

商品と購買者で構成される二部グラフの可視化の一手法

長谷川聡美 (指導教員：伊藤貴之)

1 はじめに

販売者にとって商品の購買者層を把握することは、販売促進や利益向上のために重要である。近年ではデジタル化に伴い、商品の販売者は購買者情報や購買履歴などの情報の入手が容易になった。そこで本研究は、これらの膨大なデータを二部グラフで可視化することで、商品ノードのクラスタ、購買者ノードのクラスタ、クラスタ間の関係を可視化する。そしてその可視化結果より、商品の販売者が商品の売れ行きの把握、効果的な販売戦略の提案、商品に共通する購買者層の把握を容易にすることを目標とする。

二部グラフを可視化する際、一方あるいは両方のグラフの位置を固定する手法があるが、本手法ではノード間の関係に応じて自由に配置する。これにより、エッジの交差を極力減らし、視認性を向上させる。

2 二部グラフ

グラフとは、ノードの集合と、2ノード間を接続するエッジの集合で構成される。グラフは可視化技術の重要な対象の一つである。二部グラフ G はグラフの一種であり、 $G = (V, E) = (X \cup Y, E)$ で表される。ノードの集合 V は二つの部分集合 X, Y で構成され、エッジの集合 E は X に属するノードと Y に属するノードを接続する。このとき、 X に属するノード同士、 Y に属するノード同士はエッジで接続されることはない。

本研究では、商品と、その商品を購入した人をエッジで繋ぐことで可視化する。よって、商品ノード同士、購買者ノード同士を接続するエッジは存在しないため、二部グラフによる可視化が適していると考えられる。

3 関連研究

本章では、二部グラフの可視化に関する関連研究を紹介する。Misue[1]は、二部グラフの可視化手法「Anchored Maps」を提案した。二部グラフの二つの部分集合 X, Y のうち、一方をアンカーノードとし、円周上に等間隔に配置した。もう一方の部分集合をフリーノードとし、隣接するアンカーノードに対して適切な位置に配置し、クラスタの背景を理解することを可能にした。

本手法では、二部グラフの二つの部分集合をそれぞれ自由な配置にすることで、エッジで接続されたノードを近くに配置する。これによりエッジ間の交差を減らし、視認性を向上させる。

4 提案手法

4.1 使用したデータ

本手法では、商品の購買データの一例として映画の評価データセットである The MovieLens Dataset[2] から抜粋した。使用したデータの内容を以下に記載する。

1. 1343本の映画に対する3010人の鑑賞者の内、1作品以上鑑賞した鑑賞者による5段階評価

2. 鑑賞者の人口統計情報（年齢、性別、職業、郵便番号）

本データセットは2000年にMovieLensのWebサイトを通じて収集された。また、アクション、ドキュメンタリー、ホラーといった19種類の映画のジャンル情報が含まれている。本データセットにおいて、映画作品を商品、鑑賞者を購買者として可視化する。

4.2 可視化手法 Koala

本研究で適用するKoala[3]とは、ノードをクラスタリングして配置するグラフ可視化手法である。本手法では、鑑賞者と映画をそれぞれノードで描画し、映画とその映画を評価した鑑賞者をエッジで接続することで、二部グラフとして可視化した。

Koalaはクラスタを単位として各ノードを画面に配置する。

- エッジで連結された隣接ノードの共通性
- ノードに付与された特徴量ベクタの類似性

を算出し、これらの一次結合によりノード間距離が算出される。この距離にもとづいてノードをクラスタリングすることにより、重要なノードが独立された形でのグラフ可視化を実現できる。よって、同じ商品の購買者ノード同士のノード間距離が小さくなる。

次に、ノード配置のアルゴリズムを以下に記載する。

1. クラスタをノードに置き換えたグラフを生成し、これに対してグラフ配置アルゴリズムを適用。
2. 各クラスタを構成するノード数から、各クラスタの半径を計算。
3. エッジ長が両端のノードの半径の合計値にできるだけ近づくようにスムージング処理を適用。
4. クラスタ内にそのクラスタを構成するノードを配置。

以上の手順によって、同じクラスタに分類されたノードを近くに配置する。よって、ノード間距離が小さいノード同士は同じクラスタに属する可能性が高くなる。また、類似するクラスタ同士は近くに配置される。本研究では、商品ごと、購買者ごとにクラスタを作り、そのクラスタ間の関係を可視化する。このとき、商品クラスタが縦長の楕円、購買者クラスタが横長の楕円で表される。商品同士、購買者同士を接続するエッジは存在しないため、同一クラスタ内にリンクが存在しない。以上の理由により、本研究において、Koalaが適していると考えられる。

また、同一のクラスタペアに属する2ノードを連結するエッジ群を束化する。Koalaはノードを自由に配置し、エッジの束が太く、関係が強いノードのクラスタ同士を近くに配置するアルゴリズムである。これらによりエッジ間の交差が減り、視認性が向上する。

