(続紙 １)

京都大学	博士（情報学）	氏名	岩井　律子
論文題目	Development of Methodologies and Language Resources for Acquiring Social Knowledge about Personality and Driving-related Behavior: The Synergy of Psychology and Natural Language Processing. （パーソナリティと運転関連行動に関する社会的知識獲得のための方法論および言語資源の構築:心理学と自然言語処理の共働）		

（論文内容の要旨）

人間の心理的特性に関する個人差への興味は，古くから哲学の研究対象であったが，近年，心理学の分野では急速な理論化が進められている。また，最近では，情報学の知識や技術の発展に伴い，個人特性への情報学的なアプローチが可能となってきた。そのような背景のもと，本論文は，人間のパーソナリティと運転行動の関係に関わる社会的知識に着目し，パーソナリティに関する心理学的な理論や心理学的な統計解析手法に加えて，大規模なテキストデータに対する自然言語処理の知識や技術を用いて行われた研究をまとめたものであり，９章から構成されている。

第１章では，本論文の構成について述べられ，続いて第２章ではパーソナリティ，特に本論文で取り上げたパーソナリティのBig Five理論の背景や現状，また，Big Five理論と行動との関係，さらには，それらに関する社会的知識に関する研究の概説がなされている。さらには，社会的知識が多数の著者によって記述された大規模なテキストデータ中に含まれること，また，それらが自然言語処理の方法によって解析可能であることなどを論じ，本論文のアプローチが妥当であることが議論されている。

第３章では，本論文のオリジナルな方法論として自動概念翻訳（automatic conceptual translation）アプローチが提案されている。その方法論を応用し，英語のパーソナリティを表す形容詞から，自然言語処理によって単語分散表現を用いて概念を包含する語を取得し，それらを統計的機械翻訳によって翻訳することで，逐語訳ではなく，概念的に等価な日本語単語を得た。大規模なサンプル（調査対象者）に対して調査を繰り返し（７調査,合計46,068名），ここで得られた日本語単語でも，英語単語と同様のBig Fiveの５因子構造が確認できること，また，この５因子構造は，一定期間を空けた後でも，個人内で一貫していること（再検査信頼性）に加え，基準関連妥当性，予測的妥当性も確認できたことを示すことで，構成概念妥当化の検証を行った。最終的に，選ばれた単語がBig Fiveに照らして高い信頼性と妥当性を有することが明らかにされた。この結果に基づいて，本論文では，２０項目からなる新たなパーソナリティの質問紙を構築した。

第４章では，第３章で得られた日本語のパーソナリティ単語を元に，単語分散表現を用いて，関連語を取得した。それらを，直接的にパーソナリティ特性を表す語（パーソナリティ特性語）と内容語を追加するなど表現を工夫すればパーソナリティを表現することができる語（パーソナリティ関連語）に分類した。これらに対して大規模な調査（1,938名）を実施し，各表現が５因子のパーソナリティ特性をどの程度反映するかを予測する重みを計算した。最終的に，パーソナリティの各特性を予測する重みが付与されたパーソナリティを表現する語に関する辞書を作成した。

第５章では，運転行動に着目し，心理表現を含む運転関連表現辞書の構築が行われた。これまで，自動車運転に関する表現の辞書で公開されたものはなく，特に，心理表現（目視をシッカリなど）が含まれたものは存在しなかった。そこで，テキスト解析の手法等により，運転関連語（1,173語），運転行動語（836語）を得た。運転行動語を用いて大規模テキストから述語項構造の抽出を行い，運転行動表現，運転文脈表現，心理表現，運転周辺表現の４種類のラベルが人手で付与された運転関連表現語辞書を構築している。

第６章では，運転体験が記述された文章中から，著者の行動，他者の行動，また，

その際の著者の主観体験を自動的に区別できるか否かの実験が行われた。アノテーションのガイドラインを作成し，これに基づいて人手でアノテーションされた結果を用いて学習したCRF(Conditional random fields)モデルによって，自動アノテーション実験を行った。その結果，行動表現と主観表現が高い精度で自動アノテーション可能であることが示された。

