

# 博士学位論文審査要旨

2020年7月11日

論文題目： 連想メカニズムを活用した語彙の意味推定法に関する研究

学位申請者： 後藤 和人

審査委員：

主査： 理工学研究科 教授 渡部 広一

副査： 理工学研究科 教授 下原 勝憲

副査： 理工学研究科 教授 土屋 誠司

要 旨：

本研究の目的は、機械が人間と自然な会話を行うために重要な能力である言葉を理解する方法を実現することであり、基準となる言語資源に登録されていない単語（未定義語）の意味を推定する手法の提案である。

本論文は全7章で構成される。第1章は序論であり、第2章では未定義語の意味推定を行う上で重要となる、ある単語（概念）から様々な概念を連想できる連想メカニズムについて解説している。連想メカニズムは語の概念化方法（単語に属性と重みの集合を与える方法）、意味的関連性評価方法、および、シソーラスから構成されている。本章では、語の概念化方法として、電子化国語辞書などから構築した概念ベース、インターネット百科事典である Wikipedia から構築した文書データベース、インターネット上の言語情報を活用するオートフィードバック（AF）を用いる方法について述べている。第3章では文書データベースを用いて固有名詞の意味推定を行う方法を提案している。提案方法では、文書データベースを用いて固有名詞の概念化を行い、概念化した固有名詞をシソーラスのノードに分類することで意味推定を実現している。第4章では文書データベースを用いて英字略語の意味推定を行う方法を提案している。提案方法では文書データベースを用いて英字略語の概念化を行い、Wikipedia を活用して英字略語の多義性を解消することで意味推定を実現している。第5章では AF を用いて固有名詞の意味推定を行う方法を提案しており、第6章では AF を用いて英字略語の意味推定を行う方法を提案している。最後に第7章にて本研究を総括している。

以上より本論文では、単語の概念化手法としてインターネット上の言語情報を活用する AF、および、インターネット百科事典である Wikipedia から構築した文書データベースを用いる手法を提案し、固有名詞などの概念ベースに存在しない単語に対して関連度計算を行うことができ、シソーラスノードも概念化することで、未定義語とノードとの関連度を計算し、意味推定が可能となった。また、Wikipedia の曖昧さ回避情報を概念化することなどで、英字略語の多義性解消と意味推定も可能となり、提案した未定義語の意味推定手法の有効性を示した。よって、本論文は、博士（工学）（同志社大学）の学位論文として十分な価値を有するものと認められる。

## 総合試験結果の要旨

2020年7月11日

論文題目： 連想メカニズムを活用した語彙の意味推定法に関する研究

学位申請者： 後藤 和人

審査委員：

主査： 理工学研究科 教授 渡部 広一

副査： 理工学研究科 教授 下原 勝憲

副査： 理工学研究科 教授 土屋 誠司

要 旨：

本論文の提出者は、本学大学院工学研究科知識工学専攻博士課程（前期課程）を2008年3月に修了し、2015年4月に同理工学研究科情報工学専攻博士課程（後期課程）に入学し、現在在籍中である。

本論文の主たる内容は、Lecture Notes in Artificial Intelligence 4692, pp.631-638、自然言語処理, 15巻3号, pp.91-113、International Journal of Future Computer and Communication, Vol.6, No.2, pp.53-57、および、自然言語処理, 24巻3号, pp.351-369に掲載され、すでに十分な評価を得ている。

2020年7月11日午後4時より約2時間にわたり提出論文に関する博士論文公聴会が開かれ、種々の質疑応答が行われたが、提出者の説明により十分な理解が得られた。さらに公聴会終了後審査委員により、論文に関する諸問題につき口頭試問を実施した結果、本人の十分な学力を確認できた。なお、語学に関しては英語による論文発表のほか、本学理工学研究科博士課程（後期課程）在籍中に語学試験に合格しており、十分な語学力を有しているものと認められる。以上より、本論文提出者の専門分野に関する学力並びに語学力は十分であることが確認できた。

