

レシート特化の複合AIで購買ビッグデータを市場インサイトに変えるレシートインフォマティクス

Receipt Informatics Technology to Transform Purchase Big Data into Market Insights Using Composite AI Specialized for Digital Receipts

城田 祐介 SHIROTA Yusuke 若松 友裕 WAKAMATSU Tomohiro 金井 達徳 KANAI Tatsunori 中田 康太 NAKATA Kouta

生活者の実購買行動を捉えるビッグデータとして、電子レシートへの注目が高まっている。一方、その膨大で多様なデータから市場構造や潜在ニーズを抽出するには、従来では対応が困難であった分析に、高度AIを適用することが求められていた。特に、従来のカテゴリー体系に依存せず、“買われ方”に基づいて商品を独自に再分類して市場構造を理解することや、分析結果を迅速かつ客観的に意味付けすることが課題となっていた。

これらの課題に対応するため、東芝は、購買ビッグデータに特化した複合AI技術“レシートインフォマティクス”を開発した。嗜好(しこう)クラスタリングAIによりニーズ軸を自動抽出し、従来は得られなかった高付加価値な市場インサイト(気付き)の創出を可能にした。また、AIであるLLM(Large Language Model: 大規模言語モデル)による特徴要約との統合により、クラスター(グループ)への適切なラベル付与を自動化し、市場全体を俯瞰(ふかん)した構造理解を容易にした。

Digital receipts have recently attracted attention as big data effective for understanding consumers' purchasing behavior. To extract market structures and latent needs from large volumes of diverse data, advanced artificial intelligence (AI) applications are increasingly being introduced as alternatives to conventional evaluation methods. However, challenges remain, including difficulties in understanding market structures when products are regrouped according to purchasing behavior rather than conventional category systems, as well as in rapidly and objectively interpreting analysis results.

To address these issues, Toshiba Corporation has developed receipt informatics technology using composite AI specialized for purchase big data. This technology makes it possible to generate previously unobtainable high-value market insights by automatically extracting the need axes through preference-clustering AI. By integrating large language model (LLM)-based feature summarization to automate the appropriate cluster labeling, the technology also facilitates comprehensive understanding of overall market structures.

1. まえがき

生活者の多様な実購買行動を把握できるビッグデータとして、レシートの活用が広がっている。東芝グループが提供する電子レシートサービス“スマートレシート”は、2026年4月現在で300万人を超える利用者から日々膨大な購買データを蓄積しており、社会や企業の課題解決に活用できるユニークなデータアセットである。一方で、このような大規模かつ多様性の高い購買データから、市場構造や潜在ニーズを抽出するには、従来の“モノ”の分類カテゴリー体系に基づく集計手法では限界があり、生活者が商品を選択する際の買われ方に着目した新しい分析技術が求められていた。

東芝は、これに対応するため、購買データに特化した複合AI技術“レシートインフォマティクス”を開発した(図1)。

この技術は、まず嗜好クラスタリングAIで、生活者が商品を選ぶ際に示す共通の嗜好パターンをデータから捉え、その嗜好に基づき同じ購買軸で選択される商品群を自動的に構成する。また、嗜好クラスタリングAIで抽出した構造の特徴

要約・意味付けを行うLLMを嗜好クラスタリングAIと統合することで、購買データから高付加価値な市場インサイトを創出する。

ここでは、レシートインフォマティクスにおける複合AIの全体像と代表的な構成例を示した後、特に中核となる嗜好クラスタリングAIとLLMを組み合わせた分析手法の考え方や、その実際の分析例について述べる。

2. レシートインフォマティクスの概要

レシートインフォマティクスは、購買ビッグデータの潜在価値を最大化するため、複数のレシート特化AIを複合的に組み合わせ、構造抽出・分類・予測・結果解釈といった多様な分析処理を実現する技術である。単一のAIでは捉えきれない情報を、複数AIが補完し合うことで精度と表現力を高め、購買データを市場インサイトへと昇華することを目的としている。複合AIの組み合わせ方には、幾つかのパターンがあり、用途に応じて柔軟に構成を選択できる点が特長である。また、嗜好クラスタリングAIには、当社が長年にわたり製造・