5 実行結果・考察

映画の評価データセット (The MovieLens Datasets) における映画と鑑賞者の関係を Koala で可視化した結果を以下の図 1 に示す。

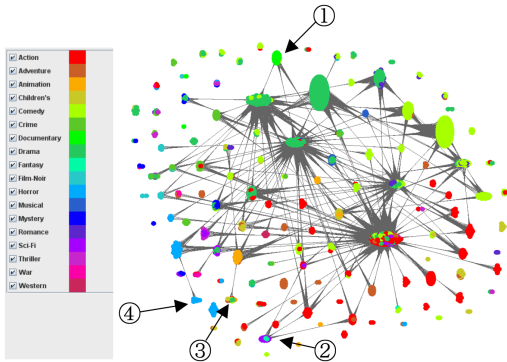


図 1: 映画と鑑賞者で構成される二部グラフ

図 1 において、縦長の楕円は同一の鑑賞者群に鑑賞される映画ノードのクラスターである。横長の楕円は同一の映画を鑑賞する鑑賞者ノードのクラスターである。左に表示されている通り、映画ノードはジャンルにより色分けがされた。鑑賞者ノードの色は鑑賞者の鑑賞履歴に基づき、Koala が自動で色分けをした。

また、エッジは映画ノードと、その映画を評価した鑑賞者ノードを接続している。映画ノード同士、鑑賞者ノード同士がエッジで接続されることはない。

図 1 の① Documentary ジャンルの映画クラスターに着目すると、エッジが多く出ている鑑賞者クラスター三つにエッジが伸びている。ここから、Documentary のみを観る鑑賞者は少なく、幅広いジャンルの映画を観る鑑賞者が、そのうちのひとつとして鑑賞しているということが読み取れた。また、②～④の鑑賞者クラスターに着目する。②は Horror ジャンルの映画クラスター、③は Animation ジャンルの映画クラスター、④は Action ジャンルの映画クラスターとエッジで接続されている。これらの鑑賞者クラスターは特定ジャンルの映画クラスターと関係が強いことから、特定のジャンルの映画のみを好んで鑑賞する鑑賞者が存在することが読み取れた。

次に、図 1 からさらにエッジを減らした可視化結果を以下の図 2 に示す。

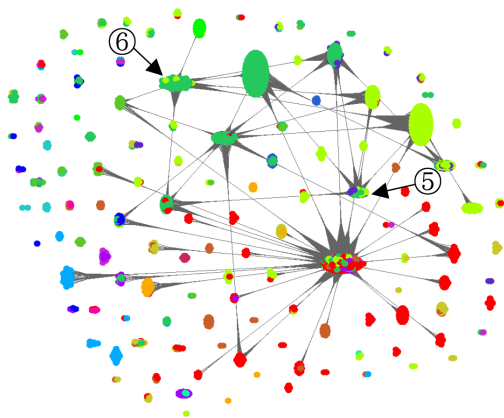


図 2: 図 1 よりエッジ数を減らした可視化結果

図 2 において、⑤の鑑賞者クラスターは Documentary, Crime, Comedy, Drama ジャンルの映画クラスターとエッジで接続されている。また、⑥の鑑賞者クラスターは Comedy, Drama ジャンルの映画クラスターとエッジで接続されている。よって、これらの鑑賞者クラスターに接続された複数のジャンルには、共通の鑑賞者が一定数存在することが読み取れた。

6 可視化結果の活用

5 章で得られた可視化結果より、考察する。図 1 において、Documentary ジャンルの映画は他の映画と共に鑑賞されていることがわかった。また、図 2 において、複数のジャンルの映画に共通の鑑賞者が一定数存在することがわかった。よって、これらのジャンルの映画には、他の映画と併せて鑑賞した際に、鑑賞者にメリットのあるプロモーションが効果的であると言える。さらに、図 1 における②～④の鑑賞者クラスターは、特定のジャンルの映画のみを好んで鑑賞することから、これらの映画ジャンルにはリピーター向けのプロモーションや、鑑賞者の属性に対応したプロモーションが効果的であると言える。

商品の販売者が可視化結果を活用する方法の一例として、以下の二つが挙げられる。一つ目に、購買者層の把握である。商品クラスターに接続された購買者クラスターの属性や購買履歴から、販売者は効果的なプロモーションを提案できると考える。二つ目に、複数の商品に共通する購買者層の把握である。異なる商品クラスターに共通する購買者クラスターを特定し、これに基づいて商品のクロスセリングやマーケティング戦略を行えると考える。

7 まとめと今後の展望

本研究では、商品と購買者で構成される二部グラフの可視化の一手法として、Koala を適用した可視化手法を提案した。ノード間距離に基づいたクラスターリング、自由度の高いノード配置、エッジの束化により、視認性の高い可視化を実現した。

今後の課題として、データセットに含まれる鑑賞者による評価、属性を可視化対象に入れることがあげられる。高評価や低評価、年齢や居住地などの属性を対応させることで、より有益な可視化結果になると考える。

参考文献

- [1] Kazuo Misue. Drawing Bipartite Graphs as Anchored Maps, In Proc. of APVIS2006, 169-177, 2006.
- [2] Maxwell F. Harper and Joseph A. Konstan. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS), 5(4), 19, 2015.
- [3] Takayuki Itoh, and Karsten Klein, Key-node-separated graph clustering and layouts for human relationship graph visualization., IEEE Computer Graphics and Applications, 35(6), 30-40, 2015.