

論文内容の要旨

博士論文題目 Computation of Semantic Equivalence for Question Answering
(質問応答のための意味的等価性の計算)

氏名 高橋 哲朗

(論文内容の要旨)

現在 Web を中心に大量の電子テキストがあふれており、それらに対する効率的な情報アクセスの技術が求められている。その手段の一つとして質問応答がある。質問応答とは、自然言語で問われた質問に対して、文書集合を情報源として回答となる語句を返す技術である。本研究では質問応答タスクを、質問文とドキュメントにおける要素間の意味的関係の等価性を認識する問題ととらえ、これを追及することによって厳密な質問応答の実現を目指す。

等価性の評価方法として本研究では、以下の3種類のアプローチを取る。

(1) 先行研究では、テキストを単語の列ととらえ単語間の類似度や近接性などを用いて類似度を求める手法が主流であったが、意味的類似性のより厳密な近似を行なうために構文構造を用いた。本研究では Collins の提案した Tree Kernel を元に、類似度の計算のアルゴリズムを拡張し、計算量を抑えながら対応ノード間の類似度の定量化やノードの飛び越えを許す照合を実現するアルゴリズムを提案した。

(2) テキストに対する語彙・構文的な操作を行なう汎用的な言い換えシステムを開発した。そして質問応答における言い換えの適用方法として、言い換えと照合を繰り返しながら回答を探索するモデルを提案し、質問応答システムを実装した。このシステムを用いて質問応答タスクの実験を行なった結果、現在の実装では我々の予想に反して構文的照合と言い換えの効果が小さいことが明かとなった。また誤り分析を通して、(a) 照応・省略解析、(b) 複数の可能性を持たせた構文森における照合、(c) より大規模な言い換え知識の獲得の3点を課題として認識することができた。

(3) 上記の(c)の問題に着目し、質問応答に必要な言い換えのバリエーションや、解かなければならない問題点について調査した結果、質問応答に必要な記号に関する言い換えや、語彙・構文的言い換えの種類を洗い出すことができた。また、語彙・構文的な言い換えの範囲を越えた言い換えについても整理を行ない、それらの分析を通して、問題の一部については属性関係を抽出することによって解決できる見込みを得ることができた。

そこで、質問応答のサブタスクとして、(対象物, 属性名, 属性値)という三つ組からなる属性関係を抽出するというタスクを設定した。このタスクはこれまでに研究が進められてきた情報抽出を一般化する新しい問題であるため、従来手法とは異なるアプローチが必要である。この問題に対して我々は、抽象化したパターンによりドメインを限定せずに三つ組の候補を抽出し、統計量を使ってそれらをフィルタリングする手法を提案した。実験の結果、この手法を用いることによりパタンのみで抽出した場合に比べ抽出精度を F 値で 6.1 ポイント (10%) 向上させられることを示した。

氏名	高橋 哲朗
----	-------

(論文審査結果の要旨)

平成16年12月27日に開催した公聴会の結果を参考に平成17年2月17日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

高橋 哲朗は、本博士論文において、従来の質問応答システムの性能を向上させるため基礎技術として、文の記述内容の等価性を判定する技術や質問応答に必要な知識を自動獲得する手法等の研究を行い、そのための手法を提案した。具体的には、次のような観点から研究により質問応答システムの性能向上の実現、および、改善のための可能性についての提案を行った。

1. 文の意味的類似性の定量化するために、Collins が提案した Tree Kernel を拡張し、対応する節点間の類似度や節点の飛び越しを許す照合を考慮したモデルを提案した。質問応答システムでは、質問文と意味的に近い内容を記述した文を大規模テキストから探索する必要があるが、これに必要となる文の間のより厳密な類似度計算が可能になり、質問応答の性能向上につなげることができた。
2. 質問文と回答候補を対応付けるために、それらを言い換え規則によって言い換えと照合を行いながら探索するモデルを提案した。それと並行して、一般的な言語言い換えシステムの開発も行った。言い換えと照合を行うことによって質問文に対する回答を求めるモデルは、直接的なシステムの性能向上にはつながらなかったが、いくつかの問題を明らかにし、大規模な知識抽出の重要性を示した。
3. 質問応答システムに必要な知識の整理を行い、その中から、大規模テキストから事前に取り出すことに価値のある定型的なパターンとして対象物、属性名、属性値の組合せによる属性関係の抽出が重要であることを示し、その自動抽出に関する方法を提案した。また、抽出された知識により、質問応答システムの性能向上が可能であることを示した。

質問応答システムの性能向上のための種々の手法を提案した本研究は、独創性が高く、しかも実用的であり、自然言語処理の分野において高い貢献があると評価する。よって、本論文は、博士（工学）の学位論文として価値あるものと認める。