

ECサイトにおけるデータ駆動型意思決定のための 非線形テンソル因子分解を用いた Visual Analyticsシステムの検討

渡辺 龍二^{1,a)} 酒井 敏彦^{1,b)} 三宅 悠介^{1,c)}

1. はじめに

本研究の目的は、電子商取引 (Electronic Commerce: EC) サイト運営におけるデータ駆動型意思決定 (Data-Driven Decision Making: DDDM) を Visual Analytics (VA)[1] の枠組みで支援することである。サービス運営を通して蓄積されるデータの、意思決定への活用という課題は EC 領域において重要である [2]。しかしながら、データの膨大さや不完全さに起因して有効な DDDM の実現は決して容易ではない。本研究はこの問題に対して、情報可視化や人間とシステムとの相互作用に着目する VA というアプローチを取る。

VA とは、人間とデータ分析システムによる視覚的な相互作用を通して、人間自身によるデータ分析・仮説検証・知識発見を支援する枠組みである。システムはデータを自動分析し、その結果を視覚的インタフェースを介して人間に可視化する。人間はその情報から分析対象に対する「仮説」を立案する。その上で人間は仮説を検証するために、次の分析の対象や観点をインタフェースを介してシステムにフィードバックする。この相互作用が繰り返されるのが VA におけるデータ分析プロセスであり、仮説検証を経て「知見」が得られる。この知見の獲得を繰り返し、最終的な意思決定に役立つ「知識」を獲得するという枠組みになっている [3]。

本稿では、EC サイト運営への VA 適用の実現性を実証するための暫定的なシステムを設計・構築する。このシステムは、運営者の仮説立案を支援するためのデータ分析に焦点を当てたものである。その上でシステムの分析結果から、DDDM を効果的に支援するシステムを設計するための課題を抽出する。

2. データ分析システムの設計

本稿では EC サイト SUZURI^{*1} を題材として、データ分析システムの設計を行う。SUZURI はクリエイターがアイテムの素材となる画像をアップロードすると、T シャツやバッグ、シールといった様々な形態でアイテムを作成しユーザに販売できるサービスである。

サービスの運用を通して蓄積されるデータからユーザやアイテムについての知見を獲得することは、運営者の意思決定にとって有用である。具体的にデータとは、ユーザが閲覧したアイテムに対して取る高評価・購入といった行動のログを想定している。このログから獲得したい典型的な知見として次の 3 つを挙げる。

- 特定のユーザの嗜好や特定のアイテムの被嗜好傾向
 - 嗜好・被嗜好傾向が類似したユーザやアイテムのグループ
 - グループ間での嗜好・被嗜好傾向の共通点・相違点
- これらの知見を獲得するには反復的な仮説の立案と検証が必要であり、本稿では前者の仮説立案に焦点を当てる。

本題材における仮説立案は決して容易ではない。その理由として次の 2 つが挙げられる。1 つ目はユーザとアイテムの数の膨大さである。このとき、仮説の対象を絞り込むことが困難となる。2 つ目はログの網羅度の低さである。全アイテム数に対して 1 人のユーザの閲覧アイテム数はごく少数である。従って有効な仮説立案のためには、既存データの理解を目的としたログの抽出や集計といった分析だけでは不十分だと考えられる。

運営者による仮説立案を支援するため、データ分析システムに対する機能要件として次の 2 つを定義する。

- 観測していないユーザとアイテムのペアに対して嗜好を予測できること
- 膨大なユーザ群とアイテム群についての概観を提供した上で、その中から注目した対象についての分析・予測結果を対話的に可視化できること

これらの要件を満たすためのアプローチは次の通りである。

¹ GMO ペパボ株式会社 ペパボ研究所
Pepabo Research and Development Institute, GMO Pepabo, Inc., Tenjin, Chuo ku, Fukuoka 810-0001 Japan

