

3.5.3 ユニバーサルコミュニケーション研究所 情報分析研究室

室長 鳥澤健太郎 ほか 14 名

情報の意味を深く分析する

【概要】

インターネット上には膨大な情報が存在し、非常に多くの人々が検索エンジンなどを用いてそうした情報にアクセスしている。一方で、情報の間には様々な繋がりがあり、本来はそうした繋がりをきちんと見ていくことで初めて本当の意味で情報を活用することが可能になる。例えば、ある情報が何らかの方法で見つかった時に、その情報と意味が同じであるような情報がどの程度存在するか、さらにはその情報と矛盾する情報がどの程度の頻度で存在するかが分かり、さらにそうした情報を誰が発信しているか（例えば、匿名ブログか、新聞等の報道か、公的機関の発表か）が把握できれば、最初の情報が信頼できるかどうか判断をする手がかりとなる。さらに、その情報の帰結（例えば、ある事件がどのような帰結をもたらすか）や原因（ある事件がおきた原因はなにか？）が分かれば、やはりその情報をもとに意思決定をする際に有効である。また、最初に見つけた情報と、帰結、原因といった情報の間の関係はインターネット上で明示的に書かれているとは限らず、情報システムが「考えて」、仮説として原因帰結をユーザに提供する必要もある。例えば、昨今、ヨーロッパにおける経済破綻や、外国における領土問題、自然災害などが日本国内に想定外の大きなインパクトをもたらすといったような事例が多数見受けられるが、その影響自体は、影響が生じてからでないと分からず、また、そうした影響によるダメージも避けようがないというのが多くの人々の受け取り方であろう。こうした影響を前もって調べる、あるいは予測して、それによるダメージを軽減するためには、少なくとも現状の情報アクセス技術は無力であり、より深い情報の分析や、先に述べたような「仮説」の生成が必要となる。現在、我々は最終的には情報のこうした高度な分析を自動化することを目標として、基盤となる情報分析技術を開発している。

また、2011年の東日本大震災では、災害時の情報アクセス、および情報の分析の機能不全、あるいは必要性がクローズアップされた。我々は上述したような情報の分析技術を災害情報にも適用し、いざと言う時に適切に必要とされる情報を取得し、また、その背景等を分析するエンジンも開発する。

【平成 23 年度の成果】

本年度は、5年間にわたる中期計画の最初の年ということもあり、基礎的な研究に集中した。本稿ではそのうち、主要な成果3点について述べる。

(1) 動詞の意味的分類に基づく、矛盾、因果関係の自動取得手法

概要で述べたように、情報の有効活用をするためには、ある情報と矛盾する情報や、その情報が表す出来事の帰結、原因を調べることが重要である。こうした調査を自動的に可能とするため、情報の非常に基礎的な単位と見なせる動詞句、特に名詞1つと動詞1つからなるものの間の矛盾関係（例：「ガンを破壊する」⇔「ガンを進行させる」）を誤りも含まれるものの100万対オーダーで自動抽出する手法を開発した。また、因果関係（例：「輸出が増える」⇒「GDPが増加する」）といった関係を、「輸出が増え、GDPが増加する」のように直接的に因果関係を表すとは言えない日本語表現から、やはり100万対オーダーで自動取得する手法を開発した。また、上述の矛盾と因果関係を組み合わせ、Web中に記載のない因果関係も仮説として生成することも可能とした。例えば、「デフレが進む」⇒「税収が減る」という因果関係と「デフレが進む」⇔「デフレを阻止する」と「税収が減る」⇔「税収が増加する」という2つの矛盾の関係を組み合わせることで「デフレを阻止する」⇒「税収が増加する」といった仮説をWebに直接記載がなくても、生成することが可能である。

こうした一連の技術の核となっているのは動詞の自動的な意味的分類である。我々は動詞と共に現れている名詞の主たる機能、効果が発揮されていることを表す一群の動詞を「活性」、機能、効果を抑制する動詞を「不活性」と呼ぶことにし、動詞を活性／不活性に自動的に分類する技術を開発した。例えば、活性の動詞の例を挙げると「Xが発生する」「Xが悪化する」といったものがあるが、これらはいずれも「台風が発生する」「ガンが悪化する」のように、名詞Xの機能、効果が発揮されていることを意味する動詞である。一方で不活性の例を挙げると「Xが消滅する」「Xが治癒する」のような動詞が挙げられる。こうした意味的分類自体がそもそも前例のない新しいものであるが、さらに我々は電子スピンのモデルを参考にして、そうした分類を数万語の動詞に対して自動的に行う技術を開発し、これを用いて、上述したような矛盾、因果関係に関して過去に前例を見ない規模のデータベースを自動構築した。

