

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 15 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量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、その情報が記述している出来事の帰結（例えば、ある事件がどのような帰結をもたらすか）や原因（ある事件が起きた原因はなにか？）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、我々は最終的には数十億件の最新の Web 文書などを対象として情報の高度な分析を自動化し、情報分析システム WISDOM の新しいバージョンとして一般公開することを目標として研究開発を進めている。

【平成 24 年度の成果】

平成 24 年度の成果のハイライトは、これまでに開発してきた様々な技術を統合した次世代情報分析システム WISDOM 2013 の稼働を開始したことである。この WISDOM2013 は毎日 1,000 万から 2,000 万ページの Web ページを収集し、最大で最新の 20 億ページに対して質問応答、評価情報分析等の様々な情報分析サービスを統合して利用可能にするものである。図 1 に示すように現在は「アトピーに効くのは何ですか」といった日本語の質問を入力として与えると、「軟膏」、「漢方」といった常識的なものから普通の人聞いたことのない珍しい薬品や物質まで、Web 中で記載されている様々な回答をタグクラウド風に表示する。これらの回答にマウスカーソルを移動すると、回答が抽出されたオリジナルの Web テキストが表示され、また、Web 文書自体へのリンクも表示されるため、オリジナルのテキストを読んだ後、リンクを辿って Web 文書を表示させることによって回答の妥当性、信頼性を判断し、またその他の関連情報を取得することができる。また、回答の各々に関して、評価表現、発信者情報の分析を行い、その結果を表示することができる。評価表現の分析とは現在公開中の情報分析システム WISDOM でも利用可能な機能であるが、各々の回答に対して肯定的評価（例：「漢方は効果的な治療法です。」）、あるいは否定的評価（例：「漢方は効果がありません。」）を下しているテキストを Web から抽出して、その分布、あるいは文そのものをユーザに提供するものである。発信者情報の分析とはそうした肯定的／否定的評価の発信者をテキストの文面から特定し、例えば、企業、匿名、大学といったカテゴリに分類し、その分布をユーザに提供するものである。例えば、ある商品に対する肯定的評価が企業による情報発信では多く、大学からの情報発信では否定的情報が多いといった場合には、企業による情報発信が宣伝等によってバイアスがかかっており、その信頼性が相対的に低い可能性があることなどが判断可能になる。

さらに質問応答に関しても、複数の質問を有機的に組み合わせることが可能になっており、効率的に有用な情報を発見することが可能になっている。例えば「アトピーに効くのは何ですか」という質問の回答のうち、適当なものを複数選択し、「【選択】を豊富に含むのは何ですか」という質問を新たに投入すると、選択された複数の回答の各々に関してそれを豊富に含む食品等が回答として表示される。こうした、質問をいわば連鎖させる機能によって、「アトピーに効果的な食品で比較的知られていないもの」を発見することも可能かもしれない。一般に通常の検索エンジンでは、アトピーという病名からその治療に役立つ物質を含む食品等に至る情報間の



図1 WISDOM 2013の利用の様子

つながりを網羅的にチェックすることはきわめて難しい。これは検索エンジンでは出力が単なる文書へのリンクであり、結局、テキストを咀嚼して知識のレベルまで高め、それらの間の関連性をシステムティックにチェックするという作業はあくまでユーザに任されているからである。WISDOM 2013の狙いの1つはそうしたWeb上の大量の情報がある意味で知識というレベルにまで高め、その相互の関係をより自動的に扱って、効率的な判断、意思決定を促すことである。

ここでもう1つWISDOM 2013の機能を例にとって、前述の情報・知識間の相互の関係を自動的に処理することの利点を示したい。図2はWISDOM 2013に対して、「少子化が進行するとどうなる？」という質問を与えたときの出力画面である。こうしたタイプの質問を我々は「どうなる」型質問と呼んでいるが、こうした質問を与えられると我々が社会シナリオ予測と呼んでいる情報分析サービスが起動される。「少子化が進行するとどうなる？」という質問に対しては、例えば、Webより自動抽出された「少子化が進行する⇒労働力人口が減少する」「労働力人口が減る⇒GDPが下がる」「GDPが低下する⇒エネルギー使用量は減る」といった3つの因果関係から「減少する⇔減る」などの言い換えも考慮して、「少子化が進行する⇒エネルギー使用量が減る」といった新たな因果関係、あるいは将来の可能なシナリオを生成することが可能になった。他にも、「少子化が進行する⇒学生が減る⇒少人数教育になる」「少子化が進行する⇒高齢化社会になる⇒耕作放棄が増える⇒(食料)自給率が下がる」「少子化が進行する⇒..⇒省力化投資が増加する」といった興味深いシナリオが多数生成されている。ここで強調したいことは、シナリオの生成では多数のWeb上に記載されている因果関係を組み合わせるが、組み合わせの結果得られたシナリオは1つのWeb文書にまとまった形で書かれているものである保証はなく、場合によっては誰も考えたことのないシナリオである可能性すらあるということである。これはつまり、システムがWeb上の情報を組み合わせることによって、人にとって想定外であるような未来のシナリオを仮説として生成、推論していることと言っても良い。本稿の冒頭でも述べたように、現代社会は様々な事象が複雑に絡み合っ、様々な想定外の事象を発生させている。今後はWeb上のビッグデータと前述したような自動的仮説生成、推論手法を組み合わせ、適切な意思決定を行う必要があると考えている。現状ではまだまだ意味をなさないシナリオも生成されてしまうが、今後は、より信頼のおけるシナリオが生成できるよう、こうした技術を情報発信者分析、社会のトレンドを認識する他の手法、また、論文等信頼のおける文書の解析結果等々と組み合わせることを検討している。

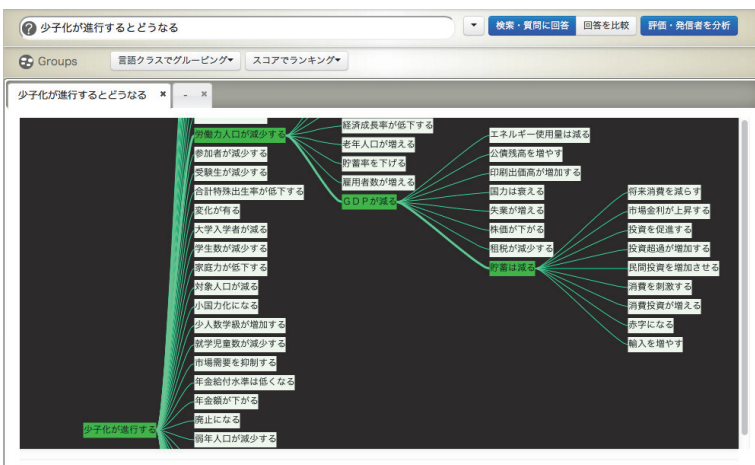


図2 WISDOM 2013における社会シナリオ予測

以上、平成24年度の成果のハイライトであるWISDOM 2013の概要について述べてきた。本年度の成果としては、トップレベルの国際学会において様々な高度な言語資源、言語処理技術の開発に関する論文が採択されたことも挙げられるが、本稿では割愛する。WISDOM 2013はそのようなハイクオリティな研究成果と、計算機クラスタにおける高効率な分散処理の実装とを組み合わせ初めて実現できた大規模なものである。このシステムはさらに改善をすすめ、平成25年度もしくは平成26年度に一般公開する他、プログラムについてもALAGINフォーラムを経由して公開を予定している。