

氏名(本籍)	いし かわ だい すけ 石川大介(茨城県)
学位の種類	博士(情報学)
学位記番号	博甲第5151号
学位授与年月日	平成21年3月25日
学位授与の要件	学位規則第4条第1項該当
審査研究科	図書館情報メディア研究科
学位論文題目	特許文献からの因果関係抽出とそれを用いた類推による仮説生成の研究

主査	筑波大学教授	理学博士	石塚英弘
副査	筑波大学教授	工学博士	田中和世
副査	筑波大学教授	理学修士	小野寺夏生
副査	筑波大学教授	博士(工学)	佐藤哲司
副査	鶴見大学教授	農学博士	長塚隆

## 論文の内容の要旨

第1章は序論で、その概略は次のとおりである。本研究の目的は、コンピュータを用いて特許文献から因果関係を抽出し、得られた因果関係を用いてコンピュータ上で類推によって仮説を生成することである。本研究で対象とする因果関係は自然科学分野の因果関係であり、特許文献に書かれた手段と効果の関係は自然科学分野の因果関係に対応する。また、類推を用いて生成する仮説は自然科学分野の研究に関する仮説を対象とする。著者は関連文献を調査した結果、科学者が発想の際に類推を頻繁に用い、また、仮説を生成する際にも類推を用いていること、コンピュータによる自動処理によって学術文献または特許文献からの自然科学の因果関係の抽出に成功した文献が見当たらず、また類推を用いて自然科学上の仮説を生成した文献も見当たらないことを述べている。

第2章は因果関係抽出の研究の章である。著者は、JAPIOの特許出願抄録の要約に類出する文型『【手段を記述した動詞句】「ことにより」【効果を記述した動詞句】』に着目して因果関係を抽出する方法を考案し、医薬分野と繊維工学分野を対象として、それぞれ抽出実験を行い、その方法の有用性を明らかにした。ここで、【手段を記述した動詞句】の例としては「特定のペプチドを含有する」等々が、また【効果を記述した動詞句】の例としては「有効な創傷治癒促進作用を発現する」等々が挙げられている。

著者は、【手段を記述した動詞句】から「手段を表すキーワード」を抽出する手順を次のとおり示している。

- 1) 形態素解析を行い、名詞、未知語、記号、接頭語を抽出し、複合語処理を行う。
- 2) 得られた結果から不要語を除去した残りを「手段を表すキーワード」として抽出する。不要語は対象分野により異なるため、著者は当該分野の特性を調査し、不要語を調べて除去のための辞書に登録している。また、【効果を記述した動詞句】からの抽出手順も上述の1)と2)は基本的に同じである。著者は実験結果として、医薬分野と繊維工学分野ではラクトフェリン、ペプチド等々の「物質を表すキーワード」が「手段を表すキーワード」になり、医薬分野の場合は「創傷治癒促進作用」、「感染防除効果」等々が、繊維工学分野の場合は「湿潤性」、「浸透性」等々が「効果を表すキーワード」になると述べている。考察では、実験結果を検討し、関連研究とも比較して本方法の有用性を示している。

第3章は類推の研究の章である。著者は、既存の類推の手法を2章で述べた方法によって得られた因果関係に適用して類推を行い、元の因果関係とは異なる因果関係が生成できることを実験（以下、類推による因果関係生成実験）によって明らかにした。また、その因果関係の検証実験を学術雑誌論文あるいは特許を用いて行った結果、類推によって得られた因果関係の中に自然科学上の仮説が含まれていることを明らかにした。

著者は類推の手法について次のように述べている。既存の類推の手法とは、目標と基底を設定し、基底の一部を類似するものに置き換えても目標を実現する可能性があるとして、それを類推結果として提示する手法である。本研究では、ある因果関係を基底として選択し、手段と効果のうち一方を目標として選択し、他方を類似するもので置き換えることによって類推を行い、因果関係を得る。得られた因果関係は仮説になり得る。

著者が行った実験のうち、医薬分野から抽出した因果関係を基底としてライフサイエンス分野を対象とする類推による因果関係生成実験の概要は次のとおりである。基底の因果関係（例えば、手段：ラクトフェリン、効果：感染防除効果）の効果を既存のライフサイエンス辞書を用いて類似の用語（英語表記）に置き換えることによって、目標（例えば、lactoferrinに関する別の因果関係）を得る。著者は得られた因果関係（例：lactoferrinにはadenovirus infectionを防ぐ効果がある）を学術雑誌論文によって検証している。なお、この例の場合は検証の結果、有効な仮説であることが実証された。

著者は実験検証結果に基づいて考察し、本方法の有用性を明らかにし、限界を論じた。

第4章は本論文の結論であり、当初の目的を達成したことが記述されている。

## 審査の結果の要旨

本研究の目的は、コンピュータを用いて特許文献から因果関係を抽出し、得られた因果関係を用いてコンピュータ上で仮説を生成することであると著者が述べている。従って、本論文が、特許文献から因果関係を抽出する研究の部分と、得られた因果関係を用いる類推によって仮説を生成する研究の部分とから構成されていることは理解できる。そのため、この2つの部分についてそれぞれ批評する。

