



HOKKAIDO UNIVERSITY

Title	Expanding Common Sense Knowledge Database using Data Mining on Reference Corpus [an abstract of dissertation and a summary of dissertation review]
Author(s)	KRAWCZYK, Marek
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12640号
Issue Date	2017-03-23
Doc URL	https://hdl.handle.net/2115/65725
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Marek_Krawczyk_review.pdf, 審査の要旨



学位論文審査の要旨

博士の専攻分野の名称 博士 (情報科学) 氏名 KRAWCZYK Marek

審査担当者 主 査 教 授 荒木 健治
 副 査 教 授 山本 強
 副 査 教 授 長谷山 美紀
 副 査 教 授 坂本 雄児

学位論文題名

Expanding Common Sense Knowledge Database using Data Mining on Reference Corpus

(参照コーパスに対するデータマイニングを用いたコモンセンス知識データベースの拡張)

今日の情報社会では、幅広い背景知識がなければ、創造的かつ楽しい会話を行うことがますます困難になっている。人々は概念やアイデアについて議論することを好み、またそれを行うためには、幅広い話題にまたがる常識と一般知識の融合が必要である。コンピュータがこのような会話を理解し、自然言語を使用する人々との対話を行うためには、構文などの言語的特徴を分析するだけでなく、意味を処理する手段、すなわちテキストの表層に表現されている概念の知識が必要である。そうした概念情報は、Cyc, YAGO, ConceptNet などの大規模な一般的な知識ベースに集約されている。

本学位論文では、人間の一般的な知識を記述する大規模なセマンティックグラフを提供する知識表現プロジェクトである ConceptNet に焦点をあてる。ConceptNet を利用した理由は、このデータベースが幅広い常識の概念や関係を捉えており、単純な意味ネットワーク構造が使いやすく操作しやすいという特徴を持つためである。ConceptNet の基本データユニットはアサーションと呼ばれ、これは2つの概念とこれらの概念間の関係を表す。ConceptNet は Open Mind Common Sense プロジェクトのウェブサイトで収集された知識だけではなく、複数の言語で一般的な知識を自動的に収集する類似のウェブサイトやオンライン単語ゲームの知識を含むように設計されている。このオープンソースの知識ベースは、トピック推定、感情認識、対話システム、日常活動の認識、ソーシャルメディア分析、手書き文字認識などの多くのアプリケーションで使用されている。

一方、人手による知識ベースの拡張は長くて労働集約的なプロセスになる。例えば、2010 年に開始されて以来、連想ゲームを使って知識を集めているオンラインプロジェクトの nadya.jp は ConceptNet にわずか 43,500 件のエントリーを投入することしかできていない。したがって、新しいデータを収集するためには自動的な知識獲得の手法を開発する必要があることは明らかである。従って、ConceptNet の知識を自動的に獲得する手法の開発を行うことが本研究の目的である。

本学位論文では日本語ウィキペディアのデータから IsA アサーション (下位関係), AtLocation アサーション (ものまたは場所の位置を示す), LocatedNear アサーション (近隣の場所を示す), CreatedBy アサーション (オブジェクトの作成者を示す), または MemberOf アサーション (グループのメンバーを示す) を自動的に抽出する方法を提案する。提案手法では、IsA アサーションと拡張した階層構造の処理結果を抽出するために、ウィキペディアの記事の定義文、カテゴリ、階層構造を分析することにより上位下位関係を抽出するツールを使用する。このツールによって拡張した階層構造の処理結果から提案手法を用いて、追加の情報、すなわち AtLocation, LocatedNear, CreatedBy および

MemberOf アサーションを抽出する.

性能評価実験では, 本研究の目的が大規模なデータで達成されたことを証明した. 96.0

提案手法および実験結果から大規模な参照コーパスから一般のおよび常識的知識を自動的に抽出することが可能であることが確認された. 一方で, より洗練された手法を適用することによって提案手法をさらに改善させることが可能であるが, このアプローチは様々な人工知能関連アプリケーションに有用なコモンセンス知識データベースを拡大するために今後の研究が必要とする方向性を示していると考えられる.

これを要するに, 著者は, コモンセンス知識データベースの構築において, 新たに参照コーパスにデータマイニングを適用することにより拡張することを提案し, 性能評価実験を行うことによりその有効性を示したものであり, 自然言語処理工学の発展に貢献するところ大なるものがある. よって著者は, 北海道大学博士 (情報科学) の学位を授与される資格あるものと認める.