

ユーザとの長期伴走を考慮したブライダル・ウェディング情報の

情報推薦システムの構築

A recommender system for wedding articles, which optimized for a long-term use

脇山 宗也^{*1}
Motoya Wakiyama櫻井 康平^{*2}
Kohei Sakurai山崎 吉倫^{*2}
Yoshinori Yamazaki^{*1} リクルートテクノロジーズ
Recruit Technologies Co.,Ltd.^{*2} リクルートマーケティングパートナーズ
Recruit Marketing Partners Co.,Ltd.

We study and implement a recommender system of the articles about wedding ceremony. Considering providing the information for a long-term, the algorithm is adopted the association and transfer learning. It is based on a recommender system composited by multiple rules; improve the usability for a long-term use.

1. はじめに

結婚は大きなライフイベントであり、それに伴い結婚式を挙げる人々は数多く存在するのに対し、事前に式に関する知識を十分に備えていることは稀である。そのため、結婚式を挙げたい本人がどんな式を挙げたいのか明確で無い事も多く、書籍やインターネットで情報を集めていく過程で、自身の望む挙式・披露宴のイメージを明確化していくか、ウェディングカウンターなどで相談をするなどが一般的である。これは結婚に関する意思決定項目が多く、且つ重要度が高いためである。結婚式に関する内容でも、式場の選択に始まり、指輪、ドレス、装花、演出、美容などが存在する。そのため、意思決定を下すにあたって十分な知識を集めることはごく自然である。最近では、スマートフォンなどの携帯端末上で、これらの意思決定に必要な情報を提供しているメディアも存在する。しかし、事前知識が無く、自らの嗜好を把握できていない状態でユーザが能動的に探索することは困難である。そこで、筆者らはメディア上で提供される記事を、ユーザ個人に対して、嗜好・時期に応じて記事を推薦・提供するシステムを開発した。

情報推薦の手法としては協調フィルタリングに類する手法が広く普及しているが、データが存在しない新規ユーザや新規アイテムに対して推薦を行うことができない Cold-Start 問題が存在することも広く知られている。また、情報推薦が必要なケースではユーザ数とアイテム数が多いためにデータが疎になっていることが多く、予測精度が低下する事が多い。Netflix Prize では、SVD や Matrix Factorization などを用いて次元圧縮を用いる手法が脚光を浴びた。しかし、協調フィルタリングでは、システムを運用していく上でどのような要素が推薦に強く寄与しているかを把握することは容易ではない。ユーザに記事を推薦するだけでなく、メディアとして情報を作成する側からすれば、ユーザがどのような情報を好むかを把握し、ニーズにあった情報を作成することも重要である。そのため、本論文で示すシステムでは協調フィルタリングとは異なる手法を用いて情報推薦を行っている。本システムでアソシエーションによって抽出されたルールを用いて情報推薦を行う手法を採用する。アソシエーションを用いた手法でも、データがスパースな場合でも十分な精度が得られること

が確認されている[Lin 02][伊藤 13]。また、アソシエーションでは演算の過程で抽出されるルールから、その内部構造が理解しやすい、またコンテンツベースの推薦が可能のため Cold-Start 問題を解決できる利点もある[Mobasher 07] [Lin 02]。

また、本システムではユーザが閲覧する業種(式場、指輪、ドレスなど)が変わっても、情報を横断的に共有して推薦を行う。前述したアソシエーションを用いた手法を業種毎に適用した場合、Cold-Start 問題により新規ユーザへの情報推薦は困難である。これはユーザが新たな業種の閲覧を始めた際に、その業種におけるデータが無いために演算を行うことが困難なためである。ユーザが閲覧する業種は長期に渡って推移していくが、個別に推薦を行うのでは、閲覧する業種が変わる度に 1 からデータを蓄積する事になってしまう。そこで、本システムでは業種それぞれをドメインとして、転移学習によって複数業種を横断的に扱うこととした。これにより、ユーザが閲覧する業種が変遷しても、ユーザの嗜好を反映した情報推薦を遂行し続けることが可能となる。転移学習は元のドメインと転移先の対象ドメインとの関係性によって様々なパターンが存在する[神島 10][Fernandez-Tobias 2012]。本システムで取り扱うパターンはユーザを共有しているので、ドメイン間で異なるのはアイテムのみである。既往研究では、協調フィルタリングに転移学習を適用した事例も存在するが[神島 09]、本システムで取り扱うドメインが類似しており、アイテム属性も直感的に関係性が強いものが多いため、アイテム属性に関する情報を共有することで転移学習を実施させることとした。

