

学 位 論 文 題 名

Commonsensical and Emotional Knowledge Retrieval from the Internet Resources

(インターネット上の資源を用いた常識及び感情的知識の抽出)

学位論文内容の要旨

著者は、社会的に多くの人が実現を望んでいる「人間らしい」ヒューマンインターフェースを実現するために、ウェブマイニングおよび統計情報を用いるモデルを提案した。

著者は、近年の社会の高齢化に伴う高齢者のコンピュータの利用者数増大を考慮し、ユーザとのより自然なコミュニケーション手段の開発を行なった。このためには、コンピュータが不足している常識的な知識（コモンセンス）の獲得及び感情的な処理ができることが必須と考えその実現に向けて研究を行なった。人間の場合まず感情（本能など）に基づいてコモンセンスを学習し、ある時点で学習されたコモンセンスが感情を抑えるようになるというプロセスに興味を持ち、そのような行動をコンピュータ上に実現するための基礎的な研究を行い、その結果を本学位論文で述べている。

30年程前から人工知能のボトルネックとも言われているコモンセンス実現の問題は種々のアプローチが試みられたが、結局必要とするデータ量の膨大さのため解決されなかったと考えられる。しかし、現在コンピュータ技術の進歩により、マシンパワーが急激に増大し、さらに今まで存在していなかった膨大な情報源であるインターネットが存在している。この有利な二点を利用し、著者は統計的なアプローチとコネクショニズムの考え方を融合し、コモンセンスと感情処理の新たなアプローチを発展させるために研究を進めてきた。本研究の最終目的は「人間らしい」対話システムであり、本プロジェクトを「GENTA」(GENeral belief reTrieving Agent)と呼んでいる。本プロジェクトの最初の段階は常識および感情的な知識がインターネットのリソースから得られるかことを確認するという試みであった。

それらに関するシステム概要及びアルゴリズムについては第2章で述べられている。従来手法では論理的な処理が主流である談話分析を(Discourse Analysis)を論理学に基づかず、インターネット上の共起頻度によって学習するシステムを提案した。感情情報を数学的に表すために Positiveness 及び情報の一般性を表す Usualness という2つの指標を導入した。Positiveness は「好き」「大好き」「嫌い」「大嫌い」のような個人の好みを表すキーワードを利用し、インターネット上でどの意見が多いというウェブマイニングから得られる数値である。例えば「ビール」の Positiveness が高いが、「冷たいビール」のほうが好まれることを数秒で発見することができる。一方、「暖かいビール」をシステムに入力

すれば、Positiveness が極めて小さくなり、インターネットユーザの好み が明確に抽出される。このようなメカニズムにより、あらかじめ膨大なデータをシステムに入力することなく、複雑な処理を行なうアルゴリズムも必要とせず、一般的な知識が獲得された。インターネットユーザの大半による意見が誤りであったとしても、それは本手法で「人間らしさ」を実現したと考えられ、人間の性格をモデリングすることによって重要な情報となる。この考え方は今までコンピュータ上での「人間らしさ」という観点からは存在していない。システムにランダムではなく、インターネットユーザの一般的な意見によって決まる「予想のできない要素」を与えることができる。本手法によりプログラマーによって予め与えられた選択の結果を利用するのではなく、インターネットの利用により社会一般が与えた選択を利用することにより「誰かに作られたものは知能ではなく、個人の知能のコピーというだけである」という人工知能の一番大きい弱点とも言われる問題が解決される可能性が示された。もう1つの指標である「Usualness」のアイデアは、Shannon の情報量の独特な解釈を源としている。普段キーワードなどの抽出に利用される情報量を発言のそれぞれの部分の面白さを測る方法に利用している。第一の実験の目的はその2つのファクターが談話をどのように変化させるかをシステムが学習することであった。帰納的学習を行ないながら、二人の話者の対話を監視し、次の発話の Usualness と Positiveness を推測する実験は3割程度の推測正解率であったが、これは一般化のために自然言語と数字が混在した出力文は評価したためであった。GENTA システムが対話を行なうためにはより深い解析が必要であることが明らかとなった。

