

修 士 論 文

題 目 評価構造における単語間の関係
性可視化に関する研究

指導教員 小山田 耕二 教授

京都大学大学院 工学研究科 電気工学専攻

氏 名 小澤 啓太

平成 28 年 2 月 4 日

目 次

第 1 章 関連研究	1
1.1 評価構造と評価グリッド法	1
1.2 テキストデータ分析	4
1.3 重複阻止	6
参 考 文 献	9

第1章 関連研究

本研究では、評価構造の文章の頻出語や関係性を可視化する。本章では、評価構造と評価構造を作成する評価グリッド法に関する研究、テキストデータ内の頻出語や関係性の可視化に関する研究、評価構造を Word Cloud で可視化する際に使用する単語の重複阻止法に関する研究に分けて関連研究を述べる。

1.1 評価構造と評価グリッド法

評価グリッド法とは半構造化インタビュー手法であり、この手法によって引き出された価値判断のネットワークが評価構造である。^{6),7)} 半構造化インタビューとは、質問の流れがある程度決まっているインタビューで、質問の流れが完全に固定された構造化インタビューと、一般的な自由形式の非構造化インタビューとの中間にあたるインタビュー手法である。評価グリッド法は、デプスインタビューでもあり、基本的に回答者とインタビュアーによる1対1の面談式インタビューを行い、回答者が自覚していない深層心理を聞く事ができる。

評価グリッド法は、レパートリー・グリッド法¹⁰⁾と呼ばれる臨床心理学の手法を元にしており、提案された当初はレパートリー・グリッド発展手法と呼ばれていた。レパートリー・グリッド法は各人が固有に持つ認知構造を抽出するために開発された面接調査手法である。レパートリー・グリッド法は、被験者の認知構造の全体を抽出するためインタビューの時間が長く被験者の負担が大きいという問題が挙げられていた。そのため、評価グリッド法は認知構造のうち評価に関与する評価構造のみの抽出を行なうことで、問題とされていたインタビューを効率化した。また、評価構造には評価に関与する要素である評価項目が存在し、この評価項目の因果関係を明らかにするためにラダーリング法を取り入れることで、評価構造を階層ネットワーク構造として表現可能にした。

評価グリッド法は、人々が持つ認知構造のうち、ものごとへの評価に関する部分、すなわち人間がある対象をどのようなプロセスで理解し、評価しているかと考えるかを効率的に引き出すことを目的として開発された。評価グリッド法は、環境心理学分野で讃井らによって開発された手法であり、評価グリッド法の利用は

環境心理学だけに留まらず、マーケティング調査や感性工学、魅力工学へと適用分野を広げてきた。マーケティング調査においては、評価グリッド法は商品企画七つ道具の一部として知られており、商品企画における顧客の潜在ニーズを調査するために実施される。感性工学及び魅力工学においては、評価グリッド法は製品の魅力を評価するための手法として位置づけられている。

評価グリッド法によるインタビュー手順はオリジナル評価項目の抽出とラダーリングの繰り返しである。オリジナル評価項目とは、インタビューの起点となる評価項目であり、ラダーリングとは、オリジナル評価項目からより抽象的な評価項目とより具体的な評価項目を引き出すための手順である。インタビューにあたって、いくつかの調査対象を刺激要素として複数個準備しておく。オリジナル評価項目の抽出では、刺激要素の中から2つを回答者に提示し、どちらのほうが好ましいかを選択してもらう。そして、何故選択した要素のほうが好ましいと思ったのか理由を尋ね、回答された理由をオリジナル評価項目として記録し、その回答をオリジナル評価項目としてグラフに描く。理由が複数個ある場合は、それらを全て記録する。刺激要素が3個以上存在する場合は、インタビューの時間短縮のため、回答者に刺激要素の好ましい順位をつけてもらい、順位が隣り合う刺激要素を提示し、オリジナル評価項目の抽出を行う場合がある。ラダーリングでは、オリジナル評価項目からより抽象的な上位概念を引き出すラダーアップと、より具体的な下位概念を引き出すラダーダウンを行う。ラダーアップでは、オリジナル評価項目としてあげられた理由「○○」について、「○○だと何故いいのですか」と質問を行い、回答された理由を上位概念として記録する。さらに、回答された上位概念についてラダーアップを行い、それ以上の上位概念が引き出せなく成るまで続ける。ラダーダウンではラダーアップとは逆に、オリジナル評価項目としてあげられた理由「△△」について「具体的にどういうところが△△なのですか」と質問を行い、回答された理由を下位概念として記録する。そして、ラダーアップと同様に、引き出された下位概念についてさらにラダーダウンを行い、それ以上の下位概念が引き出せなく成るまで繰り返す。途中で理由が複数個挙げられた場合は、上位概念または下位概念を枝分かれさせつつ上述の手順を繰り返す。ラダーリングで抽出したオリジナル評価項目の上位項目と下位項目はグラフにマッピングする。刺激要素全てのペアに対してオリジナル評価項目の抽出とラダーリングを行えばインタビューを終了する。以上の手法を取ることで、評価構造はある概念に対する評価を明らかにすることができるとされている。

