

4章 恋活・婚活のための行動履歴・ビッグデータ解析

正会員 山崎俊彦[†]

キーワード：婚活、ビックデータ、機械学習、魅力度、マッチング予測

1. まえがき

近年、各種婚活サービス（結婚相談所、婚活サイト・アプリ、恋活サイト・アプリ、婚活パーティ・イベントなど）がより身近なものとなっている。米国における調査^{*1}では、2005年から2012年に結婚した19,131組のカップルのうち、35%がインターネットなどを通じて交際を始めており、そのうち約半分がいわゆるオンライン婚活サービスを使用したものであることが報告されている。海外の航空会社機内誌を読んでいると病院、弁護士事務所と並んで広告が多いのが婚活サービス関連であることに気づく（図1）。特に、国土が広く多民族・多宗教で構成される国ではパートナー探しは容易なことではない。米国における結婚相手探しがいかに大変かを語ったTED Talk^{*2}をご覧になると、われわれ日本人からは想像できない苦労を垣間見ることができる。

それは海外でだけの話だろうと思われる方もいるかも知れないが、実は日本も例外ではない。ブライダル総研による調査^{*3}では、2015年に結婚した人のうち婚活サービスを利用したことの割合は27.6%にものぼり、実際に8.3%の人は婚活サービスで結婚相手を見つけている。筆者の身の回りでも婚活サービスを利用して結婚したカップル、今まさに婚活サービスを利用している男女が何人もいる。一昔前は「出会い系」と呼ばれネガティブなイメージが強かった感もあるが、現在の婚活サービスは真面目に結婚相手を探すための重要な選択肢の一つとなっている。

筆者は株式会社IBJ^{*4}さんとご縁を頂いてここ数年共同研究を行っており、少しずつであるが国際会議などで発表

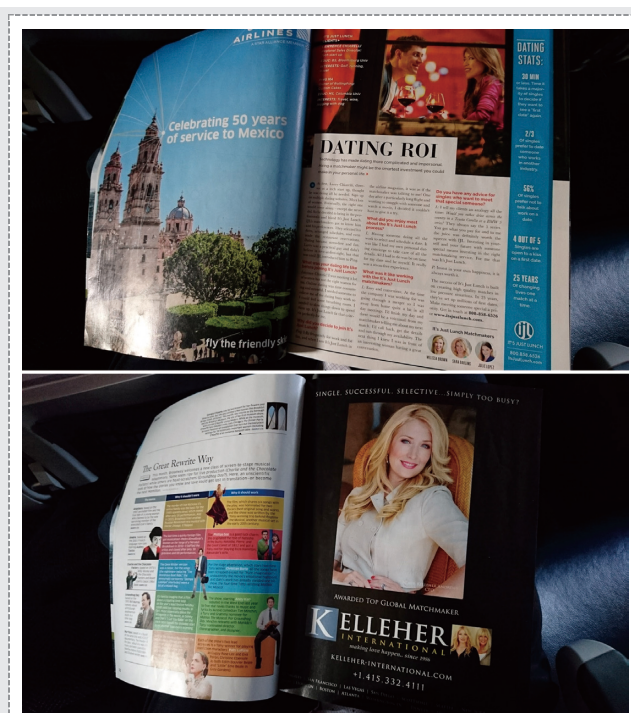


図1 筆者がUAの機内で撮影した広告

の機会を得るまでになってきた。本稿ではそのうち幾つか代表的な研究についてご紹介する。

2. 婚活研究に至った背景

人と人とのマッチング、特に「婚活」の研究をしていると必ずしも好意的な反応ばかりではなく、批判的な意見を頂戴することも少なくない。ここでは、研究に対する思いと経緯をご紹介したい。

近年、超高齢化社会に向けた医療・健康・福祉、労働人口低下を解決するためのIT・AI・ロボットなどによる労働支援、復興支援を含めた安心・安全な社会の実現などの社会的課題が議論されている。それらが重要であることは言うまでもないが、一方で個人的には置き去りにされている人たちがいるのではないかと感じている。「保育園落ちた日本死ね^{*5}」に象徴されるような、若い世代である。「若

*1 J.T. Cacioppo, S. Cacioppo, G.C. Gonzaga, E.L. Ogburn and T.J. Vander Weele: "Marital satisfaction and break-ups differ across on-line and off-line meeting venues", Proceedings of the National Academy of Sciences, 110, 25, pp.10135-10140 (2013)

*2 A. Webb: "How I hacked online dating", http://www.ted.com/talks/amy_webb_how_i_hacked_online_dating.html

