

協働的仮説形成システムによる EC サイト運営のデータ駆動型意思決定支援

Supporting data-driven decision making on e-commerce operation by cooperative hypothesis formation system

渡辺 龍二¹⁾ 酒井 敏彦¹⁾ 三宅 悠介¹⁾
Ryuji Watanabe Toshihiko Sakai Yusuke Miyake

1 イントロダクション

電子商取引 (E-Commerce: EC) サイト運営において、データ駆動型意思決定 (Data-Driven Decision Making: DDDM) が注目を集めている [1]. EC サイトではユーザの行動ログや商品情報といった大量のデータが蓄積されるため、データを元に意思決定の効果を高める試みである DDDM は、競争力向上のアプローチとして重要視され始めている。

DDDM では、データを抽出・加工・把握するためのデータ分析スキルが求められる。しかしながらこのスキルの獲得は運営者にとって大きな負担であり、組織における DDDM の浸透を阻む要因の 1 つとなっている [2]. したがって、運営者がデータ分析スキルに乏しい場合でも DDDM を実施できる支援が求められている。

本研究では DDDM 支援の中でも、データからの仮説獲得の効率化に焦点を当てる。意思決定に有用な知識をデータから獲得する過程のモデル [6] によると、

- A. データの探索による「仮説候補」の獲得
- B. ドメイン知識に基づく「仮説」の選出
- C. 複数の仮説の組合せによる「知識」の獲得

という関わり合う 3 つの階層的な工程をそれぞれ繰り返すことで、意思決定に有用な知識が得られる。つまり A と B の効率化は、数多くの仮説を揃えることにつながり、最終的な意思決定の質の向上が期待できる。

よって本研究では仮説獲得の効率化を狙いとした、協働的仮説形成システムを提案する。本研究は既存の DDDM 支援の方式が、効率的な仮説獲得に求められる次の 3 つの条件を全て満たしていないことに着目する。

- A-1. データ分析スキルの充足
- A-2. 任意の仮説に関連する多数の仮説候補の列挙
- B-1. ドメイン知識との連携。

工程 A で満たされるべき条件が A-1 と A-2、工程 B で満たされるべき条件が B-1 である。本研究の提案は、これら 3 つの条件を全て満たすシステムのアーキテクチャと、その実装アプローチから成る。

本稿は次のように構成されている。2 節では、効率的な仮説獲得のための条件を整理した上で、それらに対する既存方式の限界について述べる。3 節では、条件を全て満たす提案システムのアーキテクチャと、その実装アプローチについて示す。4 節では、実際の EC サイトのデータに提案システムを適用し、得られた仮説や知識について示す。加えて、実装アプローチの有効性を考察する。5 節では本稿をまとめ、今後の課題について述べる。

2 仮説獲得の効率化支援における課題

本研究で焦点を当てる効率的な仮説獲得の実現にあたっては、仮説獲得に取り組む主体が前述した 3 つの条件を満たす必要があると考えられる。条件 A-1. 「データ分析スキルの充足」は、工程 A. 「データの探索による仮説候補の獲得」を実施するために不可欠である。また、条件 A-2. 「任意の仮説に関連する多数の仮説候補の列挙」を満たせば、運営者が持つ仮説を元に工程 A を実行し、仮説となる見込みの高い仮説候補を数多く揃えることができ、効率化につながる。加えて条件 B-1. 「ドメイン知識との連携」は、工程 B. 「ドメイン知識に基づく仮説の選出」の実行に求められる。この条件は次の 2 つを指している。1 つ目は、運営者が容易に理解できる表現を仮説候補が取っていることである。2 つ目は、仮説候補に対する、運営者による仮説かどうかの判断が、工程 B の中に組み込まれていることを指す。特に 1 つ目は迅速な仮説の選出のために重要である。

しかしながら、運営者単独では満たせない条件 A-1. 「データ分析スキルの充足」を外部に求める場合、DDDM 支援の既存の方式では条件 A-2. 「任意の仮説に関連する多数の仮説候補の列挙」と、条件 B-1. 「ドメイン知識との連携」を両方満たすことができない。例えば条件 A-1. を運営者以外の人間 (分析者) に求める場合を考える。この場合、条件 B-1. 「ドメイン知識との連携」を満たす。その理由は、運営者と分析者を双方の持つドメイン知識やデータ分析結果をお互いが十分理解できるまで対話できる点にある。しかしながら、条件 A-2. 「任意の仮説に関連する多数の仮説候補の列挙」を十分に満たせない。その理由は、データ分析を人間が行うことから、列挙を何度も行うことは本質的に困難であり、獲得できる仮説候補の数が制限されるためである。

一方で、条件 A-1 をデータ分析ツールを持つ情報システムに求める場合を考える。このような情報システムの具体例としては、汎用的な Business Intelligence (BI) ツール [3] を備えたものや、ユーザの行動 [5] や取引履歴 [7] といった EC 領域特有のデータを分析・可視化するシステムが挙げられる。この場合、条件 A-2. 「任意の仮説に関連する多数の仮説候補の列挙」は満たすことが難しい。なぜなら、運営者から情報システムに仮説を伝達する方法が自明でないためである。また、条件 B-1. 「ドメイン知識との連携」についても満たすことが難しい。なぜなら運営者の理解が容易な仮説候補の表現が自明でないためである。

