

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 9 日現在

機関番号：15201

研究種目：若手研究(B)

研究期間：2013～2014

課題番号：25730195

研究課題名(和文) オープンイノベーション環境での企業と大学研究者のマッチングシステムの開発

研究課題名(英文) Development of a Matching System for Companies and Researchers in Open Innovation

## 研究代表者

山田 泰寛 (Yamada, Yasuhiro)

島根大学・総合理工学研究科(研究院)・助教

研究者番号：50529609

交付決定額(研究期間全体)：(直接経費) 1,300,000円

研究成果の概要(和文)：本研究は、オープンイノベーションの初期段階において、共同研究に繋がる可能性がある企業や研究者の発見のために、特許や論文のような科学技術文書から、企業と研究者の繋がりを発見する手法の開発を目的とする。文書集合において、1文書内で共起する単語の組に着目する。この単語の組が同じ2文書で共起し、どちらの単語も他のどの文書にも出現しない時、その単語の組を特徴語として抽出する。この特徴語を頂点とし、2つの特徴語の共起を辺とするグラフを出力する検索システムを作成した。辺が結ぶ2つの特徴語が出現する2文書の研究者を、その辺のラベルとして付与する。この研究者が2つの特徴語において繋がっていることを示す。

研究成果の概要(英文)：In the early stage of open innovation, it is important to find suitable partners in order to form successful collaborations between companies and university researchers. We proposed a system that finds the partners by using information retrieval and text mining techniques from the documents such as scientific papers and patents. Given a query, the proposed system outputs a graph of unique research in retrieved documents. Each vertex in the graph corresponds to a word with a document frequency of two. Two words connected by an edge occur in the same two documents, and neither word appears in other retrieved documents. The edge is labeled with the names of the researchers involved in the documents in which the two words appear.

研究分野：テキストマイニング

キーワード：テキストマイニング 情報検索 オープンイノベーション 産学連携 知的財産情報

### 1. 研究開始当初の背景

企業が、自社の技術のみで新商品や新サービスの開発を行うのではなく、他社や大学と共同で、自社の技術と他社や大学の持つ技術を組み合わせることにより、新商品や新サービスの開発を行うオープンイノベーション<sup>引用文献①</sup>が注目されている。企業にとっては、自社だけでは実現できない新商品の開発の幅が広がる可能性がある。

また、基礎研究を担う大学への期待も大きくなっている。大学は、産学連携を進めることで、大学が企業の新商品の開発に貢献したり、大学自らが開発した商品によって収入を得ることが求められている。

しかし、数ある企業や大学、研究機関の中から、自社や大学の研究開発に直接繋がる可能性がある連携効果の高い企業や大学研究者を探し出すことは難しい問題である。

### 2. 研究の目的

オープンイノベーションや産学連携の初期段階における問題点は、大学、研究機関、企業の中から、自社もしくは大学研究者と共同研究が可能である相手先を見つけることである。大学に所属する専門のコーディネータに依頼する場合は、コーディネータの持つ情報の量や知識、経験に依存する問題がある。既存の論文検索システム及び特許検索システムを用いることにより、共同研究のために必要な技術を持つ研究者を探す方法もある。しかし、必要な技術が書かれた論文や特許を検索することは時間のかかる作業であり、また、検索された論文や特許を読むことにも多くの時間が必要である。

本申請研究は、特許明細書や論文のような科学技術文書を対象として、情報検索技術やテキストマイニング技術を用いることにより、共同研究の可能性がある大学研究者と企業のマッチングを発見することを目的とする。最終的にはコーディネータの判断や仲介が必要となるが、共同研究先を探す初期段階でマッチングを機械的に支援するツールがあれば、コーディネータの作業を補助することができる。

### 3. 研究の方法

特許明細書や論文のような科学技術文書を対象として、情報検索技術やテキストマイニング技術を用いることにより、大学研究者と企業のマッチングを発見する検索システムの開発を行う。

開発した検索システムは、初めに、文書集合全体から、入力として与えられた単語の出現する文書集合を検索する。次に、検索結果としての文書集合から特徴語を抽出する。その特徴語に関連する企業や研究者を発見する

ことにより、企業と研究者の繋がりを可視化する。

### 4. 研究成果

開発した検索システムは、対象とする文書集合として、1文書に研究者の名前と研究内容に対する複数のキーワードが書かれている文書を想定している。検索システムは、初めに、検索結果としての文書集合の中から特徴語の組の集合を抽出する。次に、システムは各組の2つの特徴語が出現する文書の研究者を抽出する。

マッチングの相手を見つける時に重要なことは、共同開発にふさわしい相手を発見することと、なぜその相手がふさわしいのかという理由が分かることである。開発した検索システムでは、研究者を表示するだけでなく、特徴語も表示することにより、どの観点で研究者が繋がるのかをユーザに提示することができる。