評価グリッド法によって引き出された価値判断のネットワークを評価構造と呼ぶ。価値判断のネットワークとは、すなわち、人がある対象を評価する時に、どのような価値観を重視しているか、ある要素が満たされたときどのような価値観が満たされるか、ある価値観を満たすためにはどのような要素が必要であるかといった価値判断の接続関係である。評価グリッド法のインタビューは基本的には回答者と質問者の2人で行い、個人毎の評価構造図を作成する。その後、回答者全体の評価構造を把握する場合は、個人毎の評価構造を統合し全体の評価構造を作成する。また、評価構造中の価値判断の単位のことを評価項目と呼ぶ。評価グリッド法は、調査対象者の価値判断の全体像を把握する上で有効である。

讃井らは人々の住環境に関する評価構造を作成し、個人が持つ住環境評価の実態を明らかにした。また、評価グリッド法では個人毎の評価構造を統合し全体の評価構造を作成する場合もある。その際、評価構造内の評価項目間の因果関係について分析するために評価項目間の因果関係を定量化する分析手法が数多く提案された。尾上らは、満足度の高い学会についての因果関係を調べるため、評価項目間の因果関係モデルの構築を行った。⁸⁾ 評価グリッド法での結果からアンケート項目を設計及び実施し、アンケート結果をもとにグラフィカル連鎖モデリングを用いた因果関係モデルの構築と構造方程式モデリングによる分析を行った。他にも、アンケート結果をもとに重回帰分析やパス解析、階層分析、共分散構造分析などを行う研究も提案されている。本村らは、評価グリッド法により抽出したスケルトン構造を基にしてベイジアンネットの統計的学習により定量的なモデルを構築し、確率推論アルゴリズムの適用を行った。⁹⁾

評価グリッド法は、従来は紙ベースで実施されることが多く、分析作業などにおいて調査者の負担が大きかった。しかし、近年は評価グリッド法のインタビューと分析を支援するソフトウェアも開発され、大人数の評価構造の統合が容易になった。土田らが開発したEGM-assistは評価グリッド法支援ソフトウェアの代表例である。また、商品企画七つ道具の支援ソフトウェアであるPLANPARTERも評価グリッド法を支援している。以上のような評価グリッド法支援ソフトウェアは評価グリッド法に基づいたインタビューの効率化する。また、尾上はE-Gridという評価構造を分析するソフトウェアを提案している。以上のように評価構造の作成や統合が容易になることにより評価構造内のノード数が増え、評価構造の概観や分析が困難になった。評価構造を分析する場合ノードのラベルを読む必要があるが、限られた領域に表示する必要があるので、ノード数が増加するとノードの表

示領域が小さくなり、ラベルの視認性が低下してしまう。この問題に対して尾上らは評価項目数の削減を行った。¹¹⁾ 各評価項目のネットワーク中心性を計算し、ネットワーク中心性の低い各評価項目を表示しないことで、問題を解決した。本論文は、二つの方法から問題解決を行った。一つ目の方法は助詞や助動詞などの意味無し語を排除。二つ目の方法は評価構造内で複数回出現する単語の出現回数を一度に減少させることで、重要度が低い意味を持つ単語の領域を確保しつつ、従来より視認性の高い評価構造の可視化に成功した。

1.2 テキストデータ分析

本研究では前節で述べたように、評価構造が持つテキストデータを分析することで、頻出単語の発見、頻出語と強い関係を持つ単語の発見することを目的としている。テキストデータ分析は共起分析やテキスト要約など様々な目的のもと行われており、KH Coder* のような、テキストデータを統計的に分析するためのフリーソフトウェアも開発されている。