しかしながら情報システムの方式には、複雑なデータ分析処理の自動化という仮説の数を多数揃える観点での利点がある。これを踏まえると、効率化の支援として有望なアプローチは情報システムの方式をとった上で、従来は満たせていない条件 A-2 と B-1 を満たす、そのような情報システムの実現であると言える。

1) GMO ペパボ株式会社 ペパボ研究所

3 協働的仮説形成システム

本節では、運営者がデータ分析スキルに乏しい場合でも、効率的な仮説獲得を支援する、協働的仮説形成システムを提案する。提案システムでは、効率的な仮説獲得に必要な条件である条件 A-1. 「データ分析スキルの充足」を、情報システムが担うことで充足させる。その上で、条件 B-1. 「ドメイン知識との連携」と条件 A-2. 「任意の仮説に関連する多数の仮説候補の列挙」に対し、これを満たす運営者と情報システムの相互作用を定義する。また、この相互作用を可能とする情報システムを実装する。

3.1 項では、提案システムにおける運営者や情報システムの役割、相互作用の仕様などアーキテクチャ全体について示す。3.2 項では、情報システムを実装するアプローチについて示す。

3.1 アーキテクチャ

提案システムでは、ドメイン知識を持つ運営者とデータ分析スキルを持つ情報システムが相互作用を繰り返すことで、効率的な仮説獲得を実行する。提案システムのアーキテクチャにおいては、ドメイン知識を持つ主体とデータ分析スキルを持つ主体が分割されており、前者を Operator、後者を Analyzer と抽象化して呼ぶ。二者の相互作用によって、仮説獲得に必要な工程 A. 「データの探索による仮説候補の獲得」と、工程 B. 「ドメイン知識に基づく仮説の選出」が、効率化の条件を満たす形で実現される。また、相互作用においては、二者の間で仮説候補そのものがやりとりされる。この二者間でやりとり可能な表現を持つ仮説候補を「気づき」と呼び、Operator によってドメイン知識を元に選出されたものを「仮説相当の気づき」と呼ぶ。

工程 A. 「データの探索による仮説候補の獲得」を実現する二者の相互作用は次の通りである (図 1 上部)。Operator は手元にある仮説相当の気づきの中から、1つを選択し Analyzer に提示する。Analyzer はそれに対して、関連する気づきをデータから複数発見し、それを Operator に提示する。この相互作用は工程 A の効率化に求められる条件 A-2. 「任意の仮説に関連する多数の仮説候補の列挙」そのものとなっている。

工程 B. 「ドメイン知識に基づく仮説の選出」を実現する二者の相互作用は次の通りである (図 1 下部)。まず Operator が手元にある気づきの中から、1つを選択し Analyzer に提示する。Analyzer はその気づきに対して、データを元に定量評価を行い、その結果を Operator に提示する。これによって Operator は評価値を参考に、気づきが仮説相当か判断できる。このようにして仮説が獲得されるが、Analyzer が気づきの評価値を提示することで、Operator の気づきに対するドメイン知識に基づく理解を促し、仮説相当かどうかの判断を支援している。この支援によって、工程 B の効率化に求められる条件 B-1. 「ドメイン知識との連携」を満たしていると言える。

この工程 A と B を繰り返すことで、仮説相当の気づきが数多く揃うことになり、この中のいくつかを組合せて知識が得る工程 C が達成できれば、繰り返しが終了することになる。このように Operator と Analyzer が協働し、仮説相当の気づきが逐次的に揃えられることから、この工程 A. B の繰り返しによる仮説獲得を協働的

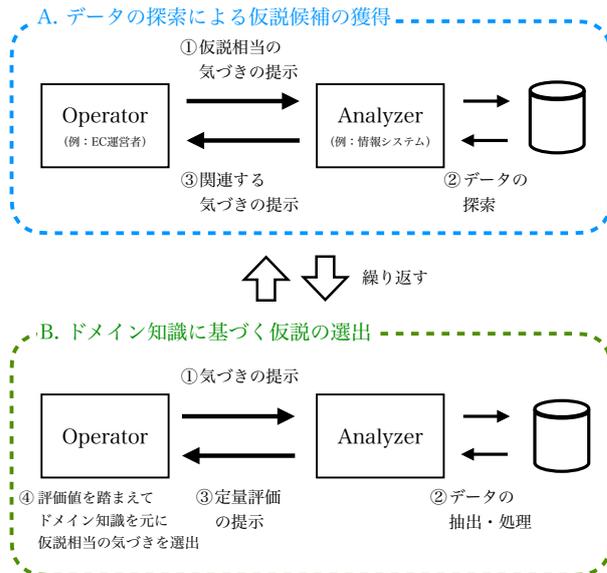


図 1 提案システムにおける「協働的仮説形成」の過程。

仮説形成と呼ぶ。

しかしこの協働的仮説形成を実現するにあたって、気づきが満たすべき要件がある。本研究では Operator を運営者が、Analyzer を情報システムが担うため、相互作用の媒介となる気づきは、運営者がドメイン知識を元に理解・判断でき、なおかつ情報システムが扱えるように表現される必要がある。

