

近傍エッジとの関係に着目したグラフマイニング手法に関する研究

稲福 和史

近年、携帯型インターネット接続端末の普及や、計算機の性能向上、ストレージの低廉化などを背景として、膨大なユーザの行動履歴データを簡便に取得し、様々に分析・応用することが容易になっている。例えば、オンラインショッピングサイトにおける購買行動履歴や、インターネットTVにおける視聴履歴、観光地における移動履歴などを分析して、商品やコンテンツの推薦に利用する応用が広がっている。

これらの収集した行動履歴データは、商品や番組、観光地をノード、ノード間の遷移関係をエッジ、遷移回数を重みとする有向グラフとして表すことができる。このため、大量の時系列データから生成した大規模なグラフを解析し、有用な知見となりうる特徴的な構造を抽出するグラフマイニング研究が盛んであるが、多くの先行研究では、グラフを大域的に捉え、高頻度に出現するパターンを対象に解析を行っている。

しかし、共購買ネットワーク等のグラフでは、高頻度でなくても局所的な構造が重要となる。このため、本研究では、グラフの局所的な構造に着目したグラフマイニング手法を提案する。提案法では、局所的な構造の抽出を目的に、隣接エッジ集合の重みからZスコアを算出し、新たなエッジの重みとして定義する。Zスコアが閾値以上を示すエッジを分析対象とする。

実運用されているオンラインショッピングサイトで投稿されたレビューデータを購買データとみなし、ユーザの購買順序から有向グラフを構築し、提案法の有効性を検証する。エッジの重みは、商品の共起頻度及び隣接エッジ集合内における共起頻度から算出したZスコアとし、閾値以上のZスコアを示すエッジについて、サブグラフを構築した。これらのサブグラフを対象に、モチーフ分析を核とする購買行動パターンを抽出し、サブグラフの持つ性質を従来の頻度に基づく分析手法と比較し、提案法の有効性を比較検討した。その結果、高いZスコアを示すエッジから重要な構造が抽出できることが明らかになった。

(指導教員 佐藤 哲司)