

令和 2 年 9 月 3 日現在

機関番号：16301

研究種目：挑戦的萌芽研究

研究期間：2016～2019

課題番号：16K12539

研究課題名（和文）隠れた好みを利用した結婚支援システムの開発と評価

研究課題名（英文）a marriage support system using hidden preferences

研究代表者

河村 泰之（KAWAMURA, YASUYUKI）

愛媛大学・教育学部・准教授

研究者番号：80369967

交付決定額（研究期間全体）：（直接経費） 2,500,000円

研究成果の概要（和文）：結婚支援システムに保存されていたボランティアとユーザー間のメール117,881通を分析した。グラフ研磨という新しい手法でデータを整理し、単語間のネットワーク情報からクリークを発見することで類似度の高い単語の集まりを知ることができる。その結果、最も多かったのは実際に会うときのスケジュールや場所の調整など事務的なものであった。その他に、相手方にプラスイメージをもつ話題、仕事にまつわる話題、自己評価が低い話題などを発見し、ボランティアとのやり取りを通してユーザーが好む話題を分析した。これらの、観察できる好みを見つける際に利用した情報表現を利用して「隠れた好み」について研究した。

研究成果の学術的意義や社会的意義

ユーザー自身が故意に、もしくは無意識に、表に出さない好みがある。それを隠れた好みとして定義し研究を行ったことは意義深い。そして、アンケートの分析により、システム以外の好みと関連付ける足掛かりを作った。また、結婚支援ビッグデータ・オープンデータ活用研究会でのディスカッションを中心的にまとめ松山市へ人口減少対策の提言を作成したことで、結婚支援に関する本研究の成果の一部を社会へ還元した。

研究成果の概要（英文）：We analyzed 117,881 emails between volunteers and users stored in the marriage support system. It is possible to know a group of words with a high degree of similarity by processing the data with a new method called graph polishing and finding cliques from the network information between words. As a result, the most common topic was coordinating communication, such as deciding when and where people actually met. In addition, we found topics that had a positive image to a lover candidate, topics related to work, topics with low self-esteem, and analyzed topics that users liked through interaction with volunteers. We researched "hidden preferences" using these information expressions used in finding observable preferences.

研究分野：情報科学

キーワード：結婚支援 アルゴリズム グラフ研磨

## 様式 C-19、F-19-1、Z-19 (共通)

### 1. 研究開始当初の背景

結婚のマッチングに関する数理的なアプローチは、Gale-Shapley のグアルゴリズムが有名であり、アルゴリズムによる解決するタイプの結婚支援システムのほとんどがこれをベースにしている。古くから知られる技術で理論的にこれより優れた方法は確立していない。実際の現場ではリコメンド機能が多く提案されるが、結婚支援システムのリコメンド機能の難しさとして、オンラインショップなどのマーケティングに基づくシステムが応用できないことが挙げられる。結婚支援に即したリコメンドサービスが求められている。

### 2. 研究の目的

これまでに、我々はグラフ理論をベースに極大クリークを列挙してクラスタリングを行って類似度を考慮してリコメンドするシステムによる結婚支援サービスを実現した。このサービスの開始により、交際申し込みの成功率が 13%から 26%に倍増し、有効な手法として注目を集めた。本研究は、この画期的なシステムの運用で蓄積され続けるデータを分析し、新たな結婚支援サービスを見いだすことを目的としている。

### 3. 研究の方法

本研究では、運用中の結婚支援システムにたまっていくテキストデータを分析して、その中に潜む隠れた好みを利用する。テキストから隠れた内容を分析するには、通常、多くの経験や言語学的な知識が不可欠で高いコストがかかるが、本研究では、機械による分析で専門家の経験や知識に相当する部分を補い、パターンを見つけ出す部分をアルゴリズム的に解決する。ここで、隠れた情報を隠れたまま利用する。テキストの中に存在する情報を曖昧にしたまま分析を続けることは人間には難しい。例えば、収入にこだわる人も、もしかしたら生活の余裕や幸福感を目指しているだけかもしれない、それならば他の条件でも良いかもしれない。しかし、それが明文化されないまま似たような好みを探することは人間には難しい。本研究ではこの隠れた情報を、情報の組合せにより扱う。具体的には、テキストの分析情報と実際の人間の考えを相互に関係させる。素朴な総当たりの組合せで表現すると数が膨大となり現実的な時間で特徴量が計算できなくなるため、アルゴリズム論的なアプローチが必要である。