本論文では、転移学習を用いたアソシエーションに基づく情報推薦手法を述べる。また、提案手法は実サービスに適用しているため、ユーザビリティを上げるために、既読記事の排除、新着記事の優先、ユーザの検討フェーズに合わせたコンテンツへの絞り込みといった複数のルールを組み合わせで運用している。これらの周辺ルールについても説明する。

2. 情報推薦アルゴリズム

2.1 アルゴリズム構成

本システムは、ユーザが結婚式の準備を進めていく間、長期間に渡って必要な情報を提供していくことを目的としている。従って、EC サイトなどで行われている情報推薦とは目的が異なる。また、ニュースサイトではユーザが読む量に対して、推薦対象となるアイテムの追加ペースが高いため、基本的に新着アイテム

を対象に情報推薦を行えば良い。これに対して、本システムで対象にしているアイテムは、ニュースほどは新規に記事が増えないため、過去の記事も含めて推薦を行っていく必要がある。しかし、カスタマの趣味・嗜好にのみ焦点を当てて情報推薦を行うと、同じ記事が連日推薦することになり、ユーザの学習機会が毀損されてしまう。本システムではそのことを考慮し、既読記事、新着記事に関しては補正を行うことで情報推薦を行っている。

2.2 アソシエーションを用いた選好度の演算

ユーザがアイテムをどの程度好むかを、予測選好度で表現する。予測選好度はユーザの記事閲覧回数(page view)、及びブックマーク(Bookmark)を元にユーザ、アイテム毎に算出する。最終的な情報推薦は、アイテムの予測選好度を元に行うが、アイテムの予測選好度は事前に算出指定おいたアイテム属性に対する予測選好度を重み付きで合算した結果である。従って、本手法はアイテム属性情報を元にしたコンテンツベースのレコメンドアルゴリズムといえる。

アイテム i が持つ属性 k はカテゴリ c に分類される。例えば、属性がホテル、レストランなどの場合、カテゴリは会場タイプである。アイテム $i \in I^k$ を属性 k が紐付けられているアイテムの集合とする。この時、ユーザ u に対するアイテム属性 $k \in \mathbf{K}$ の一時予測選好度 $a_{u,k}^{tmp}$ を以下の式で求める。PageView $_{u,i}$ 、Bookmark $_{u,i}$ はそれぞれユーザ u がアイテム i に対して行った閲覧、ブックマークの回数である。

$$a_{u,k}^{tmp} = \sum_{i \in I^k} (PageView_{u,i} + Bookmark_{u,i})$$

一時予測選好度 $a_{u,k}^{tmp}$ が高いほど、ユーザ u はアイテム属性 k を好むことになる。

ただし、閲覧を開始した初期ユーザは、自身が好むアイテム属性を持つアイテムを閲覧、もしくはブックマークしているとは限らない。その理由として、ユーザがアイテムに気付いていない、もしくはユーザ自身がどういった属性のアイテムを好むのか明確化できていないなどが考えられる。そのため、アソシエーションルールを用いてユーザのアイテム属性に対する予測選好度を補正する。アソシエーションを同一カテゴリ内で実施し、その結果得られる属性 x, y 間の確信度を $confidence_{x,y}$ とする。

$$confidence_{x,y} = N(x \cap y) / N(x)$$

$N(x)$ はアイテム属性 x を持つアイテムの閲覧ユーザ数、 $N(x \cap y)$ はアイテム属性 x を持つアイテムと、アイテム属性 y を持つアイテムの両方を閲覧したユーザ数である。

