

論理的な対話を可能とする人工知能の基礎技術を開発

賛否が分かれる議題に対し、大量のテキストデータを解析し、賛否の根拠や理由を提示

株式会社日立製作所(執行役社長兼 COO:東原 敏昭/以下、日立)は、賛否が分かれる議題に対し、大量のテキストデータを解析し、肯定的もしくは否定的な意見の根拠や理由を英語で提示する技術を開発しました。本技術は、意見を述べる際に人やコミュニティに重要と考えられる健康や経済、治安などの価値に着目し、世の中の事象とそれぞれの価値との相関関係を用いて、大量のニュース記事から、より確実性の高い根拠や理由を抽出します。複数の価値を基準にすることで、ひとつの側面に偏ることのない根拠や理由を提示します。

本技術は、人とコンピュータの論理的な対話を可能とする人工知能の実現に向けた基礎技術であり、将来、企業が持つ文書や公開されているレポート、病院の電子カルテなどを解析し、業務を支援するデータや意見を生成するシステムへの応用が期待されます。

近年、インターネットなどの情報通信技術や分析技術の発展にともない、各種センサーや POS システムなどから日々生み出されるビッグデータを解析し、有益な情報を抽出する技術が注目されています。新たな価値の創造や顧客の経営判断を支援するために、企業が持つ文書や公開されているレポート、病院の電子カルテなどのテキストデータを有効活用することが求められています。また、テキストデータを活用した人と論理的な対話が可能な人工知能の開発については、テキストデータにおける事象の持つ価値や、それらの相関関係を抽出することが難しいため、これまでその実現は困難でした。

そこで日立は、2014 年に開発した、電子カルテから病態や病気の部位などの指定された情報を高精度で抽出する技術^{*1} に今回新たに開発した技術を組み合わせることで、与えられた議題に対して、大量のニュース記事を解析し、より確実性の高い根拠や理由を英語で提示する技術を開発しました。

開発した技術の概要は以下の通りです。

(1) 賛否の根拠や理由を抽出するための基準となる価値体系辞書を作成

人は賛否が分かれる質問に対して根拠や理由を述べる際に、人それぞれが持つ価値を基準とすることが推測されます。そこで、人やコミュニティにとって重要と考えられる健康や経済、治安などの価値に着目し、多数の議題に対する賛成と反対の意見を登録したディバートのデータベース^{*2}をもとに、価値を体系的に纏めた価値体系辞書を作成しました。具体的には、人やコミュニティが判断をくだす際の本根にある価値をリスト化するとともに、それらの価値と関係が深い単語をデータベースでの使用頻度に基づいて抽出し、価値に対してポジティブかネガティブかに振り分けます。さらに、使用頻度に応じて重要度を付与することで、価値とそれに関連する単語を体系的に整理しました。例えば、「健康」という価値においては、「運動」はポジティブ、「病気」「肥満」はネガティブなどのように単語の関連性を体系的に整理しています。

(2)大量のテキストデータから事象と価値の相関関係データベース^{*3}を作成

大量のニュース記事の中で使用されている様々な文章の中から、記載されている事象がどのような価値をもたらしているかを抽出し、その事象が価値に与えた影響がポジティブかネガティブかを示す相関関係のデータベースを作成しました。例えば、「騒音が健康被害をもたらす」という記事からは、「騒音」という事象が、「健康」という価値を抑制するというネガティブな影響を抽出し、データベースで管理します。この手法により、約970万件のニュース記事から、約2億5千万からなる相関関係データベースを作成しました。