KH Coder のようなコンピュータを用いたテキストデータ分析は古くから数多く提案されてきた。1960 年台の後半には、既に 2 つのテキスト分析アプローチが提案されていた。一つは Dictionary-based アプローチと呼ばれ、分析者が分類基準を作成し、分析者の持つ理論や問題意識を操作化するためのアプローチである。Dictionary-based アプローチはコーディング規則を設けることで、データを絞り、様々な側面からデータを見ることができるという利点である。その反面、意図的に理論や仮説に都合の良いコーディングが作成される危険も存在する。もう一つは、Correlational アプローチと呼ばれるもので、分析者の理論仮説や問題意識により汚染されていない状態で、多変量解析などによってコンピュータにデータ分析してもらうアプローチである。近年では、この二つのアプローチを統合したアプローチが提案されている。¹²⁾ はじめに Correlational アプローチによりデータ全体を要約し、その結果を元にコーディングを作成する手順を踏むことで、分析者の持つ理論や問題意識の影響を極力受けない形でデータを分析することが可能となる。提案手法も同様に二つのアプローチからテキスト分析を行う。

Dictionary-based アプローチに関する研究は数多く行われている。テキスト内で頻出する特徴語を抽出することや、語と語の関係性を調査するために階層クラ

* <http://khc.sourceforge.net/>

スター分析や共起ネットワークを行う手法、内容が似た文書の群を探すクラスター分析など多岐にわたる。提案手法では、前章で述べたように評価構造内の頻出語や単語間の関係性の分析を行うための可視化を行うことを目的としている。そこで提案手法では、はじめに Dictionary-based アプローチとして評価構造内の頻出語や単語間の関係性を可視化し、次に Correlational アプローチとして選択した単語の共起語の強調表示などを行った。

単語の出現回数の可視化に関して、Word Cloud のように単語の出現回数と単語のフォントサイズを比例させる手法が多く提案されている。²⁰⁾ bubble chart では単語の中心を中心点とし円を描き、単語の出現回数と円の大きさを比例させている。また、出現回数順にリスト表示する手法なども活用されている。

テキストデータ内の単語の関係性の可視化に関して、ネットワーク可視化や多次元尺度構成法などが挙げられる。ネットワーク可視化は、共起した単語間に線を引く共起ネットワークなど、数多く提案されている。Riehmann らは WORDGRAPH というネットワーク図を用いたワイルドカードを含む keyword-in-context 検索結果を可視化する例文検索ツールを提案した。¹³⁾ WORDGRAPH では検索結果に合致した文章を図 ?? のように可視化する。検索結果では例文に下線を引き、各例文で共起する単語が存在する場合に、共起単語の下線をまとめることで見やすさを向上した。また、Wattenberg らも keyword-in-context 検索結果を可視化する WordTree を提案した。²²⁾ WordTree ではでは利用者はツリー構造上に可視化される検索結果を対話的に操作することで利用者は文章を探索することができる。Frank らは Phrase net を提案し、構造化されていないテキストの可視化を行った。²³⁾ Phrase net では、重複阻止する力指向レイアウトを用いて文書に出現する単語の組み合わせのパターンを可視化した。Carsten らは jigsaw という複数のテキスト分析アルゴリズムを統合し対話的に可視化を行うことができる文書探索支援ソフトウェアを提案した。²⁴⁾ jigsaw ではテキスト可視化画面など複数の画面を表示することで文書の要約を容易にした。多次元尺度構成法では、高次元属性を持つ分類対象物を低次元空間における点の布置で表現する手法であり、点間の距離で関係性を表す。Erick らは次元削減を用いた Web 検索結果の可視化を行った。¹⁶⁾ 検索結果のページのテキストデータからそれぞれ文書ベクトルを生成し、次元削減を行うことで文書ベクトルが近い検索ページが近くに配置される。同時に、クラスタリングを行うことで、検索結果に複数分野の情報がある場合に目的の分野の情報を発見しやすくなった。

ネットワーク図では、従来のようにノード数が増加する場合にエッジ数も同様に増加し、全てを表示することで視認性が低下するので、提案手法では Word Cloud で可視化した。Word Cloud による可視化は意味無し語の削減だけでなく、似た内容の評価項目の統合にもつながり、規模の大きい評価構造の場合でも視認性が低下せず、分析を行いやすくなった。評価構造内での単語の関係性については、評価構造内での単語の距離関係から距離行列を作成し、多次元尺度構成法を用いることで単語の座標を決定することで、単語間の位置関係から評価構造内での単語の関係性を読み取れるようにした。また、提案手法を用いたシステムでは利用者が単語を選択すると選択単語が出現する評価項目の上位概念、下位概念の単語との関係を可視化する機能を加えることで単語の関係性を読み取れるようにした。以上のように、Word Cloud を用いて必要最低限の情報のみを可視化することで、単語の関係性の情報を残しつつ視認性が低下することを防いだ。