要件を満たす気づきを定義するにあたって、本研究では EC サイトにおけるユーザの商品に対する行動に注目する。ユーザの行動はログとして残るため情報システムがデータとして扱うことも可能であり、なおかつ運営者も理解でき、なおかつ「もしこの行動パターンがあり得るなら意思決定に役に立ちそうか?」といった観点で仮説相当か判断できるため、気づきに求められる要件を満たす。

ここで、ユーザの行動ログについて表記を定義する。いま EC サイトを利用するユーザの全集合を $\Omega^{(u)} = \{\omega_n^{(u)}\}_{n=1}^{N^{(u)}}$ 、EC サイトが取り扱っている商品の全集合を $\Omega^{(p)} = \{\omega_n^{(p)}\}_{n=1}^{N^{(p)}}$ とする。また、EC サイト上でユーザが商品に対して取り得る行動の全集合を $\Omega^{(a)} = \{\omega_n^{(a)}\}_{n=1}^{N^{(a)}}$ とする。ここでの行動として想定されるのは、商品を閲覧した上で行う「お気に入りにする」「注文する」といった行動である。このとき、実際にユーザが商品に対して取った行動のログを $L = \{(n_r^{(u)}, n_r^{(p)}, n_r^{(a)}, v_r)\}_{r=1}^R$ と表す。1つのログ $(n_r^{(u)}, n_r^{(p)}, n_r^{(a)}, v_r)$ は、ユーザ $\omega_{n_r^{(u)}}^{(u)}$ が商品 $\omega_{n_r^{(p)}}^{(p)}$ を閲覧した際に、行動 $\omega_{n_r^{(a)}}^{(a)}$ を取った場合は $v_r = 1$ 、取らなかった場合は $v_r = 0$ としてユーザの行動履歴を表現している。

このようなログが観測されることを踏まえた上で、本研究では気づきを、次のようなユーザ・商品・行動についての事象として定義する。

「気づき」の定義

ユーザ $\omega^{(u)} \in V^{(u)}, V^{(u)} \subset \Omega^{(u)}$ が商品 $\omega^{(p)} \in V^{(p)}, V^{(p)} \subset \Omega^{(p)}$ を閲覧したときに、行動 $\omega^{(a)}$ を取るという事象を「気づき」と呼ぶ

この事象はユーザの部分集合 $V^{(u)}$ と商品の部分集合 $V^{(p)}$ 、および行動 $\omega^{(a)}$ が定まると一意に定まるため、本稿では $S = (V^{(u)}, V^{(p)}, \omega^{(a)})$ を気づきと同一視する。

この気づきの表現は、前述した Operator である運営者と Analyzer である情報システムの両方が扱える行動ログを元に行っているため、要件を満たす。また運営者は、気づき S の構成要素である部分集合を指定する形で、自身が持つドメイン知識を元に気づきを作ることにも可能である。これは工程 A と B の反復において、工程 A に求められる仮説の初期値として利用することもできる。

このように定義された気づきにおいて、要素数の多い部分集合について述べているものは、意思決定に有用だと考えられる。気づき S は、いかなる $V^{(u)}$ と $V^{(p)}$ の要素の組合せを取ってきても行動 $\omega^{(a)}$ を取るということを表現している。そのため、 $V^{(u)}$ と $V^{(p)}$ が要素数が大きい場合、普遍的なユーザの嗜好傾向もしくは商品の被嗜好傾向を表現していると言える。そのため新規施策の決定といった意思決定に有用だと考えられる。

3.2 実装アプローチ

本項では、Analyzer として機能する情報システムを実装するアプローチについて述べる。Analyzer には、Operator と気づきについて相互作用するためのインタフェースが必要であり、本稿では視覚的な相互作用を実現するようなグラフィカルユーザインタフェース(GUI)を実装することを考える。

この GUI が実現すべき機能要件は次の3つである。

- (i) Operator による気づきの提示
- (ii) Operator から提示された気づきに対する複数の関連する気づきの提示
- (iii) Operator から提示された気づきに対する定量評価の提示

また、これらの機能要件を満たした上で、Operator と Analyzer が効率的に相互作用できる必要がある。その理由は、定義した気づきは膨大なバリエーションを取ることにある。気づきは構成要素としてユーザや商品の部分集合を取るが、この部分集合のバリエーションは組合せ爆発を起こしそれぞれ $2^{N^{(u)}}$, $2^{N^{(p)}}$ 通りになるため、気づきのバリエーションはさらに膨大になる。この気づきの性質は、Operator が気づきを指定する際や、Analyzer から提示される複数の気づきを Operator が確認する際に、効率を損なう原因となる。したがって、気づきのバリエーションの膨大さを緩和する視覚的な表現を GUI が備えていることが望ましい。

この3つの機能を実現し、なおかつ相互作用を効率化するという条件を満たす GUI を構築するため、本研究ではテンソル自己組織化写像 (TSOM) [4] によってログをモデル化し、そのモデルの対話的な可視化方法である Conditional Component Plane (CCP)[4] を実装するというアプローチを取る。