### 特許文献からの因果関係抽出

著者が考案したテキストから因果関係を抽出する研究の特徴は、1) 抽出源を学術文献ではなく、特許文献としたこと、2) 特許文献の中でも因果関係を単文で明確に表現している要約部分としたこと、3) 考案した方法が未知語に強く、比較的単純であるため、短時間で抽出処理が可能なこと、の3点であり、これらが既存の研究に比較して著者の研究が成功した理由と言えよう。

学術文献には因果関係は記述されていないこともあり、在っても他の内容の中に埋め込まれているから自然言語処理による抽出は容易でないと考えられる。その点、特許文献は手段と効果を記述するものであり、著者の目的に適合している。しかし、特許の明細は手段と効果に分けて詳細に書いてあるため、手段と効果の照応関係把握を含む抽出は容易でない。従って、要約を抽出源に選んだ著者の着眼点が優れていたと言える。

専門領域のテキストを対象として形態素処理を行う場合の問題点は、専門用語の殆どが未知語になるが、これらの未知語を形態素処理の辞書に登録するには多くの手間が掛かることである。著者は、専門用語の辞書登録をせず、未知語もそれ自身が「手段を表すキーワード」「効果を表すキーワード」の候補、あるいは複合語処理により「手段を表すキーワード」「効果を表すキーワード」の候補になると考える実践的なアプローチを採用している。「ラクトフェリン」等々の「物質を表すキーワード」が「手段を表すキーワード」として抽出され、「創傷治癒促進作用」、「感染防除効果」等々が「効果を表すキーワード」として抽出され

た結果を見る限り、このアプローチは妥当と言えよう。

#### 因果関係を用いた類推による仮説生成

ここでは、1) 適切な類似の用語が使用されたか、2) 生成された因果関係は仮説と言えるか、3) 大量の仮説が生成されてしまうのではないか、の3点について検討した。

##### 1) 適切な類似の用語の使用について

類推の成否は適切な類似のものを用意できるか否かに掛かっている。本研究の場合は、類似の専門用語を用意することになるが、それには困難がある。例えば、当該分野の学術用語集のデータが考えられるが、それは定評のある用語に厳選されているため、類似する用語の数が少なく、類推には適していない。著者はライフサイエンス辞書を使用した。これは専門家集団である京都大学薬学部で構築されていること、欧米の有名雑誌論文から専門用語を収集していることから本研究には適した素材と言える。

##### 2) 生成された因果関係はライフサイエンス上の仮説と言えるかについて

次のとおり問題は無いと考えられる。仮説は、ある現象を合理的に説明するため仮に立てる説であり、実験・観察などによる検証を通じて事実と合致すれば定説となる。演繹、帰納の場合は得られた推論結果に論理的合理性が有るが、類推の場合は論理学上の合理性は保障されたため、著者は学術雑誌論文によって検証している。例えば「手段：lactoferrin、効果：viral infection 防除効果」と「手段：lactoferrin、効果：adenovirus infection 防除効果」の場合は、何れも *Antiviral Research* 誌の2002年の論文“Antiadenovirus activity of milk proteins: lactoferrin prevents viral infection”に該当する記述がある。基底は1997年度の特許から抽出された因果関係であるから、類推によって生成された因果関係は仮説であり、2002年の時点で定説になったと考えられる。また「手段：delta sleep inducing peptide、効果：wound healing 促進作用」の場合は、*BMC Genomics* 誌の2007年の論文に該当する記述があり、基底は1997年度の特許から得ていることから、仮説であり、2007年の時点で定説になったと考えられる。従って、著者が考案した方法は仮説の生成が十分可能であり、その点で有用性があると言える。

##### 3) 大量の仮説が生成される可能性とそれへの対応について

本論文によれば、数の予測値が出力可能であり、その結果を見て利用者が対応を考えるとのことである。それ以上の点については今後の課題と言えよう。

なお、第4章結論の中の、本研究の成果に基づく今後の展望については、対象専門領域の研究者のための研究支援システムや、研究企画者のための支援システムへの展開、因果関係を取り入れた情報検索システムへの展開、等々が考えられること、また、この種のシステムへの展開に際しては、当該分野の専門家の知識と経験に基づく研究方針などをインタラクティブに取り入れる機能の導入も必要になることから、今後の展開には対象領域の専門家との協同研究や協同作業が必要にならうとの指摘が複数の委員からあった。

以上、本研究は当初の目的を達成するとともに、新規性が高く、有用性が認められる。従って学位論文審査委員全員により論文審査合格と判定された。

よって、著者は博士（情報学）の学位を受けるに十分な資格を有するものと認める。