### 4. 研究成果

(1) 安定結婚問題で最も有名な Gale-Shapley のアルゴリズムではユーザーがすべての異性に順位を決めることで安定したマッチングが得られることが保証されているが、同順位やブラックリストを許すとマッチングできないことがあることも知られている。

実際の婚活システムではユーザーは最も良いと思う 1 人 (か数人) を決めきれない者も多く、また、事前に高順位をつけた相手同士が出会ってもカップルが成立しないことは珍しいことではない。つまり、ユーザーが自身の好みを正確に表すことは現実的には難しいことを示している。

そこで、ユーザーの直接的な意思を問うのではなく、無意識的に表れる好みについて研究した。えひめ結婚支援センターは、公共機関が行う営利目的でない結婚支援サービスの中で、最も先駆けて情報システム化をはかった内の一つである。我々の開発した独自の手法を採用しているほかに、多くのボランティアスタッフを抱えていることが特徴の一つである。ボランティアはユーザーのサポートを行うときにシステムによるメールを用いる。このメールの内容は悩み相談やスケジュール調整など広い範囲に及ぶ。本研究では、センターから提供された、2011 年 10 月から 2015 年 1 月までの 117,881 通のメールを分析することで結婚支援現場で話題とされる内容をマイニングした。

マイニングにはグラフ研磨と呼ばれる新しい技術を用いた。まず、共起ネットワークを作る。類似度の高い単語はクリークに近い構造を作るので、クリークを構成する単語を抽出し、その単語を含む文章のみでもう一度共起ネットワークを構築する。単純な実装の疑似コードは次のようになる。実際は高速化の工夫を加える。

---

#### グラフ研磨のもっとも単純な実装

---

```
1: V: 頂点集合, E: 辺集合,  $\sigma$ : 類似度下限値
2:  $E' = \phi$ ;  $V' = \phi$ ;
3: for all  $u \in V$  do
4:   for all  $v \in V$  do
5:     if  $\text{sim}(u; v) > \sigma$  then
6:        $E' = E' \cup (u, v)$ 
7:        $V' = V' \cup u$ 
8:        $V' = V' \cup v$ 
9:     end if
10:   end for
11: end for
12: return ( $V'$ ,  $E'$ )
```

---

類似度  $\text{sim}(u, v)$  には様々な尺度があり、適切なものを選ぶのが一般的である。今回は次のように定義される normalized PMI を用いた：

$$\text{sim}(u, v) = \frac{\log \frac{P(u)P(v)}{P(u, v)}}{-\log P(u, v)}$$

グラフ研磨によってできあがるクリークは類似度の高い単語の集まりで、このクリークを観察することで、ある話題に限定した単語の関連構造を観察することができる。

出現頻度が高く、意味があまり見いだせないと考えられる次の単語を除外した：

する, いる, ある, こと, れる, なる, よう, いたす, もの, せる, ため, よい, てる, やる, あと, くださる, なれる, あう, いろいろ, どちら, みたい, そう, いつ, つもり, とこ, いま, ごと, いい, 思う