TSOM と対話的な可視化法の適用によって、図2に示すような GUI が実装できる。この GUI は大きく3つの視覚的要素によって構成されている。1つ目はユーザマップ、2つ目は商品マップ、3つ目は Operator が提示したい気づきについて行動を指定するためのラジオボタンである。Operator である運営者は、視覚的要素に対

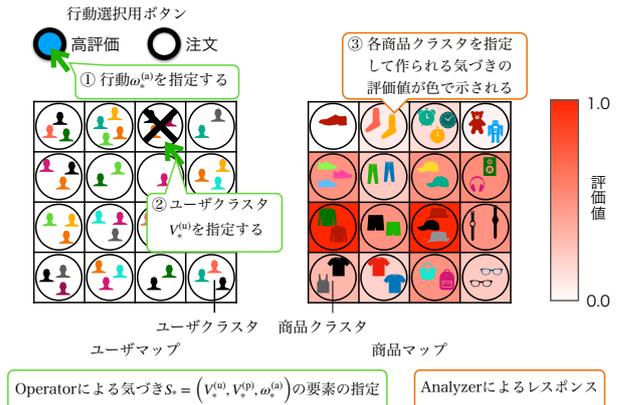


図2 実装する GUI とその上で実現される Operator と Analyzer の相互作用。

してカーソルによりクリックすることで、気づきの提示や Analyzer のレスポンスの把握ができる。

ユーザマップと商品マップは、それぞれユーザ全集合と商品全集合に対して TSOM により推定されたクラスタ構造を表現する。マップは複数のブロックに区分けされており、1つのブロックをクラスタと呼ぶ。各ユーザや商品はマップ上のクラスタのいずれか1つに配置される。マップ上で近くに配置されたユーザ同士は商品の嗜好傾向が類似しており、逆に商品同士の場合はユーザからの被嗜好傾向が類似していると言える。

図2では、3つの機能要件がどのように実現されるか示している。1つ目の機能要件である (i) Operator による気づきの提示に関しては、次のような行動選択ボタンやマップへの操作によって実現される。いま Operator が提示する気づきを $S_* = (V_*^{(u)}, V_*^{(p)}, \omega_*^{(a)})$ とすると① 行動 $\omega_*^{(a)}$ の提示は行動選択ボタンの指定、② ユーザ部分集合 $V_*^{(u)}$ の提示はユーザマップ上でのクラスタの指定、③ 商品部分集合 $V_*^{(p)}$ の提示は商品マップ上でのクラスタ $V_*^{(p)}$ を指定、となる。このように気づきを構成するユーザ部分集合や商品部分集合の指定は、マップ上に配置されたクラスタを単位として行われる。

(ii) Operator から提示された気づきに対する複数の関連する気づきの提示については、マップ上での対話的な可視化によって実現される。いま図2のように、気づきを構成する要素の一部である $\omega_*^{(a)}$ と $V_*^{(u)}$ を Operator が指定したとき、Analyzer は商品マップにおいて、 $\omega_*^{(a)}$ と $V_*^{(u)}$ に関連する複数の気づきを評価値と共に可視化する。本稿ではこの評価値として、TSOM による気づきが生じる確率の推定値を用いている。値が 1.0 に近いほど、気づきという事象が実際に生じやすいと推定されたことになる。可視化される複数の気づきとは、商品マップのクラスタの集合を $\{V_k^{(p)}\}_{k=1}^{K^{(p)}}$ とすると、 $\{S_k = (V_*^{(u)}, V_k^{(p)}, \omega_*^{(a)})\}$ のことを指す。加えてこれらの気づきの評価値を $\{g_k^{(p)}(V_*^{(u)}, \omega_*^{(a)})\}_{k=1}^{K^{(p)}}$ とすると、この評価値は対応する商品クラスタ $\{V_k^{(p)}\}_{k=1}^{K^{(p)}}$ を着色することによって示される。値が 1.0 に近いほどクラスタを赤く着色することで、Operator は提示した気づきの部分要素は固定して、残りの構成要素としてどのクラスタを指定すれば気づきとしての評価値が高まるかをひと目で把握できる。

またこの機能は (iii) Operator から提示された気づきに対する定量評価の提示も同時に実現している。ここまでの例では Operator は気づき $S_* = (V_*^{(u)}, V_*^{(p)}, \omega_*^{(a)})$ の最後の構成要素である $V_*^{(p)}$ を定めていない場合を想定していたが、これが定まっているときには、気づき S_* の評価値はマップの着色から把握することが可能である。この関連する気づきと評価値を同時に提示する Analyzer の機能は、相互作用の効率化に貢献している。

3.1 項で示した相互作用を素朴に実装すると、気づきの列挙と評価は別の操作として実現されるが、この GUI では一度の気づきの指定で可能となる。なお、図 2 の例では $\omega_*^{(a)}$ と $V_*^{(u)}$ を指定した場合を想定しているが、 $\omega_*^{(a)}$ と $V_*^{(p)}$ を指定した場合は、ユーザマップで同様の可視化が可能である。

