

博士学位論文審査要旨

2014年2月17日

論文題目： 語概念連想システムの構築とその応用

学位申請者： 芋野 美紗子

審査委員：

主 査： 同志社大学大学院理工学研究科

教授 渡部 広一

副 査： 同志社大学大学院理工学研究科

教授 下原 勝憲

副 査： 同志社大学大学院理工学研究科

准教授 土屋 誠司

要 旨：

情報処理技術において自然言語を扱うテーマは数多く存在しており、効率的な Web 検索、文書要約、更には人と円滑なコミュニケーションを行うロボットの知的会話など様々であるが、処理対象がどのような自然言語表現であれ、その意味を理解することは必要不可欠である。既存の自然言語処理においても様々な手法での「意味定義」が行われているが、その多くは単語の語彙的な意味や文書中に出現する語群の集合体で定義された処理対象に依存する意味が大半である。また、意味定義において語や事物、概念間の関係性が明確に決められているものが多く、人間らしい曖昧な意味の理解や飛躍的な連想を表現することは難しい。

本論文は全9章で構成される。第1章は序論であり、第2章では本論文で提案する語概念連想システムの全体構成について述べている。第3、4章では語概念連想システムの核となる概念ベースの構築・精錬手法について述べている。第5章では語概念連想システムの拡張として、複数語概念連想システムについて述べている。第6章から8章は語概念連想システムの応用例であり、それぞれ、新聞記事見出し文の意味具体化手法、新聞記事中の難解な語句を会話に用いることのできる程度の平易な表現に変換する手法、および、会話処理において人間の発話からそれに適した自律的な応答を自動生成する手法について述べており、語概念連想システムの有効性を示している。最後に第9章にて本研究を総括している。

以上より本論文では、「連想できる」という曖昧な関係性を定義した知識ベースを構築することにより、人間らしい連想を機械上を実現するための語概念連想システムを提案し、それを用いた複数の応用例を示すことにより、その有効性を確認しており、工学的にきわめて価値が高いものと評価できる。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

2014年2月17日

論文題目： 語概念連想システムの構築とその応用

学位申請者： 芋野 美紗子

審査委員：

主査：	同志社大学大学院理工学研究科	教授	渡部 広一
副査：	同志社大学大学院理工学研究科	教授	下原 勝憲
副査：	同志社大学大学院理工学研究科	准教授	土屋 誠司

要 旨：

本論文の提出者は、本学大学院工学研究科情報工学専攻博士課程（前期課程）を2011年3月に修了し、同年4月に同博士課程（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、自然言語処理, Vol.20, No.2, pp.105-132、Proc. of ICAI'13, Vol.I, pp.131-135、Journal of Communication and Computer, Vol.10, No.4, pp.468-473、Proc. of KES2013, pp.336-344、および、人工知能学会論文誌, Vol.28, No.2, pp.100-111に掲載され、すでに十分な評価を得ている。

2014年1月11日午後3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することができた。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： 語概念連想システムの構築とその応用
氏 名： 芋野 美紗子

要 旨：

情報処理技術において自然言語を扱うテーマは数多く存在しており、効率的な Web 検索、文書要約、更には人と円滑なコミュニケーションを行うロボットの知的会話など様々である。これらの情報処理技術において扱われる自然言語の情報は 1 単語の場合もあれば、膨大な文書集合、口語の会話文など多様であるが、処理対象がどのような自然言語表現であれ、その意味を理解するアプローチは必要不可欠である。

語や文といった自然言語表現の持つ「意味」とは、その語や文を見聞きした際に人間が理解できる内容のことだと考える。ある 1 つの語が指し示す事物・事象の内容や、ある文書に自然言語で記述されている内容といったものを、人間は「意味」として理解することが出来る。

この意味の理解を情報処理技術において行うためには、自然言語によって表現される処理対象の意味を何らかの形で機械上に定義しなければならない。例えば語を意味によって分類することで上位・下位関係や同義・類義関係を記述するシソーラス、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、既存の自然言語処理において語や文、概念の意味を理解するために様々な手法での「意味定義」が行われている。

提案されてきた多くの手法で定義される意味とは、例えば単語の語彙的な意味、つまり国語辞書の語義文により示される意味であったり、文書中に出現する語群の集合体で定義された処理対象に依存する意味が大半である。しかし人間が自然言語表現を見聞きしたときに理解する意味とは、単語の語彙的な意味や、文中に出現する語の一致度合のみで判断されるわけではない。また今まで提案されてきた意味定義においては、語や事物、概念間の関係性が明確に決められているものが多く、例えばシソーラスでは語と語の間に上位・下位や同義・類義といった関係性を定義する必要がある。しかし人間がある語と語に関連があると感じるとき、そこに必ず明確な関係性の定義があるとは言い切れない。「蝸牛」と「貝」はシソーラスにおいて上位・下位関係に分類されるが、「蝸牛」と「紫陽花」をこのような体系で定義することは至極困難である。しかし人間にとっては「蝸牛」と「紫陽花」という語の間に何かしらの関連を見出すことは容易い。

語彙的な意味のみに依存するわけではなく、また分類・体系付けられた構造でも表現しきれない、ある種の「飛躍的な発想」の元で理解する自然言語の意味や関連を私達人間は無意識に扱っている。これは人間が持つ連想能力によって支えられている。連想とは、自身の知っている事柄、情報、概念といった多種多様の「知識」から他の知識を思い浮かべることができることを言う。人間が「紫陽花と蝸牛」の間に関連を見いだせるのは、これらの語が互いを連想させるからである。

人間が「紫陽花」から「蝸牛」、もしくはその逆を連想できるのは、総括的な視点でそれぞれの語が表す内容を知識として保持しているためだと考える。人間が「紫陽花」という語の指し示す内容、つまり「紫陽花の意味」として持っている知識は語彙的なものだけではなく、自身が存在