次に、抽出する特徴語について、文書集合の中で頻度の大きい単語や、テキストマイニングでよく用いられる tf-idf という指標を用いた場合、抽出される特徴語は一般的な単語や文書集合全体を表現する単語であり、検索者に対して新しい気付きを与えるものではないことがある。そのような特徴語によって企業と研究者の繋がりを発見することも有用であるかもしれないが、申請研究では、頻度の小さい単語に着目することで、検索者が気付かないような特異な特徴語を抽出する。

開発した検索システムは、検索結果としての文書集合の中で、1文書内で共起するキーワードの組に着目する。このキーワードの組が同じ2文書で共起し、どちらのキーワードも他のどの文書にも出現しない時、そのキーワードの組を特徴語として抽出する。抽出したキーワードの組は同じ2文書でのみ出現するため、この2つのキーワードによって繋がる研究者は2文書の研究者のみである。また、このキーワードによって、他の研究者とは繋がらない。よって、この2文書の研究者は、抽出されたキーワードによって共同研究の可能性があると考えた。

このように、「この2文書にしか出現しない」キーワードの組に着目することにより、一般的ではなく、特異な特徴語の抽出を行う。また、一つのキーワードでは、偶然に同じ2文書に出現していることもあるため、キーワードの組を考えた。

検索システムは、ユーザが単語を入力として与えた時、その単語の出現する文書集合から、上記の条件を満たすキーワードを頂点とし、2つのキーワードの共起を辺として表すグラフを表示する。辺には、その辺によって繋が

れる2つのキーワードが出現する2文書の研究者がラベルとして付与される。この2文書の研究者が、頂点のキーワードに関して、マッチングの可能性があると考える。検索システムは、インデックス作成や検索などについて汎用連想検索エンジン GETA<sup>引用文献②</sup>を利用している。

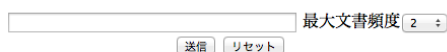


図1 検索システムの入力画面

図1は、検索システムの入力画面である。検索したい単語を入力し、送信ボタンを押す。上記の説明では、特徴語となるキーワードの文書頻度を2文書に限定しているが、右側のプルダウンメニューで文書頻度を変更することが出来る。

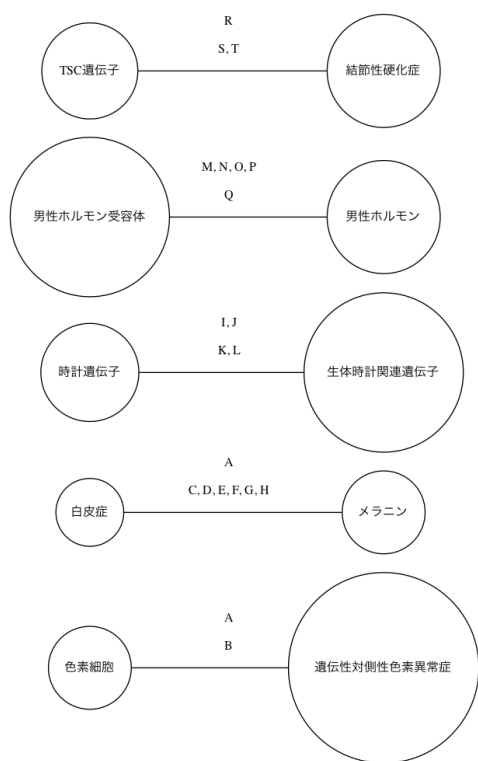


図2 「遺伝子変異」の出力結果

図2は、科学研究費助成事業データベースを対象として、「遺伝子変異」について検索を行い、検索結果の文書集合に対して、上記の条件を満たすキーワードの組と、そのキーワードが出現する文書の研究者を表示したものである。頂点としてのキーワードが、検索結果の中で文書頻度が2文書のキーワードであり、ノードを繋ぐ辺に記載されている研究者が、その2つのキーワードで繋がれる研究者である。2文書それぞれの研究者を2行で表示している。研究者は記号に置き換えてい

る。

図2では、10個のキーワード、5組のキーワードの組が抽出され、20人の研究者が表示されている。例えば、「TSC 遺伝子」と「結節性硬化症」というキーワードによって繋がれる研究者は、Rという研究者と、SとTという研究グループであることを示す。また、「白皮症」と「メラニン」というキーワードによって繋がれる研究者は、Aという研究者と、CからHまでの研究者のグループであることを示す。

図3は、「東南アジア」の出力結果である。検索結果の文書集合の中で、20個のキーワード、10組のキーワードの組が抽出され、62人の研究者が表示されている。例えば、「海賊」と「漁民」というキーワードによって繋がれる研究者は、CIとCJという研究者であることを示す。また、「南インド」と「インド化」というキーワードによって繋がれる研究者は、BXからCCまでの研究グループと、CDからCHまでの研究グループであることを示す。