1.3 重複阻止

前節では、テキストベースの可視化の一つとして、単語の位置座標を計算した Word Cloud による可視化について触れた。提案手法のように、多次元データを次元削減し、低次元空間に配置する際、複数の対象が重複することは問題とされている。このような重複問題は、ネットワーク表示でも問題とされており、重複阻止手法は古くから研究がされている。重複阻止に関する研究では、以下の美的基準のいずれかを考慮したレイアウトを生成する。

- 対象の相対的位置関係を保持する
- 対象間の距離関係を維持する
- 指定された領域からはみ出さない
- 指定領域の空白部分を減らす

重複阻止手法では大きく三種類のアプローチが提案されていた。一つ目は、均等スケーリングである。均等スケーリングでは、図形やオブジェクトを全ての方向に同じ倍率で拡大または縮小するアプローチである。²¹⁾ このアプローチでは、オブジェクトの相対的位置関係や距離関係は維持されるものの、図形やオブジェ

クトを不必要に移動させる必要があり、指定領域を大幅に拡大する必要が出てくるといった問題が発生する。

二つ目のアプローチは、力学モデルである。力学モデルでは、グラフを対象としており、グラフの頂点と辺に仮想的な力を割り当て、力学的エネルギーの低い安定状態を探すことで重複を阻止する。Xiaodi らは Force Transfer algorithm(FTA) を提案した¹⁵⁾。FTA では一定の規則に基づきオブジェクトを選択し、重複オブジェクトが存在するか調べる。重複オブジェクトを発見した場合、重複オブジェクトを水平方向か垂直方向かに移動させる力を発生させる。移動する方向は重複が解除されるまでに移動する距離が短い方向を選択する。移動したオブジェクトと重複する他のオブジェクトも同距離同方向に移動する。移動が終わったら他のオブジェクトを探し、同様の操作を行う。以上の操作を繰り返す手法である。FTA では、対象間の距離関係を維持した重複阻止が達成された。Strobelt らは RWordle-C を提案した。¹⁴⁾ RWordle-C は FTA と同様に一定の規則に基づきノードを選択し、重複ノードの有無を調べ、重複ノードを発見した場合、重複ノードを渦状に移動させる。初期位置を渦の中心とし広がるように移動させ、重複が解除される位置まで移動させる。力学モデルを用いた重複阻止では、均等スケーリングに比べ、レイアウトがコンパクトになる利点があるが、上記の力学モデルを Word Cloud に適用すると、単語の文字列の方向と重複阻止のために移動する方向の相関が強く、指定領域をはみ出すことや、指定領域内に空白部分が多く残るといった問題が発生する。そこで、提案手法では三つ目の条件付き最適化を適用した。

条件付き最適化では、重複問題を制約条件を用いた最適化計算に置き換えることで重複を阻止する。Erick らは、可視化対象の座標を変数とした目的関数を作成し座標を計算する手法を提案した。¹⁶⁾ NEIGHBORHOOD PRESERVING SNIPPET LAYOUT では、目的関数に重複阻止と単語の距離関係の維持を目的とした関数を作成することで対象間の距離関係を保持しつつ、重複の発生しない座標計算を可能とした。また、Erick らは目的関数に、制約条件を追加した最適化手法も提案している。¹⁷⁾ 可視化対象は二次元空間に配置されている点群であり、はじめに指定領域内で格子を生成し、格子内で座標点が存在する四角(細胞)だけを残し、細胞の x,y 座標、細胞の辺長を変数として最適化する。最適化計算を行う際に、相対的位置関係保持、細胞間の重複阻止、指定領域の最大利用、指定領域外への細胞のはみ出し禁止など、複数の制約条件を設けることで適切な細胞配置を可能にした。このように、条件付き最適化では複数の美的基準を制約条件に置き換えるこ

とで様々な条件を満たした重複阻止を可能とする. 提案手法では Erick らが提案したエネルギー関数を用いた重複阻止の最適化計算手法を参考にし、Word Cloud で表示する単語内の 1 文字を格子の 1 細胞と置き換え、細胞の位置の最適化計算を行った. また、Erick らの制約条件を一部修正し新制約条件を追加することで Word Cloud での単語配置に適した座標計算を可能とした.