*3 http://bridal-souken.net/research_news/konkatsu.html

*4 <http://www.ibjapan.jp/>

[†] 東京大学 大学院情報理工学系研究科 電子情報学専攻

"Big Multimedia Data Processing for Marriage Consulting" by Toshihiko Yamasaki (Department of Information and Communication Engineering, Graduate School of Information Science and Technology, the University of Tokyo, Tokyo)

*5 <http://anond.hatelabo.jp/20160215171759>

い世代が安心して結婚、出産、子育てをできるよう、自分のもつ技術で少しでもサポートしたい。そして、それが巡り巡って前述の超高齢化社会に由来する社会課題も解決できるのではないかと、そう考えて婚活^{1)~5)}、妊活⁶⁾、子育て⁷⁾の三つを目標に掲げた研究を行っている。

3. 婚活サービスの予備知識

Web上での婚活サービスの仕組みをご紹介します。多くのサービスでは、片方(ここでは男性Aさんとする)から「いいね」や「あいさつ」など名称の違いはあるものの、他方(女性Bさんとする)に対して「貴方とメッセージのやり取りがしたいです」という意思表示を行う。Bさんには「Aさんからいいね(あいさつ)が届きました」という表示がなされる。BさんはAさんのプロフィール、時にはAさんから届く短い自己PR文など参照した上、返事をするかどうか判断する。BさんがAさんに対して「いいね返し(あいさつ返し)」を行うと、初めて双方にシステムを通じたメッセージのやり取りが可能となる。逆に言うと、「いいね返し(あいさつ返し)」がなされないと、AさんとBさんは会話できないままである。

その先、実際に面会・交際・結婚に至るにはさまざまなハードルがあり、時間がかかることから研究対象とすることは難しい。そこで、「いいね(あいさつ)」や「いいね返し(あいさつ返し)」をたくさんもらえるユーザを魅力的なユーザと定義し、また、成婚に至ったかではなく、「いいね返し(あいさつ返し)」がなされことを「マッチングした」と定義することにする。もちろん、Tinderなど、お互いに「いいね」するまで誰に「いいね」されていたかわからないものや、match.comのように最初からメールのやり取りが可能なものなど、例外はある。

4. 男女会員の魅力度予測¹⁾

男女会員それぞれの魅力度予測の研究を行っている。これは「誰とマッチングできそうか」という一般に婚活研究で連想されるような内容ではなく、同性の会員たちと比較したときにどの程度異性から「いいね(あいさつ)」や「いいね返し(あいさつ返し)」を受けることができるかを示す相対的指標である。もし各ユーザの魅力度が予測できれば、丁度受験勉強における偏差値と同じように他者と比較したときの自分の「位置」を知ることができる。さらには、自分の魅力度をアップさせるポイントも知ることができる。もちろん、身長・体重や学歴、年取などは簡単に変えることはできない。しかし、例えば、趣味欄の選択や自己PR文の改善は可能である。特に、自己PR文は自由記述であることが多く、そこにあるキーワードが魅力度にどのような影響を与えるかを解析すると、非常に興味深い結果が得られる(具体的な内容は、サービスへの影響を考えて公開しない)。

そもそも「魅力度」をどのように定義すべきか、またその

「魅力度」をどのように予測したらよいか、という「鶏が先か、卵が先か」と同様の問題が発生してしまう。しかし、婚活サービスにおいては、ユーザはお相手のプロフィールだけをみて「いいね(あいさつ)」をするかしないかを判断しており、またマッチング率はユーザ間で歴然とした差がある。したがって、プロフィールを元とした何らかの「魅力度」が存在し、人間はその魅力度を判定していることは間違いない。そこで、魅力度の定義手法と魅力度の予測手法を同時に研究することにした。その結果、男性の魅力度は女性に「いいね(あいさつ)」をしたとき、女性から「いいね返し(あいさつ返し)」が来たかどうかとその女性のそもそもの返答率を考慮したスコアが男性プロフィールと相関が高いことがわかった。例えば、非常に返答率が低い女性に「いいね返し(あいさつ返し)」がもらえなくても「ダメで元々」であり魅力度が低いわけではないが、返答率が高い女性から返答がない場合は魅力度が低いと判断するわけである。一方、女性は「いいね(あいさつ)」を送ってくる男性の魅力度の平均値を用いると良いこともわかった。これは、9割以上の「いいね(あいさつ)」が男性から女性に対してなされることに起因すると考えられる。真面目な婚活では自分に不相応な相手には「いいね(あいさつ)」を送らないため、どんな男性からアプローチを受けるかが女性の魅力度の指標になるためであると考えている。魅力度の計算方法の概念図を図2に示す。