また、クラスタ構造の導入もまた効率化に貢献している。本来、ユーザ部分集合と商品部分集合の組合せの数は $2^{N^{(u)}}, 2^{N^{(p)}}$ 通りであり、この中から運営者が、要素間の行動が類似する部分集合を発見し、各要素を GUI 上で指定することは現実的ではない。実装した GUI では、TSOM によりユーザと商品をそれぞれクラスタリングし、そのクラスタを気づきの構成要素の表現単位とする。これによって、Operator による気づきを提示するときや、Analyzer から提示される気づきを Operator が確認するときの、量的な膨大さを緩和でき、効率化につながる。

これらの機能を実現するにあたっては、ユーザと商品のクラスタと、クラスタを構成要素として取る気づきの確率の推定が必要であり、TSOM ではこれを次のように行っている。まずログデータ $L = \{(n_r^{(u)}, n_r^{(p)}, n_r^{(a)}, v_r)\}_{r=1}^R$ の r 番目のログで行動を取ったか否かを表現する $v_r \in \{0, 1\}$ を次のようなベルヌーイ試行によって観測された結果であると考える。

$$v_r \sim \text{Bern}(v_r | \mu(n_r^{(u)}, n_r^{(p)}, n_r^{(a)})) \quad (1)$$

$$\mu(n_r^{(u)}, n_r^{(p)}, n_r^{(a)}) = g_{n_r^{(a)}}(\mathbf{z}^{(u)}(n_r^{(u)}), \mathbf{z}^{(p)}(n_r^{(p)})) \quad (2)$$

ただし $\text{Bern}(v | \mu) = \mu^v(1 - \mu)^{1-v}$ であり、ユーザについて $\mathbf{z}^{(u)}(n_r^{(u)}) \triangleq \xi_{k^{(u)}(n_r^{(u)})}, k^{(u)} : \{1, \dots, N^{(u)}\} \rightarrow \{1, \dots, K^{(u)}\}, \Xi^{(u)} = \{\zeta_k^{(u)}\}_{k=1}^{K^{(u)}}, \zeta_k^{(u)} \in \mathbb{R}^2$, 商品について $\mathbf{z}^{(p)}(n_r^{(p)}) \triangleq \xi_{k^{(p)}(n_r^{(p)})}, k^{(p)} : \{1, \dots, N^{(p)}\} \rightarrow \{1, \dots, K^{(p)}\}, \Xi^{(p)} = \{\zeta_k^{(p)}\}_{k=1}^{K^{(p)}}, \zeta_k^{(p)} \in \mathbb{R}^2$, と定義する。ここで $\Xi^{(u)}, \Xi^{(p)}$ はそれぞれ、ユーザマップと商品マップ上におけるクラスタの座標であり、事前に定義されているとする。このときに、全てのユーザがどのクラスタに配置されるかまとめた写像 $k^{(u)}$ と同様に商品についての写像 $k^{(p)}$, および各行動に対応するなめらかな写像の集合 $\{g_n\}_{n=1}^{N^{(a)}}, g_n : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow [0, 1]$ をログ L から推定するのが TSOM のタスクとなる。このタスクを解くアルゴリズムについては文献 [4] を参照のこと。

4 実データによる検証

本節では、実際の EC サイトのデータを元に、Analyzer となる情報システムを構築し、提案システムを実現した結果について示す。実際に得られた仮説相当の気づきや知識について示すと共に、本稿における実装アプローチの有効性について考察する。

4.1 検証条件

今回の検証では EC サイト “SUZURI”¹⁾ のデータを利用した。SUZURI はクリエイターが商品の素材となる画像をアップロードすると、その素材画像をプリントした T シャツやバッグ、シールといった様々な形態の商品を作成でき、それを別のユーザが購入できる EC サイトである。

SUZURI のサービスを運用する中で、各ユーザがどの商品に対してどのような行動を取ったかというログが蓄積される。本稿では、ある一定期間のうち閲覧している商品の多い上位 1946 人の行動「お気に入りにする」と「注文する」についてのログを抽出した。SUZURI の商品の構造に合わせて、行動ログを改めて定式化すると $L = \{(n_r^{(u)}, n_r^{(m)}, n_r^{(t)}, n_r^{(a)}, v_r)\}_{r=1}^R$ となる。1 つのログ $(n_r^{(u)}, n_r^{(m)}, n_r^{(t)}, n_r^{(a)}, v_r)$ は、ユーザ $\omega_{n_r^{(u)}}^{(u)}$ が素材画像 $\omega_{n_r^{(m)}}^{(m)}$ ・形態 $\omega_{n_r^{(t)}}^{(t)}$ からなる商品閲覧した際に、行動 $\omega_{n_r^{(a)}}^{(a)}$ を取った場合は $v_r = 1$, 取らなかった場合は $v_r = 0$ としてユーザの行動履歴を表現している。今回の場合は行動全集合は $\Omega^{(a)} = \{\text{Like}, \text{Order}\}$ となる。また、商品の全集合 $\Omega^{(p)}$ は、素材画像全集合 $\Omega^{(m)} = \{\omega_n^{(m)}\}_{n=1}^{N^{(m)}}$ と形態の全集合 $\Omega^{(t)} = \{\omega_n^{(t)}\}_{n=1}^{N^{(t)}}$ を用いて $\Omega^{(p)} = \Omega^{(m)} \times \Omega^{(t)}$ と定義する。なお、クリエイターは素材画像に対して全ての形態で商品化するとは限らないため、実際にサイト上で販売されている商品の集合は $\Omega^{(p)}$ の部分集合となる。