する現実世界において「紫陽花」とはどのような事物・事象であるか、という概念的なものだと考える。そこにはもちろん語彙的な意味も含まれるが、更に抽象的、例えば「紫陽花に蝸牛という組み合わせが一般的に共感される」という事実関係も「紫陽花」の意味として存在している。様々な事物・事象を概念的な意味で捉えることにより、単なる語彙的な意味の近さだけではない関連の有無を見出すことができる。

情報処理技術における事物・事象の概念化を行うための思想として注目されているのがオントロジーである。オントロジーでは対象となる事物・事象の存在を説明するために必要十分なカテゴリーは何かを考察し、それを書き記すことで対象物が実世界でどのように存在しているかを表現するための枠組みを構築する。しかしこのオントロジーの思考を用いても、前述した、人間が「紫陽花」という語に対して持つ概念的な意味を表現することは困難であると考えられる。オントロジーは概念化する事物・事象の存在を説明するために必要十分なカテゴリーを如何に選び、それらをどのような構造に組み合わせるかが鍵となるが、語彙的な意味にも分類・体系付けられた構造にも依存しない、「連想できる」という曖昧な関係性を記述する最適なカテゴリーを創造することは非常に難しい。

そこで本研究では、「連想できる」という曖昧な関係性を定義した知識ベースを構築することにより、人間らしい連想を機械上を実現するための語概念連想システムの提案および、それを用いた応用技術について述べる。現実世界に存在する事物・事象を表す自然言語の「語」を、「連想できる」という関係性を持っているという事実のみに着目して概念化し、その概念の知識集合である「概念ベース」を用いて人間のような連想と意味の理解を機械で表現することを目指す。また、自然言語を対象とした情報処理技術における応用事例について述べ、語概念連想システムの有用性を示す。

第2章では、連想を表現する機構である語概念連想システムについて述べる。提案する語概念連想システムとは、ある語から連想される他の語を想起する処理および、語と語の間の関連を定量的に表現することで意味の近さを測る関連度計算方式により構成される。このシステムにある語を入力すると、その語から人間が連想できるだろう他の語を出力する。また、2つの語を入力することで、語間の関連を定量的に表現した関連度という値を出力する。

第3章では概念ベースの構築手法について述べる。語概念連想システムの基礎となる概念ベースでは、自然言語で表現される様々な「語」に対して人間が持つ意味、つまりその「語」が示すものが現実世界においてどのような事物・事象であるかを、「語」から連想できる他の「語」の集合を属性として付与することで概念化する。概念と属性の間にある関係性自体を表すラベルやカテゴリーは存在せず、これはオントロジーにおける概念化のための構造定義とはまったく違うアプローチである。国語辞書の見出し語および語義文から作成される基本概念ベースの構築手法、更に属性選別のためのルールによる精練、新聞記事を用いた概念および属性の拡張、シソーラスを用いた属性追加手法について述べる。

第4章では前章までで述べた概念ベースに対して更に属性を追加し、概念の意味定義を拡張する手法を提案する。新たな属性の追加手法として、概念ベースの連鎖構造を利用した二次属性からの属性追加およびWeb上の情報から属性を取得する手法を述べる。二次属性およびWeb上から属性の候補となる語を取得した上で、各種の閾値設定により適切な属性を選択・追加することで概念ベースの精度向上を行う。

第5章では語概念連想システムの拡張として、複数語概念連想システムについて述べる。これは語概念連想システムにおける、ある語から連想できる他の語を想起する処理を複数語に対応させたシステムである。複数の概念を並列に見たとき、それらから想起されるべき概念を取得するシステムについて述べる。

第6章ではロボットとの知的会話を視野に入れた新聞記事見出し文の意味具体化手法について述べる。人間らしい会話をロボットが行うためには、あいさつのような慣用的な表現だけでなく、質問や返答、提案、更には何かしらの話題についての雑談など様々な種類の発話が必要であると考え、そのうち時事情報を話題とした会話について着目をした。時事情報を容易に取得できる新聞の見出し文を知的会話のリソースとして扱うために、見出し文を特有の書式から会話に適した表現へと変換した上で、記事本文による意味の具体化を行う手法について述べる。

第7章では前章と同じく新聞記事を会話のリソースとして扱うことを考え、新聞記事中の難解な語句を、会話に用いることのできる程度の平易な表現に変換する。難解語を変換する際には、他の簡単な一語に変換する一語変換と、一語では説明できない難解語を文によって変換するN語変換を併用することで人間にとって自然に感じる語句変換を実現している。一語変換の際には、難解語と変換を行う一語との意味的な近さを関連度により判断し、候補から適切な語を選択している。また、N語変換においては難解語のもつ意味文と変換に用いる文との関連性を、概念ベースに定義される語の意味知識を利用して算出することで適切な変換を行うことを目指した。

第8章では従来のタスク型の会話システムや膨大な対話例のコーパスに依存した応答生成とは違い、語概念連想システムによる語の連想機能およびそれを基盤として構築される常識判断システムを活用することで、人間の発話からそれに適した自律的な応答を生成する手法について述べる。提案手法では人間の発話中の場所に関する情報を起点として、場所での行動を連想して応答する「場所連想」、場所に存在する人や物への一般的な感覚による共感を応答する「形容詞連想」、人間の発話内容から連想できる他の話題を応答する「話題転換連想」の三つの処理を行い、人間が常識的と感じる応答を生成した。

本論文によって、確率に基づいた言語処理や語間の明確な関係性の定義構築といった既存の自然言語処理技術に加えて、語概念連想システムが提唱する「人間らしい曖昧性を表現する処理」が今後の自然言語処理における重要な要素と成りえる事を示せたと考える。