データ分析分野で培ってきたクラスタリング技術が活用され

ており、大規模データから高精度で特徴を抽出する当社独自のノウハウが反映されている。

レシートインフォマティクスを代表する組み合わせ方としては、3章以降で記す“嗜好クラスタリングAI×LLM”がある。これは、嗜好クラスタリングAIが購買行動パターンから市場構造を抽出し、LLMがその構造に対して特徴要約やラベリングを行うものである。従来は専門家が時間を要していた仮説立案やクラスター解釈を自動化し、分析効率と解釈性が大幅に向上する。

これとは別の組み合わせ方として、同一タスクに複数AIを適用して精度を高める方式もある。図2は、国内共通の商品バーコードであるJAN (Japanese Article Number) コードが存在しない総菜や生鮮食品などのインストア商品(店舗内で加工・調理・包装される商品)を分類する例で、ここでは、同一タスクである自動商品分類に、二つのAIモデルを使用している。一つ目のAIモデルは、レシート印字の商品名の分類に特化したLightGMAIC (Light Graph-based Multi-Angled Item Categorization) グラフニューラルネットワークモデルであり、二つ目のAIモデルは、分類に最適化したレシート分類学習済みLLMである。この2種類のAIを統合する自動分類技術を適用している⁽¹⁾。単一モデルの判定に依存せず、両AIの推論結果を機械学習で統合することで、従来より高い分類精度を実現している。

