

Web上のテキストデータを用いた  
道徳的常識の自動獲得に関する研究

2018年度

山本 眞大

報告番号	① 乙 第	号	氏 名	山本 眞大
主 論 文 題 名 : Web 上のテキストデータを用いた道徳的常識の自動獲得に関する研究				
(内容の要旨) 対話システムやコミュニケーションロボットが登場し, 人間と機械の間のインタラクションの機会が増えてきている. このようなロボットの急速な社会進出に伴い, 人工知能に道徳的な知識を獲得させる必要性が指摘されている. しかしながら, 先行研究の多くは人手により記述された知識やルールを用いた処理が主であり, コスト面等で大きな問題がある. さらに, 対象となるドメインが限られており, 日常生活などの一般的な状況における道徳性に関しては考慮されていない. これらの問題を解決するために本論文では, 道徳的常識を低コストで獲得することを目的とする. そのためのアプローチとして, (1)Web 上のテキストデータおよび評価表現, (2)単語の高次元ベクトルである分散表現, (3)自然言語処理分野で培われてきた知識獲得技術を利用した新しい手法を提案し, 上記の課題の解決を目指す. 第 1 章では, 本研究の目的, 意義を明らかにすると共に, ロボット倫理学の変遷, 自然言語処理分野の概要及びその周辺に関して関連する研究について述べる. 第 2 章ではまず, 道徳的な常識の獲得度合いを評価するための道徳判断タスクについて説明する. 次に, Web 上のテキストデータおよび評価表現を用いて道徳判断を行う手法を提案する. これは基本的には共起頻度を基にした手法であり, 評価実験により, 結果の考察や得られた課題の検討を行う. 第 3 章では, 第 2 章で提案した手法の対応言語を英語に拡張した手法の提案を行う. 具体的には, 単語を高次元ベクトルで表現する分散表現を導入し, さらに述語項構造データベースによる連想情報を用いて道徳判断を行う手法を提案する. 評価実験により, 特に第 2 章で課題であった長い入力文に対する道徳判断の精度が向上することを示す. 第 4 章ではまず, 評価表現, 接続表現, 構文情報などの言語的パターンを用い道徳に関する擬似ラベル付きデータを作成する手法を提案する. 次に, 注意機構付き LSTM(Long short-term memory)を導入した深層学習による新しいシステムを提案する. 評価実験により, 高い精度で道徳判断が可能になることを示す. 第 5 章では, 本研究において取り組んだ道徳判断タスクについてまとめ, その成果を要約する. また, 今後の展望や残された課題についても述べる.				

## Thesis Abstract

No. \_\_\_\_\_

Registration Number	<input checked="" type="checkbox"/> "KOU" <input type="checkbox"/> "OTSU" No.                                      *Office use only	Name	Masahiro Yamamoto
Thesis Title A Study on Automatic Acquisition of Moral Common Sense using Web Corpus			
Thesis Summary <p>In recent years, communication opportunities between humans and robots are becoming more and more popular. When we take into account this situation, it is considered that common sense about morality plays an important role. However, implementation costs of the previous methods are very high because these methods require hand-crafted rules. Furthermore, the general ethics such as daily life morality is not considered.</p> <p>To tackle this problem, we present the study of the automatic acquisition of moral common sense. In order to achieve the goal, we employ (1) text data on the web and evaluation expressions, (2) distributed representation of words, and (3) knowledge acquisition methods of natural language processing (NLP).</p> <p>In chapter 1, we describe the historical background of studies related to our study. Starting from robot ethics, we introduce artificial intelligence, natural language processing, and related topics.</p> <p>In chapter 2, we explain a moral judgment task to evaluate morality of the method. We then propose a moral judgment method using text data on the web and evaluation expressions. This method is based on the co-occurrence frequency. In evaluation experiments, we show results of moral judgment and analyze it.</p> <p>In chapter 3, we aim to extend the method for another language, English. In concrete, we introduce distributed representations which can express words as high dimensional vectors and the associative information in a database of the predicate argument structure. We then propose a moral judgment method using them. Experimental results show that the accuracy is improved especially when the input is a long sentence.</p> <p>In chapter 4, we propose an automatic acquisition method of the moral common sense from the web text data using language patterns such as evaluation expressions, connection representations and syntax information. We then propose to use these data as training data for supervised learning and apply logistic regression model based on N-gram features and attention-based LSTM (Long short-term memory). Experimental results show that the accuracy of moral judgment is significantly improved compared to the other methods.</p> <p>In chapter5, we summarize each chapter and discuss the future work of this study.</p>			