

審査の結果の要旨

氏名 シェック モスタファ アル マスム

本論文は「An Analytical Approach for Affect Sensing from Text (テキストからの感情センシングのための解析的アプローチ)」と題し、英文で記されており、7章から成る。

第1章「序論(Introduction)」では、テキストからの感情抽出を従来とは異なるアプローチで行っているという、本論文の特徴について述べている。ここでのアプローチは、まず第1フェーズとしてテキストの著者(場合によっては話者)の情緒(sentiment)がポジティブ/ネガティブあるいはニュートラルであるかの度合いを算出する。次いで第2フェーズで、認知心理学に基づく感情理論(具体的にはOCCモデル)に基づき、感情の検出を行うとしている。なお、対象とするテキストは英文テキストである。

第2章「テキストの情緒分析(Sentiment Analysis of Text)」では、まずテキストからポジティブ/ネガティブの情緒を識別する従来手法についてまとめている。従来手法として、感情語のキーワードスポッティング、語彙類似性、統計的学習法、常識ベース、領域依存の識別法の利用などを挙げている。しかし、これらの手法は不十分なところがあるとし、感情の認知的評価構造、自然言語の意味解析、実世界知識の利用により、より包括的な感情種別を対象として、より頑健性を持つ新手法を考案、開発したとしている。

第3章「自然言語学リソースと SenseNet (Linguistic Resources and SenseNet)」では、本研究のテキストからの感情認識の第1フェーズの中心となる、ポジティブ/ネガティブの程度 (Valence 値と呼ぶ) を算出する SenseNet について論じている。SenseNet の基礎構造は、単文単位(1述語から成る)の述語フレーム(主語, 述語, 目的語)の3つ組データに対して Valence 値を計算することである。この計算は、動詞, 形容詞, 副詞, 名詞の各単語の Valence 値を、WordNet, ConceptNet (総体として常識を表す集合知による70万文を意味ネットワーク構造としたもの)、及び Web 上の判断検索サイトを利用して求め、それらを否定的表現の扱いなども含めて定めた複数の規則で統合することにより、行っている。単語への Valence 値の割り当てを ConceptNet の利用により、ポジティブ/ネガティブの情緒に直接的に結びついた単語でなくても、通常良く用いられる文脈から判断して可能にしている。また人名などの固有名詞に対しても、Web 上の評判検索を利用することにより、Valence 値の割り当てを可能としている。1文は複数の述語フレーム3つ組データで表されることになるが、各述語フレームの依存関係に基づいて各 Valence 値を統合する規則を定め、文の Valence 値を算出している。更に、複数文から成るパラグラフに対して、文脈としての Valence 値を計算する方法を示している。テキスト文から述語フレームや依存関係構造の抽出には、意味解析パーサを用いている。

第4章「SenseNet の評価(SenseNet Evaluation)」では、3章の SenseNet の性能について人間の判断を基準にした評価と、既存手法との比較による評価を行っている。評価用データセットには、映画や製品の評判に関する4種を用いている。文章レベル、パラグラフレベルの評価で、提案手法はポジティブ/ネガティブの Valence 値の判定に関し、多様な文に対する頑健性及び判定精度等について、既存手法よりも優れていることを実証している。

第5章「テキストの感情分析(Emotion Analysis of Text)」では、SenseNet による述語フレームの3つ組構造データ、及びポジティブ/ネガティブの Valence 値算出を利用し、テキストか

ら従来よりも詳細で包括的な認知的評価理論に基づく OCC モデル (22 種の感情を定義している) の感情を検出する手法について論じている。このために、テキスト著者の認知的状態を表す 16 種の認知変数を導入し、テキストの意味解析結果からルールによりこれらの認知変数の値を求める方法を提示している。そして、SenseNet から得られる Valence 値とこれら認知変数の値から、OCC モデルの 22 種の感情を検出するルールを構成している。結果は 1 種の感情を確定的に出力するのではなく、可能性のある複数の感情を強度付きで出力する。既存手法より、より詳細なレベルで、かつ精度高く感情検出が可能であること、人間の判断とおよそ 80% の精度で一致する感情検出性能が得られていることを実験的に実証している。

第 6 章「開発した応用(Developed Applications)」では、開発した 4 種のシステムについて記している。第一は第 3 章の SenseNet を Web サービスのシステムとして実現したものであり、テキストを入力すると SenseNet による述語フレーム 3 つ組データの分析結果のグラフィカル表示や、テキストの Valence 値の算出結果を表示する。第二の応用は ASNA(Affect Sensitive News Agent)と呼ぶシステムの開発である。この ASNA システムは、RSS(Rich Site Summary)により最新の Web ニュースを収集し、ニュース文を 8 種の感情(happy, sad, hopeful, fearful, admirable, shameful, loveable, hatred), 及びニュートラルを加えて 9 種に分類して、関連画像も付加して提示する。第三の応用は ESNA(Emotion Sensitive News Agent)と呼ぶ ASNA の拡張としてのシステムの開発である。この ESNA システムでは、固有名詞に対するユーザの嗜好を考慮したテキストの感情分類が行われる。(例えば、イタリアサッカーチームのワールドカップ優勝のニュースは、イタリアに対して嫌いの嗜好が与えられていれば sad として分類されることになる。) 第四の応用はオンライン・テキスト感情検出システムであり、ユーザ入力テキストから検出される感情を提示する。

第 7 章「要約と結言(Summary and Conclusion)」では、本論文の研究を要約している。

以上を要するに、本論文はテキストからの感情抽出に関し、認知心理学に基づく感情理論を背景として、既存手法より深くかつ詳細に分析する解析的アプローチとして、第 1 フェーズとしてポジティブ/ネガティブの情緒(sentiment)を算出し、第 2 フェーズで包括的な OCC モデルに基づく 22 種の感情を判別する手法を提示している。第 1 フェーズに対しては、Web リソースとして利用可能な実世界知識を利用することにより、直接的な感情語ではない単語に対しても情緒を求め、述語フレームの 3 つ組データ構造 (主語, 述語, 目的語) を基礎単位にして、規則により統合することでより大きな単位のパラグラフ、文章の情緒を頑健に算出する SenseNet を創案、開発している。第 2 フェーズに対しては、OCC モデルの 22 感情の判別に必要なテキスト著者の認知的状態を表す認知変数を導入し、テキストの意味解析結果からこれら変数値を求める方法を示し、SenseNet 分析結果とあわせて 22 感情を検出するルールを構成し、感情抽出を実現している。提案手法の性能は実験により既存手法に対して分析の詳細度、精度について優れたものであることを実証すると共に、応用として Web ニュース文を感情により分類するなどのシステムを開発し、実用的にも有用であることを示している。これらの成果はテキストからの感情抽出の研究の新段階を切り拓いたものとして意義が認められ、電子情報学上貢献するところが大きい。

よって本論文は博士 (情報理工学) の学位論文として合格と認められる。