更に、ほかの組み合わせ方として、構造抽出を担う嗜好クラスタリングAIと、販売反応を推定するためのLLMを組み

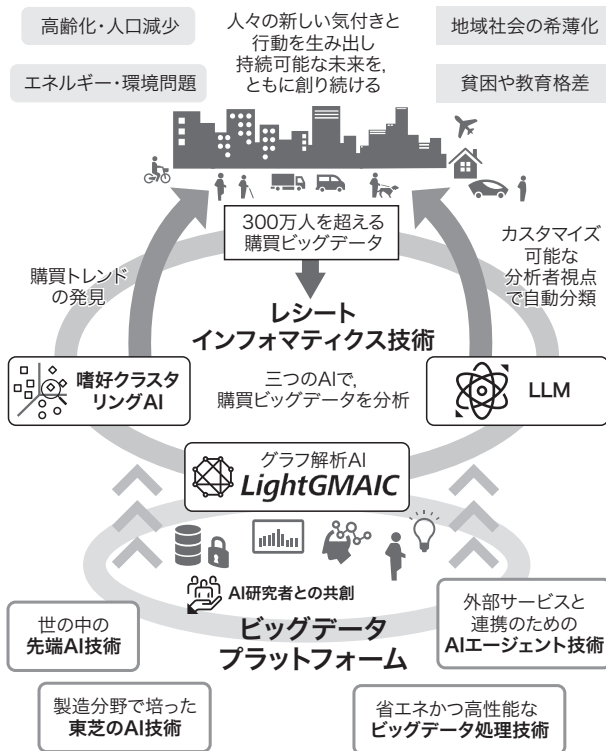


図1. レシートインフォマティクスの概要

複数の独自AIを複合的に組み合わせることで、購買ビッグデータに内在する市場構造や潜在ニーズを自動的に抽出する。

Overview of receipt informatics technology

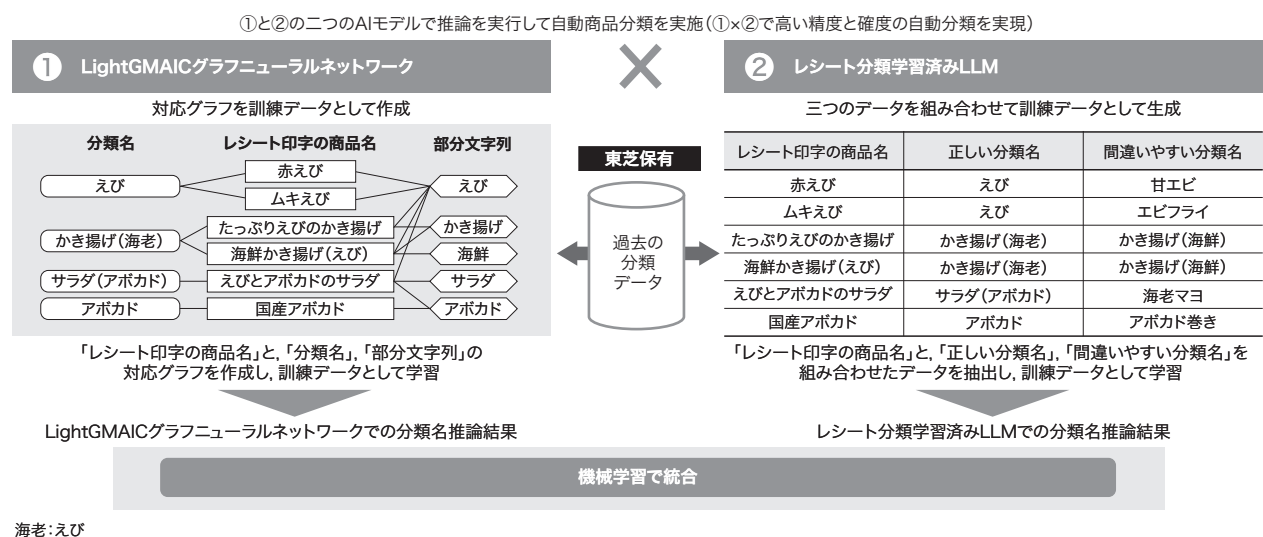


図2. 二つのAIを同一タスクであるインストア商品の自動分類に適用した場合の概要

レシート印字の商品名の特徴を捉えるLightGMAICと、レシート分類に特化したLLMのそれぞれで、新たに入力したレシート印字の商品名で分類名を推論後、最終的に機械学習で統合し、高精度の分類を実現している。

Application of two AI technologies for automated classification of in-store product names

合わせた予測型の複合AIもある。この構成では、まず嗜好クラスタリングAIによりユーザーを購買行動に基づいてクラスター分割し、そのクラスター特性を入力として、反応推定用のLLMが新商品に対する反応を推定することで、販売数量を予測する⁽²⁾。このように、嗜好クラスタリングAIと反応推定LLMという役割の異なる二つのAIを組み合わせることで、評価指標が改善し、予測値と実績値の整合性も向上したことを確認できた。これは、複合AIの枠組みが市場構造の理解にとどまらず、実務的な予測タスクにも有効であることを示した例である。

また、これらの複数AIの開発・運用には、モダンなビッグデータ分析基盤を活用し、大規模データに基づくAIの開発・運用を継続的かつ安定的に実行できる環境を整備している⁽³⁾。

3. 嗜好クラスタリングAIによる市場構造の発見

嗜好クラスタリングAIは、購買ビッグデータから生活者の共通する嗜好パターンを抽出し、同じ購買軸で選択されている商品群を自動的にグルーピングする技術である。従来のように“モノのカテゴリー”に基づいて商品を分類するのではなく、実際の購買行動に着目し、どのような基準で商品が選択されているのかという買われ方の軸をデータから直接抽出できる点に特長がある。これにより、カテゴリーを横断した商品同士の意外な関係性や、表面的な売上ランキングでは把握が難しい“隠れた購買トレンド”などを可視化することが可能となる。その二つの適用事例を述べる。

3.1 シャンプー市場におけるお試し需要の可視化

シャンプー市場は、ブランド数や商品バリエーションが多く、売上ランキングだけでは生活者の選択理由や商品間の関係性を十分に把握することが難しい。そこで、特定の仮説を設けずに、購買データに対して嗜好クラスタリングAIを適用した結果、“お試しサイズ”を重視する（お試しサイズを購入することが多い）購買軸に基づく特徴的なクラスターが抽出された⁽⁴⁾。