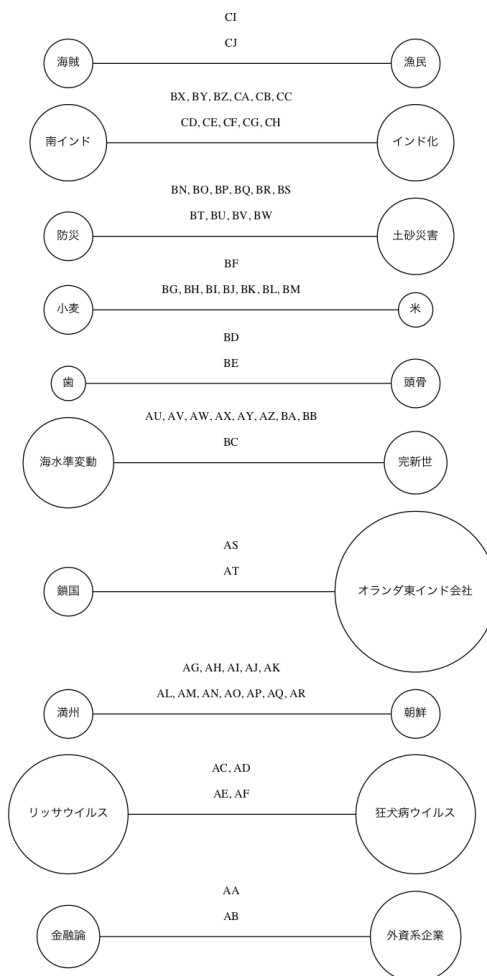


図3 「東南アジア」の出力結果

図4は「エピジュネティクス」の出力結果である。検索結果の文書集合の中で、6個のキーワード、7組のキーワードの組が抽出され4人の研究者が表示されている。これまでのグラフとは異なり、研究者AとBは、「糸状菌」、「酵素阻害剤」、「天然物」、「二次代謝物」の4つのキーワードによって繋がっていることが分かる。

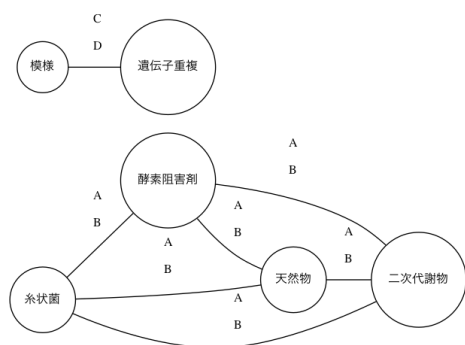


図4 「エピジュネティクス」の出力結果

実験により、特徴語として抽出するキーワードの総数は、全体のキーワード総数と比較して少量であった。これは、文書中に共起する2つのキーワードが同じ2つの文書にしか出現しないという条件が厳しすぎたためであった。抽出する特徴語の条件を緩和し、特異性を表す特徴語の定義について改良が必要であることが分かった。

システムによって発見したマッチングが共同開発可能なマッチングであるか、実際的な観点からの評価が必要であるが、今後の展望として、特許明細書や論文など大量の科学技術文書を検索の対象に含めることで、より多くの技術や企業、研究者を対象としたシステムが実現できると考える。

#### 〈引用文献〉

- ① H. Chesbrough, Open Innovation: The New Imperative for Creating and Profiting from Technology, Harvard Business School Press, 2003.
- ② 高野明彦, 西岡真吾, 今一修, 岩山真, 丹羽芳樹, 久光 徹, 藤尾正和, 徳永健伸, 奥村学, 望月源, 野本忠司, 汎用連想計算エンジンの開発と大規模文書分析への応用.  
<http://geta.ex.nii.ac.jp/pdf/itx2002.pdf>

#### 5. 主な発表論文等

##### 〔雑誌論文〕(計3件)

- ① Yasuhiro Yamada, Daisuke Ikeda and Sachio Hirokawa, Unique Links as Weak

Ties, Proceedings of the 4th International Congress on Advanced Applied Informatics, 2015. (査読有)(to appear)

- ② 山田泰寛, 丹生晃隆, 八塚一成, 廣川佐千男, 特許情報を活用した産学連携支援ツールの提案, 産学連携学会第12回大会講演予稿集, pp. 82-83, 2014. (査読無)
- ③ Yasuhiro Yamada, Terutaka Tansho and Sachio Hirokawa, Proposal of a Matching System for Companies and Researchers Using Patents and Scientific Papers, Proceedings of the 2013 Second IIAI International Conference on Advanced Applied Informatics, pp. 397-398, 2013. DOI 10.1109/IIAI-AAI.2013.55 (査読有)

##### 〔学会発表〕(計3件)

- ① Yasuhiro Yamada, Unique Links as Weak Ties, 4th International Congress on Advanced Applied Informatics, 2015年7月14日, 岡山コンベンションセンター(岡山県岡山市).
- ② 山田泰寛, 特許情報を活用した産学連携支援ツールの提案, 産学連携学会第12回大会, 2014年6月25日, 下諏訪総合文化センター(長野県下諏訪町).
- ③ Yasuhiro Yamada, Proposal of a Matching System for Companies and Researchers Using Patents and Scientific Papers, 2013 Second IIAI International Conference on Advanced Applied Informatics, 2013年9月3日, くにびきメッセ(島根県松江市).

#### 6. 研究組織

##### (1) 研究代表者

山田 泰寛 (YAMADA, Yasuhiro)  
島根大学・大学院総合理工学研究科・助教  
研究者番号: 50529609