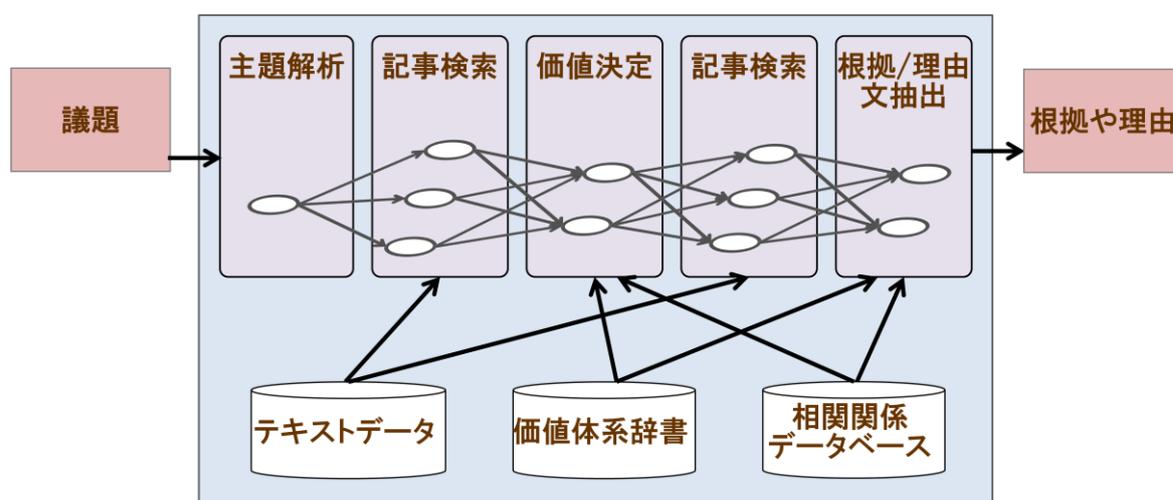
与えられた議題に対して、この相関関係データベースと(1)の価値体系辞書を活用し、大量のニュース記事の中から、議題と関連性の高い価値を複数選定します。全てのニュース記事の中から、この選定された価値のいずれかを含む文を検索することで、議題に対して根拠や理由となる可能性のあるものを抽出します。

(3)抽出した根拠や理由となる可能性のある文について確実性を算出

(1)の価値体系辞書と(2)の相関関係データベースを活用して抽出した文を、引用元の記載や数値データの有無、使われている表現などの指標を用いて数値化することで、議題に対して関連性の高いものであるかどうかを判定します。根拠や理由となる可能性のある全ての文にこの処理を行い、数値を算出することで、より確実性の高い文を選出し提示することができます。

(4)多数のアルゴリズムを非同期かつ分散的に実行するアーキテクチャ^{*4}を構築

議題を与えられてから、根拠や理由を提示するまでの処理速度を速め、途中で停止することなく、指定した時間にある程度の回答を提示するために、主題の解析や価値の決定、記事の検索、根拠や理由を提示するまでの各プロセスにおいて、多数のアルゴリズムを非同期・分散的に実行できるアーキテクチャを構築しました。本アーキテクチャは、一つのアルゴリズムを並列に分散処理するとともに、次のプロセスへの非同期な処理を行うことで、指定した時間内に根拠を抽出することができます。



議題の解析から根拠/理由を提示するまでのプロセス

本技術は、東北大学(総長:里見進)大学院情報科学研究科の乾・岡崎研究室の協力を得て開発したものです。日立は、2014年に開発した技術にこれら4つの技術を組み合わせることで、大量のテキストデータを解析し、肯定的もしくは否定的な意見の根拠や理由を提示する技術を開発しました。今後、更なる研究開発を推進し、人とコンピュータの論理的な対話を可能とする人工知能の実現をめざします。

なお、本成果は2015年7月26日～31日に中国で開催される国際会議 ACL-IJCNLP 2015 (53rd Annual Meeting of the Association for Computational Linguistics and 7th International Joint Conference on Natural Language Processing)にて発表する予定です。

- *1 日立製作所ニュースリリース「電子カルテから病態や病気の部位などの指定された情報を高精度に抽出する技術を開発」
2014年9月17日発表
- *2 ディベート普及団体が提供する、トピックごとに肯定側・否定側の争点を纏めた巨大データベース「Deatabase」を使用
- *3 相関関係をメタデータとして管理したもの
- *4 情報システムの構成に関する基本設計および設計思想

■照会先

株式会社日立製作所 研究開発グループ 情報企画部 [担当:有金、安井、木下]
〒185-8601 東京都国分寺市東恋ヶ窪一丁目 280 番地
電話 042-323-1111(代表)

以上

このニュースリリース記載の情報(製品価格、製品仕様、サービスの内容、発売日、お問い合わせ先、URL 等)は、発表日現在の情報です。予告なしに変更され、検索日と情報が異なる可能性もありますので、あらかじめご了承ください。