このクラスターに属する購買者には、通常サイズではなく、お試しサイズを繰り返し購入し、短期間に複数ブランドを比較する行動が見られた。更に、当該クラスターの購買者が最終的に購入する通常サイズ商品の一人当たりの購入価格平均値を分析した結果、お試しサイズを重視しないクラスターに属する購買者の約2倍であった（図3）。すなわち、お試しサイズを活用して複数商品を比較した上で、最終的に自分にあった高価格帯の商品に移行する購買トレンドが確認された。

この結果は、従来のカテゴリー単位の集計では捉えづかった“お試し需要”という潜在ニーズを可視化したものである。メーカーにとっては、新商品設計におけるお試しへ導く

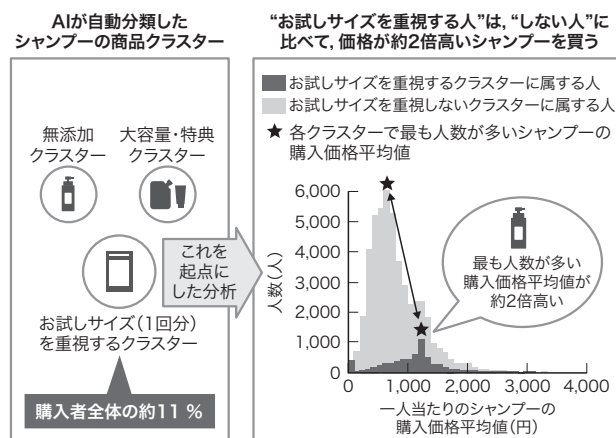


図3. シャンプーの購買データの嗜好クラスタリングAIによる分析

嗜好クラスタリングAIにより抽出された“お試しサイズを重視する”クラスターから、“お試し需要”という潜在ニーズを示唆する結果が得られた。

Analysis of shampoo purchase data using preference-clustering AI

流れの重要性や、価格帯ごとの購買行動の違いを理解する上で有用な示唆となり、販促企画や商品開発方針への活用もできる。

3.2 グルテンフリー市場の潜在規模推定

グルテンフリー食品は、国内ではニッチ市場とされてきたが、健康意識の高まりに伴い注目が高まっている。当社は、ミツカングループの株式会社ZENB JAPANと共同で、嗜好クラスタリングAIを用いたグルテンフリー食品の潜在市場規模の推定を実施した⁽⁵⁾。

嗜好クラスタリングAIを購買ビッグデータに適用することで、まず健康志向に関連する購買軸が自動抽出された。更に、その購買軸と、様々なカテゴリーの食品との関係性の強さに基づいて商品が再分類され、グルテンフリー食品と類似の購買軸で選択されている食品群を一つのクラスターとして抽出できた。

分析の結果、このクラスターにおけるグルテンフリー食品の購買軸は、糖質オフ、たんぱく質摂取オン、食物繊維摂取オンの三つの健康系ニーズが中心であることが明らかになった（図4）。

こうして抽出されたクラスターに含まれる三つの購買軸と、それらのニーズと共通性を持つ非グルテンフリーの類似商品群の購買データを基に、既存の様々なカテゴリーの市場からどの程度取り込めるかを推定することで、グルテンフリー食品の潜在市場規模を定量的に予測することが可能となる。