このデータを元に Operator と相互作用する Analyzer の GUI として実装したものが図 3 である。構成する視覚的要素は図 2 で示した GUI と基本的には同様であるが、SUZURI の商品が持つ構造を考慮し、図 2 の商品マップを素材画像マップと形態マップの 2 つに拡張している。Operator が気づきを提示するために商品部分集合 $V_*^{(p)}$ を指定するときは、素材画像マップ上のクラスタ $V_*^{(m)}$ と形態マップ上のクラスタ $V_*^{(t)}$ の両方を指定することになる。このとき、指定された商品部分集合は $V_*^{(p)} = V_*^{(m)} \times V_*^{(t)}$ となる。これによって、どの形態の商品について気づきを構成するか、Operator が明示的に指定することができる。また、Operator が提示した気づきに対して、Analyzer が関連した気づきを提示する際も、商品マップが 2 面に拡張されたことにより、評価値が高い商品がどの形態か、把握が容易になる。なお、この商品マップの素材画像と形態への拡張を実現するために、式 (1)(2) による TSOM のモデル化も拡張している²⁾。

このように実装した GUI 上での Operator と Analyzer の相互作用の例を図 3 に示す。ここでは Operator が気づきを構成する行動 $\omega_*^{(a)}$ と商品部分集合 $V_*^{(p)}$ を提示し、Analyzer がレスポンスするという相互作用を示している。Operator によって提示された気づきの部分的な要素に対して、Analyzer は気づきに不足するユーザ部分集合をユーザマップ上のクラスタで補った時の気づきの確率を、マップ上での色付けで提示する。また、素材画像マップや形態マップにおいても同様の色付けがなされるため、その色付けを参考に Operator は次の気づきを提示できる。

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

2) 式 (1)(2) は分析対象がユーザ・商品の 2 種類の場合のモデル化であるが、今回の検証では分析対象がユーザ・素材画像・形態の 3 種類であり、モデルの拡張が必要となる。拡張されたモデルの詳細は [4] を参照のこと。

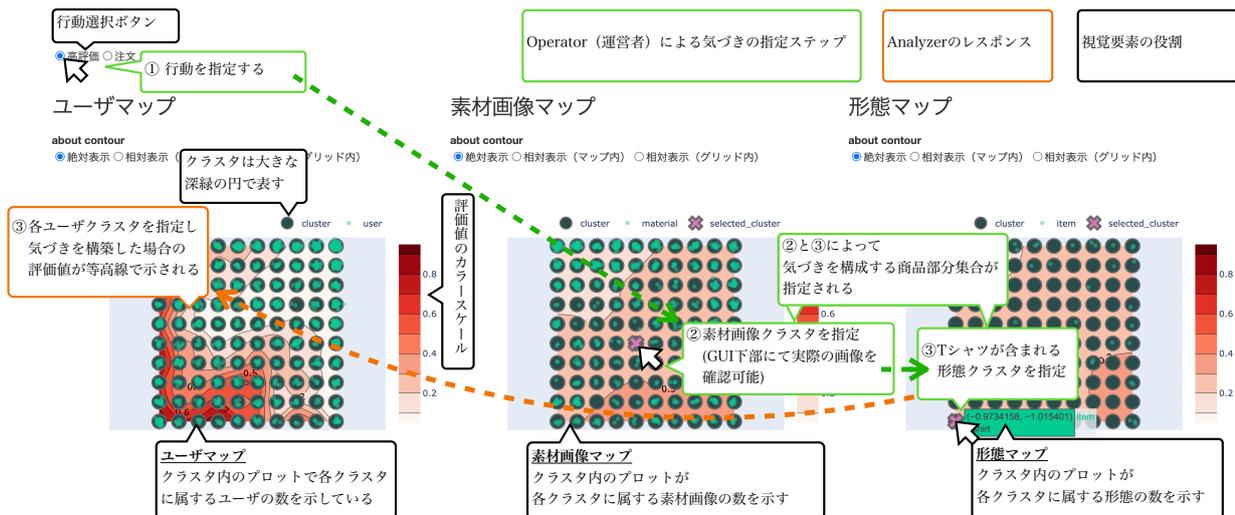


図 3 SUZURI データを元に構築した GUI. 商品マップは SUZURI の商品の構造に合わせて素材画像マップと形態マップに拡張されている.

4.2 検証結果

Operator と Analyzer が相互作用を通して、組合せて知識となる、2 つの仮説相当の気づきを得る過程の例を示したのが図 4 である. (a) は仮説相当の気づきその $1S_{*1}$ を得る過程, (b) は仮説相当の気づきその $2S_{*2}$ を得る過程である. (a) では、まず、Operator は素材画像マップで各クラスタがどのような素材画像で構成されるか確認した上で、同一のクリエイタの素材画像が集まったクラスタ $V_*^{(m)}$ を発見し、クリックする. するとユーザマップでは、 $V_*^{(m)}$ を用いた商品と、ユーザマップ上の各ユーザクラスタを気づきの構成要素として取ったときに、どれぐらいの確率が期待できるかが色付けによって表示される. これを元に、Operator は確率の高いユーザクラスタ $V_{*1}^{(u)}$ を指定する. すると、素材画像が商品になった際に、大多数の形態 $V_{*1}^{(t)}$ に対してお気に入りにすることが、形態マップから分かる. ここまでの過程をもって、気づき $S_{*1} = (V_{*1}^{(u)}, V_*^{(m)} \times V_{*1}^{(t)}, \text{Like})$ が得られたことになる. ここではこの S_{*1} を、Operator が仮説相当であると判断したとする.