参考文献

- 1) 奥西智哉, 炊飯米を生地に添加したパンの官能評価. 日本食品科学工学会誌, 56, 424-428, (2009).
- 2) 入江正和, 豚肉質の評価法. 日本養豚学会誌, 39, 221-254, (2002).
- 3) 来田宣幸, 赤井聡文. 野球における球速と球速感の関係. 日本認知心理学会発表論文集, 42-42, (2009).
- 4) 中前光弘, 順位法を用いた視覚評価の信頼性について: 順序尺度の解析と正規化順位法による尺度構成法. 日放技学誌, 56, 725-730, (2000).
- 5) 大山正, 瀧本誓, 岩澤秀紀. 順位法を用いた視覚評価の信頼性について: 順序尺度の解析と正規化順位法による尺度構成法. 行動計量学, 20, 55-64, (1993).
- 6) J. Sanui, Visualization of users' requirements: Introduction of Evaluation Grid Method, Proceedings of the 3rd Design and Decision Support System in Architecture and Urban Planning Conference, 365-374, (1996).
- 7) 讃井純一郎, 乾正雄. レパートリー・グリッド発展手法による住環境評価構造の抽出: 認知心理学に基づく住環境評価に関する研究 (1). 日本建築学会計画系論文報告集, 367, 15-22, (1986).
- 8) 尾上洋介, 久木元伸如, 小山田耕二. 可視化情報学会における会員満足度の因果関係分析. 可視化情報学会論文集, 34, 43-51, (2014).
- 9) 本村陽一, 金出武雄. ヒトの認知・評価構造の定量化モデリングと確率推論. 電子情報通信学会技術研究報告, 104, 25-30, (2005).
- 10) G. A. Kelly, The Psychology of Personal Constructs, 1 and 2, (1955).
- 11) Y. Onoue, N. Kukimoto, N. Sakamoto, K. Koyamada, Network Coarse-Graining for Evaluation Structures, In Proc. of International Conference on Simulation Technology, 34, 447-450, (2015).

- 12) 樋口耕一, テキスト型データの計量的分析. 理論と方法, 19, 101-115, (2004).
- 13) Riehmann. P, Gruendl. H, Potthast. M, Trenkmann. M, Stein. B, Froehlich. B, WORDGRAPH: Keyword-in-Context Visualization for NETSPEAK's Wildcard Search. IEEE transactions on visualization and computer graphics, 18.9, 1411-1423, (2012).
- 14) Strobel. H, Spicker. M, Stoffel. A, Keim. D, Deussen. O, Rolled- out Wordles: A Heuristic Method for Overlap Removal of 2D Data Representatives, Computer Graphics Forum, 31, 1135-1144, (2012).
- 15) Huang. X, Lai. W, Force-transfer: a new approach to removing overlapping nodes in graph layout, Proceedings of the 26th Australasian computer science conference, 16, 349-358, (2003).
- 16) Gomez-Nieto. E, San Roman. F, Pagliosa. P, Casaca. W, Helou. E. S, Oliveira. M. C. F, Nonato. L. G, Similarity Preserving Snippet-Based Visualization of Web Search Results, IEEE transactions on visualization and computer graphics, 20, 457-470, (2014).
- 17) Gomez-Nieto. E, Casaca. W, Motta. D, Hartmann. I, Taubin. G, Nonato. L, Dealing with Multiple Requirements in Geometric Arrangements, IEEE transactions on visualization and computer graphics, 1, (2015).
- 18) T. Kuo, K. Yamamoto, Y. Matsumoto, Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 conference on empirical methods in natural language processing, 230-237, (2004).
- 19) 尾上洋介, 評価構造のビジュアル分析に関する研究, 博士論文, 2016.
- 20) J. Steele, L. Noah, Beautiful visualization (O'Reilly Media Inc, 2010) 40-60.
- 21) E. Peter, W. Lai, Algorithms for disjoint node images, Key Centre for Software Technology, 1991.
- 22) M. Wattenberg, B. V. Fernanda, The Word Tree, an interactive visual concordance, IEEE transactions on visualization and computer graphics, 14, 1221-1228, (2008).

- 23) F. V. Ham, M. Wattenberg, B. V. Fernanda, Mapping text with phrase nets, IEEE transactions on visualization and computer graphics, 15, 11691176, (2009).
- 24) C. Gorg, J. Stasko, Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw, IEEE Transactions on Visualization and Computer Graphics, 19, 16461663, (2013).