4. 嗜好クラスタリングAIとLLMを組み合わせた商品クラスターの自動ラベリング

嗜好クラスタリングAIが抽出した市場構造は、そのままでも

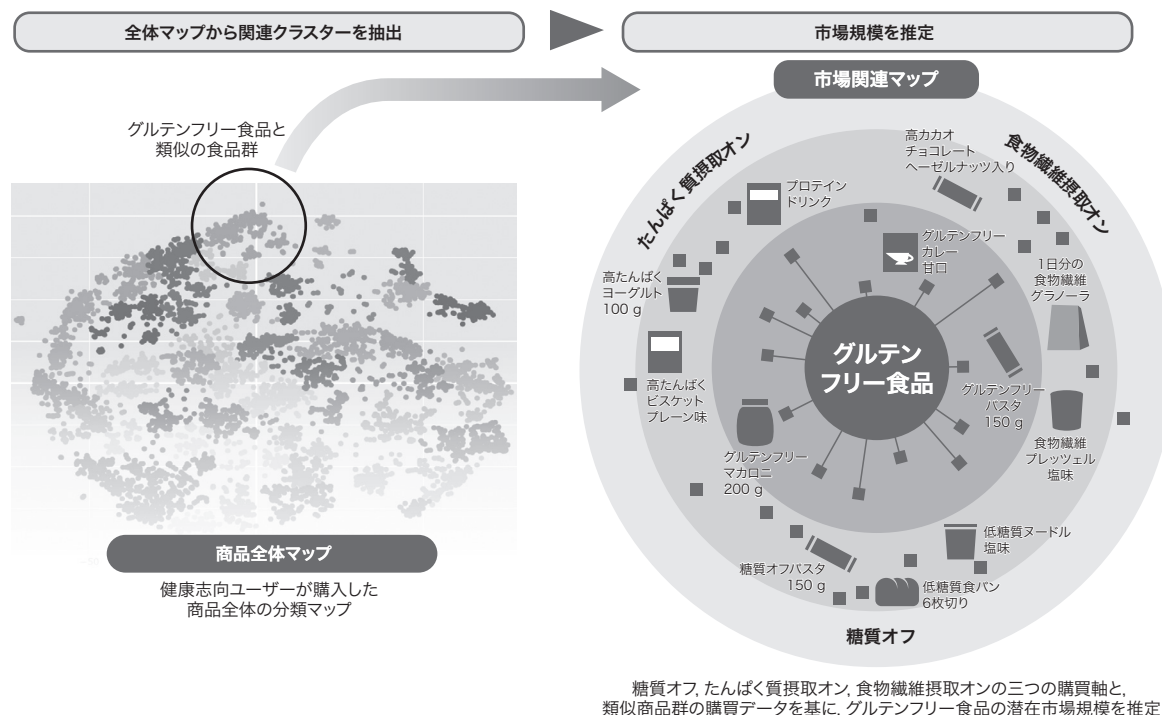


図4. 嗜好クラスタリングAIで分析したグルテンフリー食品の潜在市場

嗜好クラスタリングAIにより、潜在市場予測に必要となる、グルテンフリー食品に関連する三つの購買軸と、これらの軸と共通性を持つ類似食品群が自動抽出される。

Estimation of potential market size for gluten-free foods using preference-clustering AI

は専門家でなければ解釈が難しい場合がある。そこで、クラスターの特徴量をLLMが要約し、名称や象徴的な特徴を自動生成する仕組みを備えている。これにより、クラスターの名称と特徴とが明確になり、市場をニーズ軸で俯瞰しやすいインサイトとして提供できるようになった。

図5は、お菓子市場の購買データに嗜好クラスタリングAIを適用して抽出された20個のクラスターについて、LLMによって自動ラベリングを行った結果を示している。各クラスターは、購買行動に基づくニーズ軸で整理され、それぞれのクラスター名と特徴が簡潔に表現されている。

この自動ラベリングにより、従来は担当者がクラスターを手作業で読み解いていた工程が大幅に短縮され、分析結果の標準化も進んだ。また、商品の買われ方に基づいて整理された20個のクラスターにより、市場全体の構造やトレンドを誰でも理解しやすくなる。メーカーでは、これらの市場インサイトを活用し、成長余地の大きい市場セグメントの特定や商品ポートフォリオの再構築など、様々な意思決定に役立てている。

5. あとがき

ここでは、購買ビッグデータの価値を最大化するために開発した複合AI技術であるレシートインフォマティクスの構成と、実購買データに基づく市場分析例を述べた。嗜好クラス

LLMによる自動ラベリングの結果

お菓子市場の購買データに嗜好クラスタリングAIを
適用して抽出された20個のクラスター

1. スパイシー&刺激系クラスター	11. 定番スナック王道クラスター
2. ファミリーパック系クラスター	12. パラエティー中容量クラスター
3. あっさり系定番クラスター	13. 大容量ファミリークラスター
4. 駄菓子・低価格クラスター	14. プレミアム少量クラスター
5. 定番ポテトスナック系クラスター	15. スティック/筒形クラスター
6. プレミアム・限定クラスター	16. 小容量お試しクラスター
7. ハード食感・濃厚系クラスター	17. プチサイズ統一クラスター
8. 大容量・シェア系クラスター	18. ポテチ基本形クラスター
9. ベーシック・大衆系クラスター	19. こだわりポテト上質クラスター
10. プレミアム・こだわりクラスター	20. 特殊企画・大容量クラスター