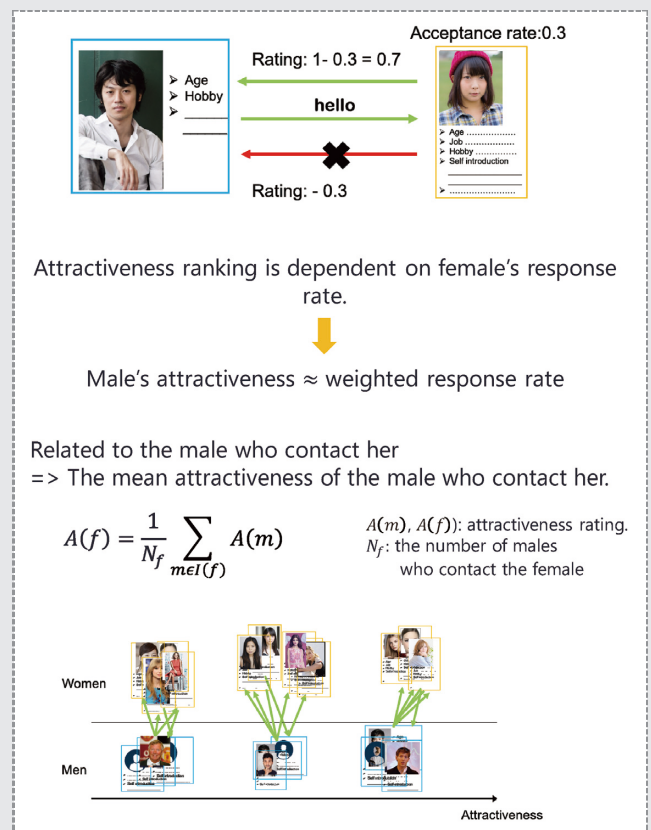


図2 男女の魅力度予測指標

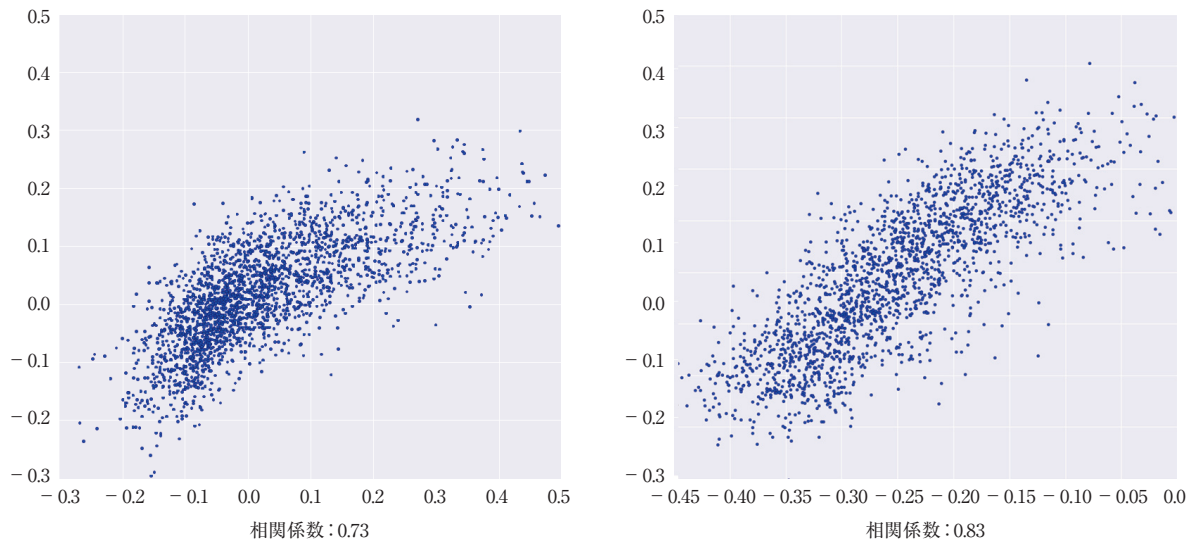


図3 男女の魅力度予測値と実際の魅力度スコアの相関
左が男性、右が女性の魅力度。

アクティビティの高い男女それぞれ1万人程度の会員に対して、プロフィール情報を用いてSupport Vector Regression (SVR) による線形回帰を行ったところ、男女それぞれの魅力度スコアと魅力度予測値の相関係数は0.73と0.83と非常に高い精度を実現している(図3)。また、線形回帰を使っているのでプロフィールのどの項目がどの程度重要度を持っているかも知ることができる。上位項目は当然のことながら年齢や年収など世の中で言われていることと同じ結果となっているが、意外と海外滞在経験の有無や園芸趣味が魅力度アップのキーワードとなっていることなどが判明した。