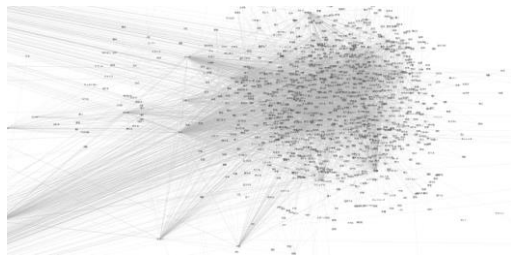


図 1：研磨前の共起ネットワーク

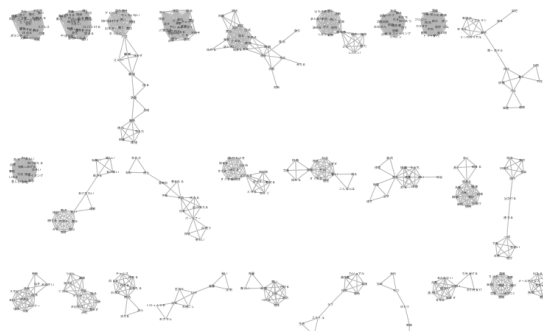


図 2：グラフ研磨を適用した共起ネットワーク

図 1 に研磨前のグラフ、図 2 に研磨後のグラフの一部を示す。サイズが 4 以上のクリークが 94 個抽出された。それらの中からいくつかを次に示す：

- { あたる, 乱文, 用件, 長文 }
- { いか, お過ごし, 元気, 区切り }
- { くれる, 折り返し, 折り返す, 留守電 }
- { スタート, 伺い, 解除, 開始 }
- { お年, お正月, お迎え, 迎える }
- { 回復, 帰れる, 段階, 気長 }
- { おく, 縮まる, 縮める, 置く }
- { ひく, 寝込む, 引く, 風邪 }
- { もてる, 好感, 持てる, 自信 }
- { いただき, 下す, 心遣い, 足労 }
- { 乏しい, 生かす, 豊富, 離婚 }
- { 助かる, 嬉しい, 有り難い, 有難い }
- { きく, 利く, 効く, 融通 }
- { パートナー, 出逢う, 探す, 新しい }
- { 充実, 有意義, 短い, 過ごせる }
- { 実家, 帰省, 年明け, 未定 }

これらから意味を見出すことは素人には難しいので、各クリークに対して、その中の1つの単語でも含んでいるような文のみで共起ネットワークを作成すると、内容が見え隠れする。図3に{すごい、人柄、優しい、明るい、真面目、素敵、誠実}のいずれかの単語を含む分から共起ネットワークを作り、これらの語とその近傍を抽出したものを示す。

「優しい」、「明るい」、「すごい」の共通接続ノードに「楽しい」、「感じ」があることが観察でき、また、「素敵」の近傍ノードに「場所」、「お話」、「笑顔」がある。次数の高いノードは納得できるものが多い。このような近傍ネットワークからいくつかの話題を特定することができた。

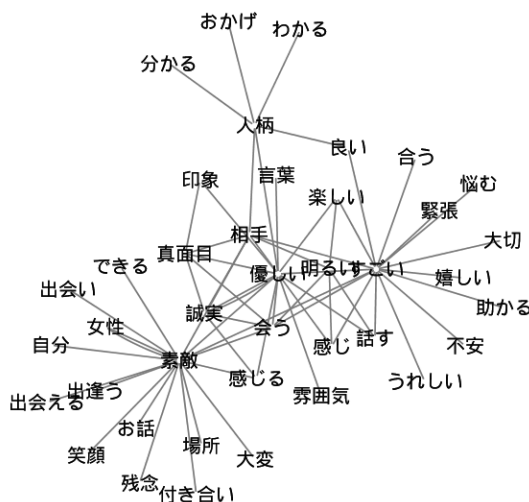


図3：{すごい、人柄、優しい、明るい、真面目、素敵、誠実}とその近傍のネットワーク

本研究では、共起ネットワークからクリークに近い構造を抽出することで意味のある情報を示していると考えられ、人間には容易に意味を見出すことのできないネットワークでもそれらを隠れた情報として扱うことで、新たなサービスへつなげられる。

隠れた情報を具体的に抽出する方法と、それを表す表現方法を提案したことは、テキストマイニングとしても価値のあることで、結婚支援システムでこの方法をとったことは世界でも先駆けた研究である。

(2) 愛媛県内でまつやま人口減少対策推進会議が行った結婚に関するアンケート（回答者604人）のデータを分析し、松山市で開かれた結婚支援ビッグデータ・オープンデータ活用研究会にて議論を行った。例えば、園田の分析を図4に示す。10～40代の既婚者（離婚死別含）の世帯における子どもの人数を収入別にみると、年収が900万円を超える世帯の割合は19.8%もあり、少ないサンプルではないが、子どもが3人以上いる世帯はなかった。400～900万円には10%前後あることとは対照的である。「少子化は収入が少ないから起こる」という論調は正しいとは言えないことを示している。

