

関連性理論による文章の図解化

An automatic illustration of texts based on Relevance Theory

石塚 隆男

Takao Ishizuka

亜細亜大学経営学部

Faculty of Business Administration, Asia University

We propose a new method for automatic illustration of texts based on Relevance Theory. Generally, illustration of texts promotes understanding of the meanings of the texts written by natural language. However, there are no tools which convert texts into figures or diagrams because of large degree of freedom. Relevance Theory is a persuasive hypothesis of pragmatics, and introduces the concept of labor or cost to accounting for recognition behavior of discourses. We apply Relevance Theory to get an optimal illustration of texts. Our method quantizes information content by the ratio of the number of old words to the number of new words in a paragraph. Old words mean words emerging in the predecessor paragraphs. We can illustrate a tree diagram by the distribution of the information content semi-automatically. Our method extracts paragraphs with high information by Relevance Theory. Relevance Theory is expected to a new paradigm of text illustration norm or criterion.

1. はじめに

今日、「図解」なるタイトルの書籍は言うに及ばず、図解を取り入れた多くの出版物が日々刊行されている。このように、図解化は文章の内容を容易に理解するための常套手段である一方、人手による高度な知的作業でもあり、あらゆる文章全体で図解化されているのはごく一部にすぎない。そこで、コンピュータにより文章を自動的に図解化できれば知識や情報の普及に大いに役に立つと考えられる。

一般に、文章とは概念の形容や概念間の関係を自然言語により叙述的に記述したものである。したがって、文章を図解化することは、文章から重要な概念を抽出し、それらの概念間の関係を叙述的でなく、図的に表現する（たとえば、包含関係や因果関係等）ことであり、要約の一種と見なすことができる。しかし、ひとつの文章にたいして図的表現の方法は多様であり、自由度が大きい。また、一口に図解化といっても、文章の全体像や流れを把握するために用いる図解化と、ある文の意味を理解するのに用いる図解化では方法に違いがある。

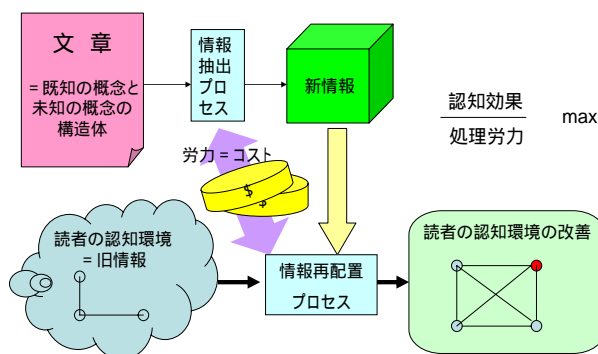
このように、文章の図解化を機械的に行うには多くの課題があるが、どんな図を描くか(how)の問題よりも何を(what)図にすべきかの方がより重要と考えられる。そこで、本研究では、文章から図解の要素となるべき概念を抽出するための基準として語用論の理論として知られる関連性理論の可能性について検討を行う。

今回、文章を構成する各パラグラフの情報量を定義し、情報量による図解化について関連性理論による解釈を示す。

2. 関連性理論について

関連性理論は、「人間は情報処理にあたって最大の関連性を目指す」、つまり「最小の処理労力で、できるだけ多くの認知効果を得ることを目的とする」(Dan Sperber 1995)ことを前提とした理論であり、高い説得性をもち、今後の発展が注目されている。関連性理論は、人間の情報処理行為を合理性の観点から説明したものであり、コストの概念が導入されている。しかしながら、処理労力、すなわち、コストをどのように定量化すべきかについては答えを待たず、本理論の数量的な扱いに関する研究はほとんど見当たらない。図1は、関連性理論に基づく認知プロセスを図示したものである。

図1. 関連性理論による認知プロセス



人間の認知行為には、情報抽出と情報再配置の2つのプロセスが関与しており、時間や労力を惜しまなければ無限のコンテキストや仮説を想定あるいは生成しうるが、フレーム問題に陥るため現実にはサイモンのいう限定された合理性あるいは満足度規準が働いていると考えられる。

私たちは本来、無限の解釈が可能と言う意味で無限次元の文章をそのまま認知しているのではなく、既存知識の空間に射影し、文脈を構成している。その際、多くの説明変数を用いたり、説明変数間の複雑な関係を想定すると、元の文章とのずれ=情報量の損失は小さくなるかもしれないが、多くの組合せの中から最適な説明変数とその関係性を特定するのに大きな“計算量”、すなわち、処理労力を要する。一方、単純であればよいかというと、たとえば、ある文章を何らかの基準によりひとつの単語で表現するのは究極の要約であるが、組合せ計算量はないが、情報量の損失は大きい。メトカーフの法則によれば、n 個の中から2 個取ってくる組合せの数からネットワークの価値はメンバー数 n の2 乗に比例するが、組合せの計算量とともに情報量の増減も考慮することが必要であろう。今回は、これらのコスト要因を総合化した評価関数を構築するまでには至らず、以下ではパラグラフの情報量を定義し、関連性理論からの解釈を行う。