次に (b) では、(a) で注目した $V_*^{(m)}$ についてさらなる気づきを得るべく、 $V_*^{(m)}$ を指定した上で (①) ユーザマップに注目し、 $V_{*1}^{(u)}$ とは対照的に、 $V_*^{(m)}$ を元にした商品にお気に入りにする見込みの低いユーザクラスタ $V_{*2}^{(u)}$ を発見し指定する (②). すると形態マップ上では、 $V_{*2}^{(u)}$ がお気に入りにする確率が可視化される. このとき、複数の形態の中で形態 $V_{*2}^{(t)} = \{\text{BigShoulderBag}\}$ が確率が高いことを発見する. これによって気づき $S_{*2} = (V_{*2}^{(u)}, V_*^{(m)} \times V_{*2}^{(t)}, \text{Like})$ が得られたことになる. ここではこの S_{*2} を、Operator が仮説相当であると判断したとする. 加えて、素材画像 $V_*^{(m)}$ が $V_{*2}^{(t)} = \{\text{BigShoulderBag}\}$ として商品化されていないことも把握したとする.

よって、仮説相当の気づきその $1S_{*1}$ とその $2S_{*2}$ を総合すると、「 $V_*^{(m)}$ を用いた BigShoulderBag を出品すれば、既存の商品ではリーチできないユーザ $V_{*2}^{(u)}$ に働きかけることができるのでは？」という予想を立てることができ、「 $V_*^{(m)}$ のクリエイタに BigShoulderBag の出品を促す」という意思決定につながる. よって、仮説相当

の気づき S_{*1}, S_{*2} の組合せによって「知識」が得られたと言える.

この結果から、提案システムは次のように意思決定を支援すると言える. Operator と Analyzer の相互作用を繰り返すことで、あるユーザ部分集合、もしくはある商品部分集合について、複数の仮説相当の気づきを得ることができる. この相互作用は仮説獲得の効率化の条件である A-1. 「データ分析スキルの充足」、A-2. 「任意の仮説に関連する多数の仮説候補の列挙」、B-1. 「ドメイン知識との連携」を全て満たすため、仮説獲得を効率的に行うことができる. 複数の仮説相当の気づきを獲得した結果として、知識につながる良い組合せを発見しやすくなる. Operator はこの良い組合せを発見するまで、Analyzer との相互作用を繰り返すことができる. そして知識が獲得されればそれを元に意思決定できる.

4.3 実装アプローチの有効性の考察

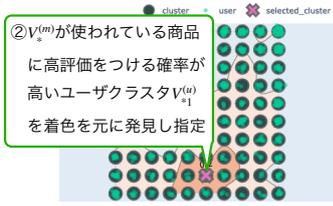
今回の GUI の利点は、気づきの構成要素である部分集合をクラスタに限定することで、膨大にある気づきの効率的な探索を可能としている点である. 部分集合のバリエーションがユーザ、商品それぞれで $2^{N^{(u)}}$, $2^{N^{(p)}}$ あることを考えると、選択可能な部分集合のバリエーションを機械学習によって中規模な $K^{(u)}, K^{(p)}$ まで削減しておくことは必要である. また今回の SUZURI の状況の場合、1 つの素材画像を元に形態の総数 $N^{(t)}$ 個の商品を出品できるため、総商品数が増加しやすく、このような対策は必要である.

しかしながら、運営者がドメイン知識を元に考えるクラスタと、TSOM で推定したクラスタが一致するとは限らないため、現状の GUI によって提示できる気づきが、本当に運営者のドメイン知識を齟齬なく反映できているかは追加検証が必要である. 運営者が持つドメイン知識をクラスタに反映させるための方法として、クラスタリングを対話的に行うことが考えられる.

また、可視化している確率の信頼性の担保も課題としてあげられる. TSOM がユーザ・素材画像・形態・行動を入力として、その行動を起こしたことを示す $v = 1$ と、起こさなかったことを示す $v = 0$ の 2 値を出力とする分類問題を解いているとみなし、モデルの学習時に

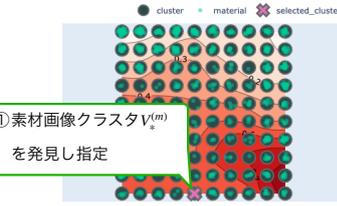
ユーザマップ

about contour
○ 絶対表示 ○ 相対表示 (マップ内) ● 相対表示 (グリッド内)



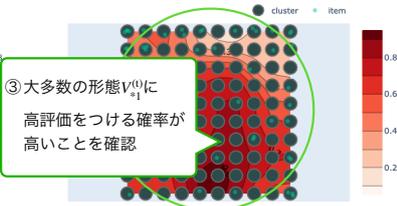
素材画像マップ

about contour
● 絶対表示 ○ 相対表示 (マップ内) ○ 相対表示 (グリッド内)