従来の対話システムは常識が不足しているために「当たり前の知識」の分析が非常に困難である。しかし、アルキメデスが主張していたように、コモンセンスとカテゴリー化は知能への道であると考え、コモンセンスを抽出しながらカテゴリー化を実現する Bacterium Lingualis という手法を提案した。本手法はヨーグルト菌のような働きをモデル化したものである。本手法で実現される機能は WWW が頭脳と仮定すると、ばらばらに収集された知識をまとめて整理をするものである。他の手法と異なり、日本語の助詞と共起頻度（何と何との共起頻度ですか？）によってカテゴリーが作成されるという考え方である。例えば、インターネット上で札幌とニューヨークは「に・で・まで」という助詞とセットでもっとも高頻繁で現れるので、NiDeMade という大きいカテゴリーに分類され、そのなかでこのセットに頻繁に付属する動詞の頻度がサブカテゴリーを決定する。また、Bacterium はカテゴリー化以外に Positiveness がどんな状況で変わるのかという知識を抽出することができる。例えば、「雰囲気」という単語が感情的にニュートラル（P の値が3）であって、「試合の前の、ぴりぴりした」雰囲気や「静かなバーの落ち着いた」雰囲気の場合などでは、Positiveness レベルが変化することがある。Bacterium はそのそれぞれの状況を抽出することができる。このことを第3章において実験とその結果として述べ、第4章では Bacterium Lingualis 手法のロボット用の命令分析などの工学的なアプリケーションへの導入方法について述べる。Positiveness を測定しながら、利用者の発言から状況を判断し、状況が好ましくなければ、Positiveness を変化させる方法とそれに必要な手段の情報を自動的にインターネットから抽出するアルゴリズムの提案及び本手法の有効性を確認する実験とその結果について述べたものである。

学位論文審査の要旨

主 査 教 授 荒 木 健 治

副 査 教 授 青 木 由 直

副 査 教 授 北 島 秀 夫

学 位 論 文 題 名

Commonsensical and Emotional Knowledge Retrieval from the Internet Resources

(インターネット上の資源を用いた常識及び感情的知識の抽出)

著者は、多くの人が実現を望んでいる「人間らしい」ヒューマンインターフェースを実現するために、ウェブマイニングおよび統計情報を用いたモデルの提案を行った。

著者は、近年の社会の高齢化に伴う高齢者のコンピュータの利用者数増大を考慮し、ユーザとのより自然なコミュニケーション手段の開発を行なった。このためには、コンピュータに不足している常識的な知識（コモンセンス）の獲得及び感情的な処理を行うことが必須のことであると考え、その実現に向けて研究を行なった。著者は人間の場合、まず感情（本能など）に基づいてコモンセンスを学習し、ある時点で学習されたコモンセンスが感情を抑制するようになるというプロセスに興味を持ち、そのような行動をコンピュータ上を実現するための基礎的な研究を行い、その結果を本学位論文で述べている。

30年程前から人工知能のボトルネックとも言われているコモンセンス実現の問題はこれまで種々のアプローチが試みられたが、結局必要とするデータ量の膨大さのため解決されることはなかった。しかし、近年、コンピュータ技術の進歩により、マシンパワーが急激に増大し、さらに今まで存在していなかった膨大な情報源であるインターネットが存在している。このような状況を最大限生かすために、著者は統計的なアプローチに基づいた、コモンセンスと感情処理の新たなアプローチを発展させるために研究を行った。本研究の最終目的は「人間らしい」対話システムであり、本プロジェクトを「GENTA」(GENeral belief reTrieving Agent)と呼んでいる。本プロジェクトの最初の段階は常識および感情的な知識がインターネット上のリソースから得られるかことを確認することである。