*各クラスターラベルは、AIが自動付与したもの

*各商品は、買われ方が類似したクラスターのいずれか一つに分類

図5. お菓子市場の嗜好クラスタリングAIとLLMを組み合わせた分析結果

嗜好クラスタリングAIでお菓子市場の多様な購買軸を抽出し、LLMで各クラスタの意味付けとして自動ラベリングを行うことで、理解しやすい分析結果が得られる。

Results of preference-clustering analysis and LLM-based cluster labeling (in snack market)

タリングAIとLLMの組み合わせにより、購買行動の背後にある買われ方の構造を抽出し、それらを誰もが活用可能な市場インサイトとして提供できる技術を開発した。これにより、従

来のカテゴリーや売上指標だけでは把握できなかった市場の本質的な構造や潜在ニーズを、データドリブンに理解できることを示した。

この技術を通じて、物価高騰などにより市場環境が大きく変化した際にも、生活者の購買行動や需要の動きをデータに基づいて分かりやすく把握することが可能となる。市場で起きている変化をタイムリーに可視化できるため、企業による価格設定や商品提供の見直しを支援するだけでなく、行政による物価対策や支援施策を検討する際の参考情報としても活用できる。これらを通じて、生活者への新たな価値提供と、社会全体のデータ利活用の促進に貢献していく。

文 献

- (1) 東芝データ, “レシート印字名に基づきJANコードがない商品をAIで自動分類する技術を開発～商品の分類作業を省人化し、レシートデータの分析による商品開発やマーケティング活動の活性化に貢献～”, ニュース, <<https://www.global.toshiba/jp/news/data-corp/2024/11/20241107.html>>, (参照 2026-04-21).
- (2) 布川大知, ほか, “クラスタリングと生成AIを用いた新商品販売予測モデルの提案: スマートレシート[®]の購買ビッグデータを用いた評価実験”, 2026年度人工知能学会全国大会(第40回)論文集, 群馬, 2026-06, 人工知能学会, 4Yin-A.
- (3) 東芝データ, “東芝データ、購買データのリアルタイム分析基盤としてデータブリックスの「レイクハウス・プラットフォーム」を採用～日々発生する大量データの低遅延で効率的な分析環境を構築～”, ニュース, <<https://www.global.toshiba/jp/news/data-corp/2023/06/20230628.html>>, (参照 2026-04-21).
- (4) 東芝データ, “210万人のレシートデータから見る「買い物動向」調査 AIで導き出したシャンプーの購買傾向 シャンプー購入者の”お試し”“需要が明らかに!”, ニュース, <<https://www.global.toshiba/jp/news/data-corp/2024/09/20240911.html>>, (参照 2026-04-21).
- (5) 東芝データ, “ZENBと共同で購買ビッグデータに基づくグルテンフリー食品の市場分析を実施～AI技術を活用して消費者の購買特性と潜在市場のつながりを明らかに～”, ニュース, <<https://www.global.toshiba/jp/news/data-corp/2024/04/20240424.html>>, (参照 2026-04-21).



城田 祐介 SHIROTA Yusuke, Ph.D.
ICTソリューション事業部 データイノベーション技術部
博士(工学)
情報処理学会会員
Data Innovation Engineering Dept.



若松 友裕 WAKAMATSU Tomohiro
ICTソリューション事業部 データイノベーション技術部
Data Innovation Engineering Dept.



金井 達徳 KANAI Tatsunori, Ph.D.
ICTソリューション事業部 データイノベーション技術部
博士(情報学)
Data Innovation Engineering Dept.



中田 康太 NAKATA Kouta, Ph.D.
総合研究所 AIデジタルR&Dセンター アナリティクスAI研究部
博士(理学)
人工知能学会会員
Analytics AI R&D Dept.