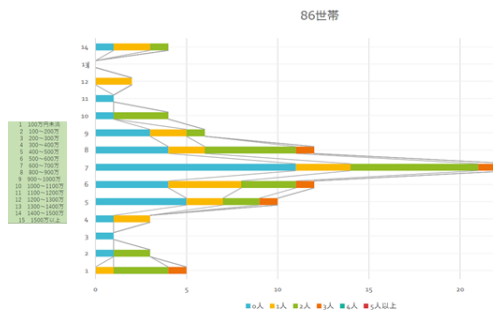


図4：世帯年収別子供人数（園田）

その他に、例えば、次の8つの選択肢から結婚に対する不安を答える調査に対して、既婚者と未婚者に分けて主成分分析を行った。

- 1 経済面、2 年齢、3 自由、4 仕事面、
- 5 健康面、6 離婚、7 親族・身内の問題
- 8 姓が変わる

主成分分析の結果を、図5に未婚者、図6に既婚者が答えたデータに対するものをプロットした。

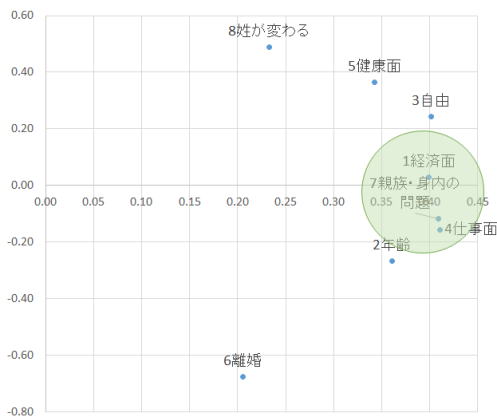


図5：未婚者の結婚に対する不安に対する主成分分析の結果

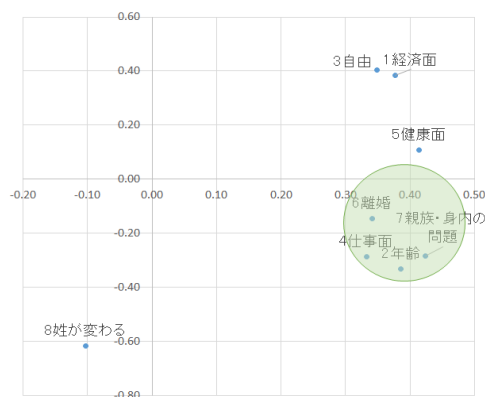


図6：既婚者の結婚に対する不安に対する主成分分析の結果

結婚に対する全般的な不安は似通っているものの、未婚者では「3 自由」と「6 離婚」が、既婚者では「1 経済面」が第2軸で大きな値を取っていることがわかる。経済的な心配は既婚者が特に気にすることであり、未婚者にとっては他の心配事とあまり変わらないように見受けられる。

この研究会でのディスカッションをもとに、申請者を中心に松山市に人口減少対策の提言をまとめた。これによって本研究によって得られた知見の一部を社会に還元できたと言える。

(1) で得られるシステムからの隠れた情報と、(2) で議論されたシステム以外から得られる情報とを関連させることができた。特にわかりやすいのは収入に関することで、システムで検索条件に必ず上位に上がってくるが、実際には、収入が高い A と相手に高収入を望む B が会って、A が B を気に入り交際を申し込んだとしても B は収入以外の何かで断ることがしばしばある。この例から分かるように、ユーザー自身が故意に、もしくは無意識に、表に出さない好みがある。それを隠れた好みとして定義し研究を行ったことは分野として意義深い。また、システム以外の好みと関連付ける足掛かりを作った。例えば、現実世界では結婚支援には金銭面のサポートをする支援がしばしば打ち出されるが、金銭はあまり効果がないようであることがわかった。では、どんなサポートやサービスが必要か、それを考えるために、今後、隠れた好みを活かされることが期待される。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 越智麻生汰、河村泰之、宇野毅明
2. 発表標題 研磨された共起ネットワークからの内容マイニング
3. 学会等名 第15回情報科学技術フォーラム
4. 発表年 2016年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

河村泰之、結婚に関するアンケートの分析、結婚支援ビッグデータ・オープンデータ活用研究会、2019年2月
---

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	宇野 毅明  (UNO Takeaki)  (00302977)	国立情報学研究所・情報学プリンシプル研究系・教授   (62615)	
研究分担者	富田 英司  (TOMIDA Eiji)  (90404011)	愛媛大学・教育学部・准教授   (16301)	