それらに関するシステム概要及びアルゴリズムについては第2章で述べられている。従来手法では論理的な処理が主流である談話分析を(Discourse Analysis)を論理学に基づかず、インターネット上の共起頻度によって学習するシステムの提案を行った。感情情報を数学的に表現するためにPositiveness及び情報の一般性を表すUsualnessという2つの指標の導入を行った。Positivenessは「好き」「大好き」「嫌い」「大嫌い」のような個人の好みを表すキーワードを利用し、インターネット上でどの意見が多いかというウェブマイ

ニングから得られる数値である。例えば「ビール」の Positiveness は高いが、「冷たいビール」のほうが好まれることを数秒で発見することができる。一方、「暖かいビール」をシステムに入力すれば、Positiveness が極めて小さくなり、インターネットユーザの好みが明確に抽出される。このようなメカニズムにより、あらかじめ膨大なデータをシステムに入力することなく、さらに複雑な処理を行なうアルゴリズムも必要とすることなく、一般的な知識の獲得を行うことができる。インターネットユーザの大半による意見が誤りであったとしても、それは本手法で「人間らしさ」を実現したと考えられ、人間の性格をモデリングすることによって重要な情報となる。このような考え方は今までコンピュータ上での「人間らしさ」という観点からは存在していない。このようなメカニズムによりシステムにランダムではなく、インターネットユーザの一般的な意見によって決まる「予想のできない要素」を与えることができる。本手法によりプログラマーによって予め与えられた選択の結果を利用するのではなく、インターネットを用いて社会一般が与えた選択を利用することにより、誰かに作られたものは知能ではなく、個人の知能のコピーというだけであるという人工知能の一番大きい弱点とも言われる問題が解決される可能性が示された。もう1つの指標である「Usualness」のアイデアは、Shannon の情報量の独特な解釈を源としている。普段キーワードなどの抽出に利用される情報量を発言のそれぞれの部分の面白さを評価する方法として利用している。

次に、従来の対話システムは常識が不足しているために「当たり前な知識」の分析が非常に困難である。しかし、アルキメデスが主張していたように、コモンセンスとカテゴリー化は知能への道であると考え、コモンセンスを抽出しながらカテゴリー化を実現する Bacterium Lingualis という手法の提案を行なった。本手法はヨーグルト菌のような働きをモデル化したものである。本手法で実現される機能は WWW が頭脳と仮定すると、ばらばらに収集された知識をまとめて整理をするものである。他の手法と異なり、日本語の助詞と共起頻度によってカテゴリーが作成されるという考え方である。例えば、インターネット上で札幌とニューヨークは「に・で・まで」という助詞とセットでもっとも高い頻度で出現するので、NiDeMade という大きいカテゴリーに分類され、そのなかでこのセットに頻繁に付属する動詞の頻度がサブカテゴリーを決定する。Bacterium がカテゴリー化の能力以外に、Positiveness の指標がどんな状況で変化するという抽出ができる。例えば、「雰囲気」という単語が感情的にニュートラル (P の値が 3) であっても、「試合の前の、ぴりぴりした」雰囲気や「静かなバーの落ち着いた」雰囲気の場合などでは、Positiveness レベルが変化することがある。Bacterium がそのそれぞれの状況を発見できる。このことを第3章において実験とその結果として述べ、第4章では Bacterium Lingualis 手法のロボット用の命令分析などの工学的なアプリケーションへの導入方法の有効性を示した。Positiveness を測定しながら、利用者の発言から状況を判断し、状況が好ましくなければ、Positiveness を変化させる方法とそれに必要な手段の情報を自動的にインターネットから抽出するアルゴリズムの提案及び本手法の有効性の実験による確認を行った。

以上を要約すると、著者は Positiveness 及び Usualness という指標を導入し、様々なアプリケーションに利用できる新たな常識及び感情情報抽出手法の提案を行い、その有効性の確認を行った。さらに、オリジナルのカテゴリー化を行ない、知識・言語獲得ができる

Bacterium Lingualis という生物学的なアイデアに基づいたロボットが感情情報の変化の理由の抽出を行なうことができることを示した。本研究を通じて、情報メディア工学、自然言語処理工学への貢献は大なるものがある。よって、著者は博士（工学）の学位を授与される資格あるものと認める。