この時、アソシエーションを適用する業種は単一業種内に限らない。業種に関係なく横断的にアソシエーションを適用することで、業種が異なるアイテム属性間の確信度を得ることで、ユーザが未閲覧の業種に対するスコア演算を可能とする。これを用いて、ユーザ u のアイテム属性 k に対する最終予測選好度を以下の式で算出する。

$$a_{u,k}^{fn} = \max(a_{u,k}^{tmp}, \max_{i \in \mathbf{K}_c} (confidence_{i,k} \cdot a_{u,i}^{tmp}), \max_{j \in \mathbf{K}_c} (confidence_{j,k} \cdot a_{u,j}^{tmp}))$$

最終予測選好度は 3 つの項の最大値を採用するが、2 項目と 3 項目は、他の属性の予測選好度に確信度を積算することで求めた該当属性の予測選好度である。2 項目はドメイン内での関係性、3 項目は別ドメインの関係性を表している。ユーザ u が、自身が好む属性 k を持つアイテムをよく閲覧していれば $a_{u,k}^{tmp}$ が高くなり、情報推薦では属性 k を持つアイテムが推薦されやすくなる。しかし、潜在的には属性 k を好んでいるが、ユーザが属性 k を持つアイテムに気付いていない場合は、属性 k を持つアイテムが推薦されることはない。そのため、他属性から属性 k への確信度を用いて、属性 k の予測選好度を求めるのである。

属性に対する最終予測選好度 $a_{u,k}^{fn}$ が求められたら、これを元にアイテムに対する予測選好度を求める。アイテムの予測選好度 $s_{u,i}$ は以下の式にて算出する。

$$s_{u,i} = \sum_{k \in \mathbf{K}} a_{u,k}^{fn} \sigma_{i,k}$$

$\sigma_{i,k}$ はアイテム i が属性 k を持つとき 1、持たない時は 0 を取るダミー変数である。

2.3 既読記事の優先度変更

推薦順序を考慮していない手法の場合、同じユーザに繰り返し推薦を行っても類似したアイテムが頻繁に推薦されることが多い。本システムでは、ユーザが結婚式の準備段階において長期間利用することを想定しているため、何度も同じ記事を推薦してもユーザの知識の総量は増えず、学習機会が失われてしまう。そのため、長期的にユーザに情報推薦を繰り返し提供の際は、既読記事を推薦リストから除外する方が望ましい。しかし、既読記事を単純に排除すると、ユーザが閲覧可能な記事が減ってしまい、ユーザが最終的に全記事を閲覧すると表出できる記事がなくなってしまう。そのため、本システムでは以下の方法でアイテムの予測選好度を補正する。

$$s'_{u,i} = (s_{u,i} - \min_{i \in I} (s_{u,i})) / (\max_{i \in I} (s_{u,i}) - \min_{i \in I} (s_{u,i})) - \sigma_{u,i}^r \gamma$$

$\sigma_{u,i}^r$ はユーザ u がアイテム i の既読状態を表すダミー変数である。ユーザ u がアイテム i を閲覧済みの場合は 1、そうでない場合は 0 である。

本ルールでは厳密にはアイテムを除外せず、アイテムの優先度を下がるだけである。そのため、ユーザが全アイテムを既読にしても推薦するコンテンツがなくなることは無く、また未読だから故に選好度が低い記事が表出されることを防ぐ事ができる。

2.4 新着記事の優先度変更

ユーザに提供される記事は定常状態を保っているわけではなく、ニーズやトレンドに応じて新規に記事が追加されている。ここまではユーザの行動に応じて、アイテムに対する予測選好度の算出・補正を行ってきた。しかし、ユーザの行動結果のみから予測を行うと、記事が新規に増えたとしても閲覧される機会は推薦アルゴリズム次第になってしまう。推薦アルゴリズムに委ねることも可能だが、記事自体が流行に合わせたものである場合は旬を逃してしまう。そのため、新規に追加された記事が閲覧される機会を増やすために、記事の公開日に応じて既読情報で補正後の予測選好度 $s'_{u,i}$ を補正した。