5. マッチング予測²⁾

「いいね(あいさつ)」はほとんどの場合男性から女性に対してなされ、その場合のマッチング率(いいね返し率)はユーザの魅力に大きく依存するものの、おおよそ5~15%とかなり低い。マッチング率が5~15%もあるのか、と好意的に取られる方もいらっしゃるが、10~20人に「いいね(あいさつ)」をしてその中のたった1人から「いいね返し(あいさつ返し)」を貰えるかどうか、という作業は非常に心の折れるものである。しかも、「いいね返し(あいさつ返し)」は単にメッセージのやり取りをする権利を得ただけで、交際や結婚に至るかどうかは別問題である。また、一定期間に「いいね(あいさつ)」できる回数が限られていたり、追加で「いいね(あいさつ)」するためポイントは有料であったりすることを考えると、マッチング率をいかに向上させるかがユーザ体験を向上させるための課題となる。

そこで、

- (1) ユーザが記入したプロフィール情報(魅力度予測に用いたものと同じ)

表1 各種法のマッチング成否予測性能

	Profile features	Topological features	Preference features	Topological+ preference features	Profile+ Topological+ preference features
MCC	0.266	0.344	0.316	0.368	0.372
Accuracy	0.682	0.775	0.721	0.760	0.763
Precision	0.237	0.310	0.268	0.306	0.310
Recall	0.713	0.675	0.736	0.753	0.756

- (2) ユーザがプロフィールを見た相手、いいね(あいさつ)をした相手、いいね返し(あいさつ返し)をした相手などの情報

- (3) 自分と同じ傾向を持っている同性群の情報、自分のことを好みと思ってくれる異性群の情報などを組合せることによって高精度にマッチングの成否を予測する手法を開発した。

アクティビティが高い会員を中心に集めた約180万件のいいね(あいさつ)行動を元にSupport Vector Machine (SVM) にて学習したところ、今回提案するすべての特徴量を用いたときにMatthews Correlation Coefficient (MCC)⁸⁾ 0.37、マッチング成功予測の適合率0.31、同適合率0.76を実現した(表1)。マッチング予測は、先述の通り成功率は非常に低いので「すべて失敗する」と予測するだけで精度としては90%程度になってしまう。MCCは成否予測精度のバランスを考慮した指標となっており、表1に示した通り提案手法が最大である。また、SVMが成功すると予測したものうち実際に成功する確率が31%、成功したすべてのケースのうち成功すると予測できたものが76%あり、人間が自律的に行動した場合に比べて、大幅な成功率向上が期待できる。以上に述べた精度は検証データに対するものであるが、

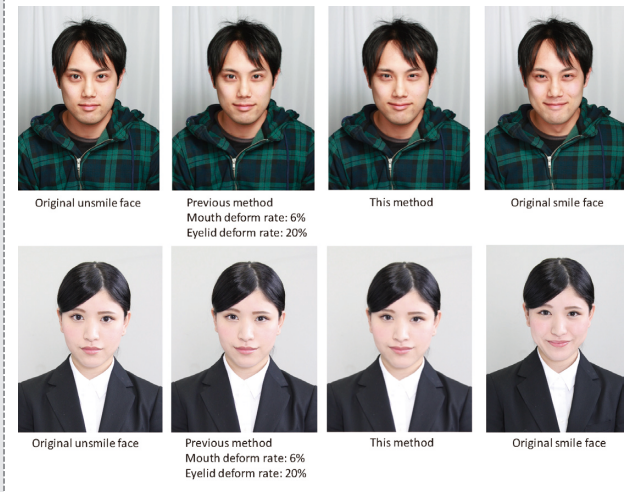


図4 データ駆動型笑顔処理の結果

一番左が真顔、一番右が本物の笑顔、中央二つが提案手法により一番左の真顔画像を笑顔にした結果。

今後実証実験を行っていきたい。

6. 会話の成否予測

Web上の婚活では、マッチングすることがゴールではない。最終的にはメッセージのやり取りではなく、実際に面会して話をしないと結婚には到達できない。そこで、メッセージのやり取りから、会うための（本人のプライベートな）連絡先交換に至るかどうかを予測する取り組みを行っている。まだ研究に着手したばかりであるが、Long Short-Term Memory (LSTM) などを用いることによって70%を超える精度で連絡先交換の成否を予測するに至っている。