第７章では，大規模テキストデータから社会的知識が取得可能であるかの検討がなされた。まず，大規模テキストから，第４章で作成したパーソナリティ特性語を抽出した。また，同様に大規模テキストに対して，第６章で構築したCRFモデルによって，行動表現を抽出した。さらに，同一テキスト内におけるパーソナリティ特性語と行動表現の共起の抽出を行った。ここで得られたパーソナリティ特性語と運転行動表現が，一般に共有されている社会的知識に相当するか否かを，クラウドソーシングを用いて検証した。その結果，パーソナリティと運転行動に関する社会的知識として266項目が得られた。

第８章では，結果をまとめるとともに，今後の課題や今後の応用に関する議論を行っている。９章では，本論文の結論が述べられている。

注）論文内容の要旨と論文審査の結果の要旨は１頁を３８字×３６行で作成し、合わせて、３，０００字を標準とすること。
論文内容の要旨を英語で記入する場合は、４００～１，１００wordsで作成し
審査結果の要旨は日本語５００～２，０００字程度で作成すること。

（続紙　２　）

（論文審査の結果の要旨）

人工知能を搭載したシステム等がユーザの行動からその背後にある心理状態や心理特性を理解するためには，システムが，我々人間が共有している他者の行動と心理状態・特性に関する知識（社会的知識）を持つ必要がある。一方で，社会的知識が，具体的にどのような内容を含んでいるのかを明らかにする研究は未開拓である。本研究は，パーソナリティと運転行動を対象に，心理学の理論や概念妥当化の方法論と，自然言語処理や統計的機械翻訳の技術を融合することで，具体的な社会的知識の内容を明らかにすることを試みた。また，その過程で，新たなアプローチの提案や言語資源の構築を行った。得られた成果は以下の通りである。

（１）自動概念翻訳というアプローチを提案した。これは，多義的な英語のパーソナリティ表現語を日本語に逐語訳するのではなく，概念的に的確に意訳された語を，自然言語処理と統計的機械翻訳の手法によって獲得しようとするものである。この方法で得られた日本語のパーソナリティ語が，本来の英語で想定されているものと概念的に同等であることが，心理学的に厳密な妥当化の方法によって示され，本アプローチの有効性が検証された。また，簡易で信頼性，妥当性の高い日本語のパーソナリティ質問紙の構築を行った。自動概念翻訳アプローチは，パーソナリティ以外の心理学的な構成概念の翻訳にも適用できるものである。

（２）大規模テキストデータから，パーソナリティと運転行動に関する社会的知識を抽出するための言語資源（辞書等）の構築，整備を行った。我々は，任意のテキスト中からパーソナリティに関連する表現や運転に関連する表現を容易に抽出できるが，コンピュータによる自然言語処理では，そのようなことを可能にする事前知識を十分に有していないため，同様のことを行うのは非常に困難である。そこで，本論文では，自然言語処理や機械学習の方法を用いて，6つの言語資源の構築を行った。これらは，いずれも有用なものであり，関連研究の発展に大いに資することが期待される。

（３）大規模テキストデータから，パーソナリティと運転行動に関する社会的知識の抽出を行った。パーソナリティと運転行動について我々が有している社会的知識の内容を具体的に記述した初の成果である。本論文が用いた一連の方法は，運転行動とパーソナリティ以外の分野にも適用可能である。

以上，本論文では，パーソナリティと運転行動に関する社会的知識に関して，心理学と情報学の知識と技術の融合によってアプローチするという，世界的に見ても先駆的な試みがなされている。特に，自動概念翻訳というアプローチや，我々の潜在的な知識を大規模テキストデータから抽出するという本論文全体のアイディアは心理学における様々な分野に適用可能なものであり，当該分野の研究の発展にも貢献することは想像に難くない。本研究は，その成果のみならず、情報学の他分野の研究に対する有効性を実証した点で学術上、寄与するところが少なくない。よって，本論文は博士（情報学）の学位論文として価値あるものと認める。

また，令和元年８月２６日，論文内容とそれに関連した口頭試問を行った結果合格と認めた。

なお，本論文は，京都大学学位規程第１４条第２項に該当するものと判断し，公表に際しては，当分の間，当該論文の全文に代えてその内容を要約したものとすることを認める。

要旨公開可能日：　　　　年　　　月　　　日以降