補正する値はユーザに関係なく、アイテムの公開開始日のみで行う。公開開始日と推薦実施日と比較し、実施日が公開開始当日、公開開始日から 1 週間以内などでアイテムをセグメントに分割し、セグメントごとにスコア補正を行う。

2.5 適用先のアプリケーション

リクルートが提供するスマートフォン向けアプリケーションのゼクシィにて、前述のアルゴリズムを適用した。記事の提供は、業種ごとにウェディング・ブライダル情報が掲載された記事を提供する 6 つのタブ(お金・常識、結婚式スタイル、指輪、ドレス、ビューティー、演出アイテム等)と、全業種の記事をまとめて提供する 1 つのタブ(ピックアップ)の計 7 タブにて行われる。

パーソナライズによって提供される記事は、1 日 1 回更新され、ユーザに提供される。

3. まとめ

筆者らは、結婚式に向けて長期的に情報収集を行うユーザに対して、ユーザの嗜好に沿った記事を長期にわたって繰り返し推薦を続けるシステムの構築を行った。また、構築したシステ

ムを実サービスに適用し、現在も多くのユーザにブライダル・ウェディング情報を提供し続けている。本サービスの提供により、ユーザの来訪頻度、滞在時間は伸びており一定の効果が出ている。

しかし、本システムには改善可能な箇所が複数存在する。第一に、アイテムの予測選好度はアイテム属性を元にして算出しているため、属性が同じ場合は予測選好度が等しくなってしまう。これを解決するためには、アイテム属性を介さずに選好度を求める方法が必要となる。具体的な解決手法としては、テンソル分解の手法を用いて、ユーザ、アイテム、アイテム属性からなるデータマトリクスの補完を行うことで選好度を求める方法が考えられる。

また、アイテム属性は既にアイテムに付与されているものを利用しているため、属性で表現可能な範囲には限界がある。そのため、自然言語処理を用いてアノテーションを行うことで、予測精度を上げていくことも可能である。

既読に関してはユーザが記事を読んだか否かで判定を行っているが、実際には記事のタイトルでユーザは判別しており、興味がない記事を読まないことも十分にある。そのため、記事タイトルを確認したが閲覧を行わなかったインプレッションのログを蓄積することで、ユーザが興味を弱いと推測される記事の優先度を下げる方法もユーザの利便性には大きく関わってくると考えられる。

参考文献

- [Lin 02] Weiyang Lin, Sergio A. Alvarez: Efficient Adaptive-Support Association Rule Mining for Recommender Systems, Data Mining and Knowledge Discovery, Vol.6, No.1, pp.83-105, 2002
- [伊藤 13] 伊藤寛明:アソシエーションルールを用いたアイテム推薦におけるアイテムベースとユーザベースの性能比較, 人工知能学会全国大会(第 27 会) 3E1-3, 2013
- [Mobasher 07] Bamshad Mobasher: Data mining for Web personalization. The Adaptive Web: Methods and Strategies of Web Personalization. Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York, 2007
- [Lin 02] Weiyang Lin: Association Rule Mining for Collaborative Recommender Systems. Master's Thesis, Worcester Polytechnic Institute, 2000
- [神寫 10] 神寫 敏弘: 転移学習, 人工知能学会誌, vol.25, no.4, pp.572-580, 2010
- [Fernández-Tobías 2012] Ignacio Fernández-Tobías, Iván Cantador, Marius Kaminskas, Francesco Ricci: Cross-domain recommender systems: A survey of the state of the art, In Spanish Conference on Information Retrieval, 2012
- [神寫 09] 神寫 敏弘, 赤穂 昭太郎: 転移学習を利用した集団協調フィルタリング, 人工知能学会全国大会(第 23 回)論文集 1A4-1, 2009