3 . 関連性理論による情報量の解釈

関連性理論における文脈効果 = 認知効果とは、読者の想定あるいは保有する旧情報を更新する新情報との相互作用であり、処理労力は新情報である未知の概念数に比例すると考えられる。そこで、文章の処理労力を最小にし、かつ文脈効果が最大になるようにするためには、当該文章上のあらゆる未知の概念をとりこむのではなく、絞ることにより労力を抑え、同時に最も文脈効果のある部分に焦点を当ててそこを中心に読めばよいことになる。

各パラグラフの文章は、第1パラグラフ～直前のパラグラフまでの旧情報と当該パラグラフにおける新情報が共起した相互作用の結果として表現されている。

新情報は、文章中の新語によってもたらされる。文脈効果上、有意なパラグラフとは新情報の追加が顕著なパラグラフであり、そうしたパラグラフを抽出することは読者の処理労力を最小にすることにつながると考えられる。各パラグラフの新語数と総単語数を調べ、情報量を次式により定義する。

$$H_i = -\log(1 - p_i)$$

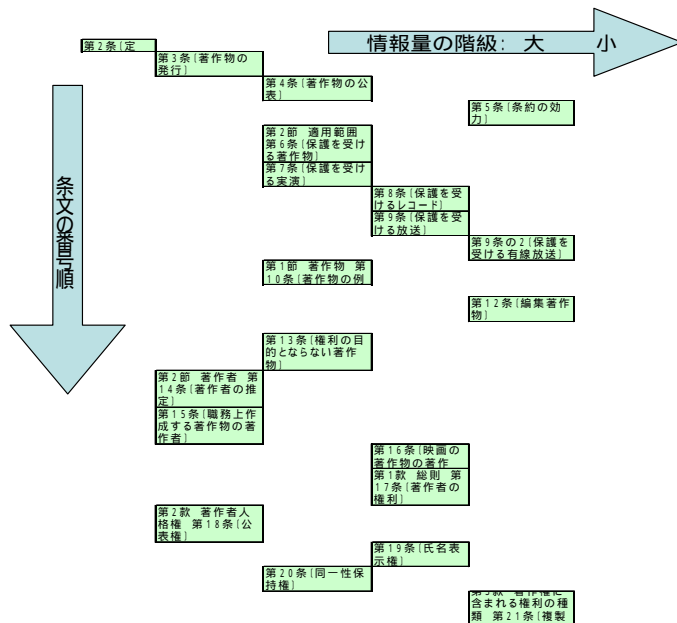
ここで、 p_i : パラグラフ i の新語率

$$p_i = \frac{\text{パラグラフ i の新語数}}{\text{パラグラフ i の総単語数}}$$

とする。

情報量の増加のみられたパラグラフだけを抽出することにより、処理労力を最小にし、しかも文脈効果のある新情報のみを得ることができよう。また、情報量が単調減少にある連続するパラグラフをグループ化し、意味パラグラフ化することも可能である。情報量の変化を調べることでより文章構造の同定と可視化が可能になる。

図2 . 情報量による著作権法のマクロ的図解化の例



4 . 図解化の例

法律条文に情報量による図解化を適用した例を図2に示す。各条文が文章のパラグラフに相当し、左寄りの条項ほど情報量が大きく、それらをフォローすることで当該法律の全体像を知ることができよう。

5 . 考察並びに今後の課題

本研究では、人間の認知行為を説明する理論である関連性理論が文章の自動図解化に合理的な根拠を与えてくれるのではないかと仮説をもとに検討を行った。情報量の変化により文脈効果のある情報の抽出を行い、法律条文へ適用した。今回はマクロ的な図解化を行ったが、文レベルのミクロ的図解化では、主語 + 目的語 + 述語といった統語構造から図解化することが考えられ、関連性理論による解釈を試みたい。さらに、処理労力と情報量を総合化した評価関数を構成することも今後の課題である。

参考文献

[Dan Sperber 1999] Dan Sperber and Deirdre Wilson : *Relevance: Communication and Cognition*(2nd ed.), Blackwell (内田聖二他訳: 『関連性理論 - 伝達と認知 - 第2版』研究社, 1999)