものではなく、筆者がウェブ上の情報を参考に判断したものであるため、今後、ライフサイエンス分野の専門家に蛋白質・薬物とその機能の正解の組み合わせデータを作成してもらい、今回の推定結果と比較することで評価することを計画している。

謝辞

本研究の一部は、科学研究費補助金・基盤研究(B) 25280081 および基盤研究(A)26240033 の支援による。

参考文献

- [櫛田 17] Kushida, T. et al, Efficient construction of a new ontology for life sciences by sub-classifying related terms in the Japan Science and Technology Agency thesaurus, ICBO2017, Newcastle, (2017).
- [増田 17] 増田壮志, 古崎晃司:概念階層の比較に基づくオントロジー品質向上支援手法, 人工知能学会論文誌, Vol. 32 No. 2, (2017).
- [Masuda 17] Masuda T., Kozaki K., Kushida T., Tateisi Y., Watanabe K., Matsumura K., Kawamura T., and Komatani K., Extending A Bioscience Ontology Based on Comparison between Sibling Concepts, 8th International Conference on

-
- Internet Technologies & Society, Sydney, Australia, Dec. 11-13, 10pages, (2017).
- [Noy 01] Noy, N. F., Sintek, M., Decker, S., Crub´ezy, M., Ferguson, R. W., and Musen, M. A.: Creating semantic web contents with Prot´eG´e-2000, *IEEE Intelligent Systems*, Vol. 16, No. 2, pp. 60–71, (2001)
- Osborne, F., Motta, E. 2015, Klink-2: Integrating Multiple Web Sources to Generate Semantic Topic Networks, *The Semantic Web-ISWC 2015(Part1)*, Lecture Notes in Computer Science, Springer International Publishing, Cham, Bethledem, Pennsylvania, US, pp. 408–424.
- [Poveda-Villalon 12] Poveda-Villalon, M., Suarez-Figueroa, M. C., and Gomez-Perez, A.: Validating Ontologies with OOPS!, *Knowledge Engineering and Knowledge Management Lecture Notes in Computer Science*, Vol. 7603, pp. 267–281 (2012)
- [Sirin 07] Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., and Katz, Y.: Pellet: A practical owl-dl reasoner, *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 5, No. 2, pp. 51–53 (2007)