よって、総合試験の結果は合格であると認める。

# 博士學位論文要旨

論文題目： 連想メカニズムを活用した語彙の意味推定法に関する研究  
氏名： 後藤 和人

## 要 旨：

近年、情報処理技術の進展に伴い、コンピュータをはじめ様々な機械の高度化・知的化が著しい。これらの発展に共通する未来像の一つは「人間と共存する機械」であると言える。人間と共存する機械の一例としては人工知能ロボットが挙げられる。人工知能ロボットは搭載されたコンピュータがセンサから獲得した情報を処理してアクチュエータを動作させることで、ロボットの身体を駆動する構成になっている。人工知能ロボットは知覚、運動、思考、記憶、学習といった機能から構成されている。上記機能のうち、知覚と運動に関する機能については、身体能力に長けたロボットが数多く開発されてきたことにより、その一部が実現されつつある。しかし、ロボットが人間と共存するためには、当該機能に加え、人間と自然な会話ができる機能が必要である。人間は会話をする際に、意識的または無意識のうちに、様々な常識的な概念（場所、感覚、知覚、感情など）を会話文章から判断し、適切な応答を行っている。自然な会話を実現するためには、思考、記憶、学習に関する機能に基づき、ロボットが人間と同じように、常識的に物事を理解し、応答できることが必要である。

ロボットとの自然な会話を実現するにあたり、重要となる能力の一つが言語を理解する能力といえる。ロボットに言語を理解する機能を与える主要な技術分野が自然言語処理である。自然言語処理の分野において、語（単語）の意味を理解するために、単語の意味を何らかの形で定義した様々な言語資源が構築されている。例えば、単語を意味的に分類したシソーラス、概念（ノード）を関係（リンク）で結ぶ意味ネットワーク、文書中の語群を出現頻度からベクトル化するベクトル空間モデル、if-then 形式の記述で条件文による推論の知識を記述するプロダクション規則などが存在する。しかし、会話文に上述の言語資源に含まれていない単語（未定義語）が含まれている場合、当該会話文を理解することは困難である。そのため、未定義語が持つ意味を取得できる仕組みが必要となる。

これまでに、未定義語の意味を取得することに関連する方法として、言語資源を活用した多くの研究が行われている。例えば、ISAMAP やコーパスなどの言語資源に存在し、かつ、シソーラスや概念ベースに存在しない単語に対して、単語の共起頻度などを利用することで、当該単語をシソーラスや概念ベースに分類・登録する方法が提案されている。しかし、これらの方法は言語資源に登録されていない未定義語を対象とする場合、共起頻度を獲得することができないため、対応できないという問題を抱えている。

また、未定義語を言語資源に分類・登録するわけではないが、単語（語義）の曖昧性を解消する方法が研究されている。例えば、ハイパーリンクと品詞情報を利用することで単語の曖昧性を解消する方法や、ニューラルネットワークを構築して文書内の単語を知識ベースのエントリにマッピングすることで曖昧性を解消する技術（Entity Disambiguation）が提案されている。しかし、これらの方法は有効に機能する単語（品詞）が限定されることや、対象領域に応じたニューラルネットワークのトレーニングやメンテナンスが要求されることが問題である。

以上の問題に対処するためには、多くの単語（未定義語）の意味を取得できる方法に加え、当該方法が未定義語の品詞や対象領域に限定されず汎用的に適用できることが必要である。このような未定義語の意味取得方法を実現することで、人間と共存する機械に言語を理解する能力を与え、会話文の意味を理解させることを支援できるようになると考えられる。

本研究の目的は、人間と共存する機械に与える仕組みの一環として、人間と自然な会話を行うために重要な能力である言語を理解する方法を実現することである。より具体的には、基準となる言語資源に登録されていない単語の意味を提示するシステムを実現することを目指す。本研究では、基準となる言語資源に登録されている単語と登録されていない単語（未定義語）を分けて扱う。これは、単語の意味を推定・取得する際にいくつかの言語資源を活用することになるが、対象となる単語の特徴に応じた言語資源を適用することで、より正確に意味の推定・取得が可能になると考えられるためである。具体的には、本研究では、基準となる言語資源として人間が持つような常識を機械に与えるために構築された概念ベースを使用し、概念ベースに登録されていない単語を未定義語と定義する。概念ベースを使用することで、概念ベースに登録されている単語に対して意味を付与することが可能である。一方、辞書などから構築された概念ベースに登録されていない未定義語については、辞書に掲載されていないような単語に関する情報を入手できる Web 上の情報を活用することで意味推定を行う。なお、基準となる言語資源として概念ベースを使用する理由は、人間が言語を理解するために活用している連想という能力を実現することを目的として構築されているためである。