a) ryujisan@pepabo.com

b) toshihiko.sakai@pepabo.com

c) miyakey@pepabo.com

*1 <https://suzuri.jp/>

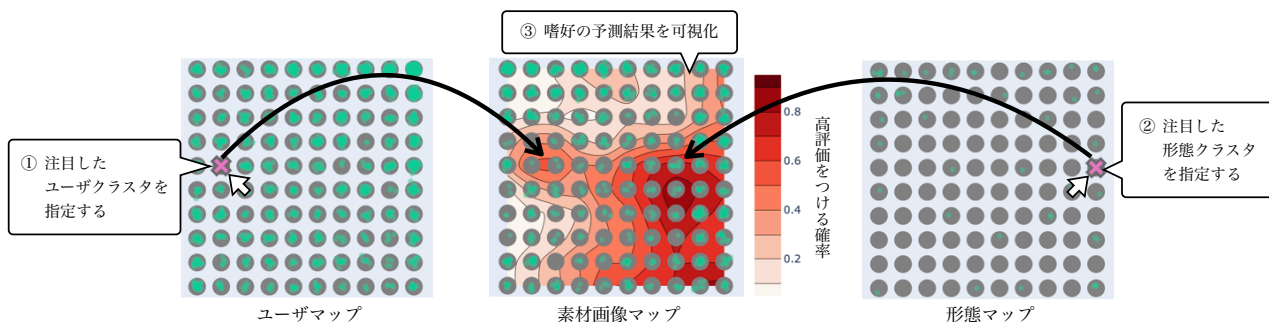


図 1: 構築したデータ分析システムの視覚的インターフェース。ユーザ・素材画像・形態という複数のオブジェクトに対してそれぞれマップを備える。個々のオブジェクト（緑の点）は有限個の等間隔に並んだクラスタ（グレーの円）に割り当てられる。運営者がユーザマップと形態マップでクラスタを指定すると、その嗜好傾向の予測結果が素材画像マップ上の赤いヒートマップとして示される。赤いほど各素材画像を元にしたアイテムに対して高評価をつける確率が高いことを示す。

テンソル因子分解による行動ログのモデル化

機能要件 (1) を満たすために、テンソル因子分解 [4] によって行動ログをモデル化する。まず行動ログ全体を $X = (x_{ijkd}) \in \mathbb{R}^{I \times J \times K \times D}$ というテンソル（多次元配列）で表現する。ただしここで I, J, K, D はそれぞれユーザ数、素材画像数、形態数、行動種別数であり、テンソルの要素 $x_{ijkd} \in \{0, 1\}$ は i 番目のユーザが、 j 番目の素材画像の k 番目の形態によるアイテムを閲覧した際に、 d 番目の行動を取ったか否かを表現している。このテンソルは実際に観測された要素にのみ値を持ち、その他の要素は欠損として扱われる。次にテンソルの生成過程を次のように仮定する。

$$\forall i, j, k \quad x_{ijk} = f(\mathbf{z}_i^{(1)}, \mathbf{z}_j^{(2)}, \mathbf{z}_k^{(3)})$$

ここで $\mathbf{x}_{ijk} = (x_{ijk1}, \dots, x_{ijkD})^\top$ である。このような潜在変数 $\mathbf{Z}^{(1)} = (\mathbf{z}_i^{(1)})^\top \in \mathbb{R}^{I \times L_1}$, $\mathbf{Z}^{(2)} = (\mathbf{z}_j^{(2)})^\top \in \mathbb{R}^{J \times L_2}$, $\mathbf{Z}^{(3)} = (\mathbf{z}_k^{(3)})^\top \in \mathbb{R}^{K \times L_3}$ と写像 $f: \mathbb{R}^{L_1} \times \mathbb{R}^{L_2} \times \mathbb{R}^{L_3} \rightarrow \mathbb{R}^D$ を観測されている要素から推定する。ここで L_1, L_2, L_3 は潜在変数の次元であり、本稿では可視化のために全て 2 とする。また今回はテンソル因子分解の非線形手法の 1 つであるテンソル自己組織化写像 (Tensor Self-Organizing Map: TSOM) [5] を用いる。

モデルを対話的に可視化する視覚的インターフェースの実装

機能要件 (2) を満たすために、TSOM を元にした視覚的インターフェースである Conditional Component Plane [5] を実装する。実装したインターフェースとその上での対話的な可視化の様子を図 1 に示す。インターフェース上ではユーザ、素材画像、形態の潜在変数集合がそれぞれ二次元マップ上で可視化され、各集合内での類似関係を概観できる。またマップ上で運営者の関心に合わせて対話的に予測結果を可視化できる。対話を繰り返すことで、前述した 3 種の知見に対する仮説を立案できる。

3. 今後の課題

データ分析システムを構築し、対話的な可視化の様子を確認した上で、今後の課題として以下を設定した。

- 予測精度の向上による可視化結果の信頼性の担保。
- 解釈可能性の向上。潜在変数の推定方法や可視化方法についての検討。
- 仮説の検証方法の確立。実際のデータの提示や追加のデータの収集による仮説検証支援についての検討。
- 運営者によるフィードバック機能の実現。マップの結果が運営者の直感や経験に反する場合に、運営者によるユーザ間やアイテム間の類似関係についてのフィードバックを可能とすること。

今後はこれらの課題を解決するためにシステムの再構築や EC サイト運営者へのレビューテストを行い、VA による運営者の DDDM 支援を引き続き目指していく。

参考文献

- [1] J J Thomas and K A Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, Vol. 26, pp. 10–13, 2006.
- [2] K Moorthi, Gaurav Dhiman, P Arulprakash, C Suresh, and K Srihari. A survey on impact of data analytics techniques in e-commerce. *Materials Today: Proceedings*, 2021.
- [3] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey Ellis, and Daniel A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, pp. 1604–1613, 2014.
- [4] Pan Zhou, Canyi Lu, Zhouchen Lin, and Chao Zhang. Tensor factorization for low-rank tensor completion. *IEEE Transactions on Image Processing*, Vol. 27, pp. 1152–1163, 2018.
- [5] Tohru Iwasaki and Tetsuo Furukawa. Tensor som and tensor gtm: Nonlinear tensor analysis by topographic mappings. *Neural Networks*, Vol. 77, pp. 107 – 125, 2016.