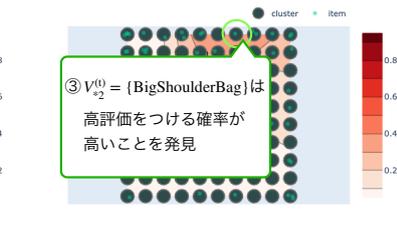
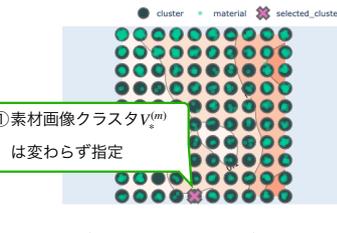


形態マップ

about contour
● 絶対表示 ○ 相対表示 (マップ内) ○ 相対表示 (グリッド内)



(a) 気づきその1 $S_{*1} = (V_{*1}^{(u)}, V_*^{(m)} \times V_{*1}^{(i)}, \text{Like})$ を得るまでのインタラクション



(b) 気づきその2 $S_{*2} = (V_{*2}^{(u)}, V_*^{(m)} \times V_{*2}^{(i)}, \text{Like})$ を得るまでのインタラクション

図4 知識となる複数の仮説相当の気づきの組合せを得る過程の例。(a) 素材画像クラス $V_*^{(m)}$ が使われている商品全般を好むユーザクラス $V_{*1}^{(u)}$ が存在するという仮説相当の気づきの獲得過程。(b) 素材画像 $V_*^{(m)}$ において、実際には出品されていない $V_{*2}^{(i)} = \{\text{BigShoulderBag}\}$ であればお気に入りにするユーザクラス $V_{*2}^{(u)}$ が存在するという仮説相当の気づきの獲得過程。この2つの気づきは「クリエイタに素材画像 $V_*^{(m)}$ で $V_{*2}^{(i)} = \{\text{BigShoulderBag}\}$ を出品することを促す」といった意思決定につながると考えられる。

は利用しなかったテストデータで評価した。結果は正例 $v = 1$ に対する precision が 0.61, recall が 0.37 程度であった。なおテストデータのうち正例が含まれる割合を踏まえると、この結果はチャンスレートより優れている。しかしながら別手法の利用など精度向上の試みが、可視化結果の信頼性担保のために必要だと考えられる。また、この確率以外にも、その推定の根拠となるデータ数といった評価指標も可視化することで、Operator の気づきに対する判断材料を増やすことも必要である。

5 まとめと今後の課題

本研究では EC サイト運営者がデータ分析スキルに乏しい場合における、データからの効率的な仮説獲得を支援するために、協働的仮説形成システムの提案を行った。提案システムはドメイン知識を持つ運営者とデータ分析スキルを持つ情報システムが相互作用しながら逐次的に仮説を獲得するアーキテクチャを持つ。また、このアーキテクチャを実現するための情報システムの実装も併せて行い、実データを用いた検証によって、知識に相当する複数の仮説の組合せや、本稿の実装アプローチの利点や欠点を明らかにした。

今後は知識につながる見込みの高い仮説や仮説の候補を獲得するために、提案システム内における気づきに対する評価指標として、今回利用したもの以外にも検討し、協働的仮説形成の有効性をさらに高める情報システムの開発を進めていく。また、既存のデータ可視化システムと比較し、提案システムのアーキテクチャの有効性・効率性を検証していく。加えて、実際に EC サイトの運営者によるテストやレビューを行い、提案システムの意思決定に対する有効性を実践的な観点で評価していく。

参考文献

- [1] Shahriar Akter and Samuel Fosso Wamba. Big data analytics in e-commerce: a systematic review and agenda for future research. *Electronic Markets*, 26:173–194, 2016.
- [2] Abhishek Behl, Pankaj Dutta, Stefan Lessmann, Yogesh K Dwivedi, and Samarjit Kar. A conceptual framework for the adoption of big data analytics by e-commerce startups: a case-based approach. *Information Systems and e-Business Management*, 17:285–318, 2019.
- [3] Jorge Bernardino and Marco Tereso. Business intelligence tools. *Intelligent Systems, Control and Automation: Science and Engineering*, 61:267–276, 2013.
- [4] Tohru Iwasaki and Tetsuo Furukawa. Tensor SOM and tensor GTM: Nonlinear tensor analysis by topographic mappings. *Neural Networks*, 77:107–125, 2016.
- [5] Shangsong Liu, Di Peng, Haotian Zhu, Xiaolin Wen, Xinyi Zhang, Zhenghao Zhou, and Min Zhu. Muluba: multi-level visual analytics of user behaviors for improving online shopping advertising. *Journal of Visualization 2021 24:6*, 24:1287–1301, 8 2021.
- [6] 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*, 20:1604–1613, 2014.
- [7] Cong Xie, Wei Chen, Xinxin Huang, Yueqi Hu, Scott Barlowe, and Jing Yang. VAET: A visual analytics approach for e-transactions time-series. *IEEE Transactions on Visualization and Computer Graphics*, 20:1743–1752, 12 2014.