本論文では、未定義語の意味推定を 2 段階で行う。1 段階目は人間であれば表記から意味を特定できる未定義語に対して意味推定を行う。1 段階目の意味推定では、未定義語の代表的な存在である固有名詞をターゲットとする。また、システムへの入力単語（未定義語のみ）とし、単語の関係性を定義するシソーラスを活用することで意味の推定・取得を行う。2 段階目はより難解な対象として、多義性を有する未定義語に対して意味推定を行う。2 段階目の意味推定では、多義性を持つ語として代表的な存在である英字略語をターゲットとする。システムへの入力は英字略語を含む文章とし、世界で最も収録語数が多いとされ、かつ、語の曖昧さを回避するための情報を持つ Wikipedia を活用することで意味の推定・取得を行う。上述の能力を機械に持たせることができれば、様々な未定義語の意味を推定・取得することができるようになり、人間と自然な会話を行うために必要となる言語を理解する能力の実現につながると考えている。

未定義語の意味を推定・取得するためには、ある単語から概念を想起し、さらに、その概念に関係のある様々な概念を連想できる能力が重要な役割を果たす。そこで、すでに提案されている連想能力を表現するメカニズムである連想メカニズムを活用して、未定義語の意味を推定・取得する方法を提案し、その有効性を評価する。

第 1 章では、序論として、人間と共存する機械を実現するために、連想メカニズムを活用した語彙の意味推定法に関する研究に至った背景、目的、および、各章の要旨について述べる。

第 2 章では、未定義語の意味推定を行う上で重要となる、ある単語（概念）から様々な概念を連想できる連想メカニズムについて解説する。連想メカニズムは語の概念化方法（単語に属性と重みの集合を与える方法）、意味的関連性評価方法、および、シソーラスから構成されている。本章では、語の概念化方法として、電子化国語辞書などから構築した概念ベース、インターネット百科事典である Wikipedia から構築した文書データベース、インターネット上の言語情報を活用するオートフィードバックを用いる方法について述べる。また、意味的関連性評価方法として、概念と概念の関連の強さを定量的に評価できる関連度計算と Earth Mover's distance を説明する。さらに、単語を意味的に分類した分類体系であるシソーラスについても述べる。

第 3 章では、文書データベースを用いて固有名詞の意味推定を行う方法を提案する。提案方法では、文書データベースを用いて固有名詞の概念化を行い、概念化した固有名詞をシソーラスのノードに分類（当該固有名詞と最も関連が高いノードを算出）することで意味推定を実現する。

第 4 章では、文書データベースを用いて英字略語の意味推定を行う方法を提案する。提案方法では、文書データベースを用いて英字略語の概念化を行い、Wikipedia を活用して英字略語の多義性を解消することで意味推定を実現する。

第5章では、オートフィードバックを用いて固有名詞の意味推定を行う方法を提案する。提案方法では、オートフィードバックを用いて固有名詞の概念化を行い、概念化した固有名詞をシソーラスのノードに分類することで意味推定を実現する。

第6章では、オートフィードバックを用いて英字略語の意味推定を行う方法を提案する。提案方法では、オートフィードバックを用いて英字略語の概念化を行い、Wikipedia を活用して英字略語の多義性を解消することで意味推定を実現する。

第7章では、論文全体の要点をまとめ、今後の展望について述べる。

本論文によって、未定義語として代表的な存在である固有名詞、および、多義性を持つために意味の推定が困難な英字略語に対して、それぞれの意味を推定する方法を提案し、連想に基づく意味推定の有効性を示すことができた。今後、高度化・知的化されたコンピュータなど様々な機械が人間と共存するようになる状況が到来すると考えられるが、本研究で提案した連想メカニズムを活用した語彙の意味推定法は、人間と共存する機械に与える仕組みの一環として、人間と自然な会話を行うために重要な能力である言語の理解に向けて重要な役割を果たすものである。