7. その他の取り組み

本稿では、主に婚活を成功させるための魅力度予測、マッチング予測、連絡先交換予測について述べてきた。それ以外にも、より魅力的な顔写真を生成するための笑顔処理の試み³⁾ (図4) や、プロフィール写真と前述の魅力度の関係解析⁴⁾、深層学習を用いた顔の魅力度判定や魅力化処理⁵⁾ などに関する研究も行っている。

8. むすび

本稿ではWeb婚活における研究事例についてご紹介した。ご紹介した技術は、「これさえあれば必ず結婚できる」というようなものではない。Web上の婚活では、前述の通りマッチング率ですら必ずしも高いとはいえない現状があり、研究・開発中の技術で少しでも出会いの可能性が高まればよいと望んでいる。

婚活と一口にいってもスピードデーティングと呼ばれるようなパーティー型のものからカウンセラーがつくじゅくり型のものまでさまざまな形態がある。当然それらにも求められる必要な技術は互いに異なり、まだまだ多くの課題

が存在しており、研究の余地は大いにある。

また、婚活研究に求められるのは結局のところコミュニケーション支援や相性予測といった、人間がいればどのような状況でも必要になる技術である。これから人事、教育、スポーツコーチングなどさまざまな分野への応用展開を考えている。

本稿で紹介した婚活研究は主に株式会社IBJさんとの共同研究にて行ったものである。顔の研究については笑顔学・感性工学で著名な早稲田大学・菅原徹氏との共同研究である。本研究の一部は科学研究費助成事業(26700008)、JST-CREST(JPMJCR1686)、Microsoft IJARC Core 13、栢森情報科学振興財団などの支援を受けている。

(2017年7月28日受付)

〔文 献〕

- 1) X. Zang, T. Yamasaki, K. Aizawa, T. Nakamoto, E. Kuwabara, S. Egami and Y. Fuchida: "How competitive are you? - analysis of people's attractiveness in an online dating system", Prof. IEEE International Conference on Multimedia and Expo (ICME), pp. 973-978 (2017)
- 2) X. Zang, T. Yamasaki, K. Aizawa, T. Nakamoto, E. Kuwabara, S. Egami and Y. Fuchida: "You will succeed or not? - matching prediction in a marriage consulting service", Proc. IEEE International Conference on Multimedia Big Data (BigMM), long paper, pp.109-116 (2017)
- 3) S. Shen, T. Yamasaki, K. Aizawa and T. Sugahara: "Data-Driven Geometric Face Image Smilization Featuring Moving Least Square Based Deformation", Proc. International Workshop on Attractiveness Computing in Multimedia in conjunction with IEEE International Conference on Multimedia Big Data (BigMM-ACM), pp.220-225 (2017)
- 4) X. Zang, T. Yamasaki, K. Aizawa, T. Nakamoto, E. Kuwabara, S. Egami and Y. Fuchida: "Prediction of users' facial attractiveness on an online dating website", Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW) 2017, International Workshop on Multimedia Artworks Analysis (MMArt 2017), pp. 255-260 (2017)
- 5) S. Shen, R. Furuta, T. Yamasaki and K. Aizawa: "Fooling Neural Networks in Face Attractiveness Evaluation: Adversarial Examples with High Attractiveness Score but Low Subjective Score", Proc. IEEE International Conference on Multimedia Big Data (BigMM), short paper, pp.66-69 (2017)
- 6) 中村達介, 山崎俊彦, 相澤清晴, 石川勇介: "生活習慣アンケートを用いたk近傍探索とカーネル密度関数による妊娠までの期間予測", 映像情報メディア学会冬期大会, 13C-1 (2016)
- 7) 大淵友暉, 山崎俊彦, 相澤清晴, 島海哲史, 林幹久: "IoTカメラを用いた保育施設における子育て見守り支援", 第20回画像の認識・理解シンポジウム (2017)
- 8) B.W. Matthews: "Comparison of the predicted and observed secondary structure of T4 phage lysozyme", Biochimica et Biophysica Acta (BBA) - Protein Structure, 405, 2, pp.442-451 (1975)



やまざき しゅん 山崎 俊彦 東京大学工学系研究科電子工学専攻博士課程修了。現在、同大学情報理工学系研究科電子情報学専攻准教授。2011～2013年、米国・コーネル大学Visiting Scientist。「魅力」の予測・要因解析・増強を行う魅力工学に関する研究を精力的に行っているほか、大規模マルチメディアデータ処理、物体認識・機械学習、最適化、3次元映像処理などの研究に従事。博士(工学)。正会員。