

区間イベント系列からの頻出部分グラフマイニング

鈴木 湧人

要旨

近年、情報技術の発展により様々な種類の大規模データの取得、蓄積が容易となった。これらの大規模なデータから有益な情報や頻出パターンなどを抽出する研究が盛んに行われている。特に、商品の購買履歴や株価など、時間とともに値が変化する時系列データから特徴的な規則性や傾向などを抽出する技術を時系列データ分析という。複数の時系列データからなる多次元時系列データを対象としたパターンマイニングに関する既存研究の多くは、各時系列データに繰り返し現れる箇所を記号化し、他の系列との関係性をパターンとして抽出する。これに対し本研究では、系列同士の類似性に着目することで、既存手法では捉えることのできないパターンの抽出を試みる。

まず、系列同士の類似区間を開始時刻と終了時刻を持つイベントとして捉えることで、多次元時系列データを区間イベント系列化する。その上で区間イベント系列を辺系列とみなし単一のグラフを構成する。そこからイベント時間を考慮したパターンを列挙する。本研究では、パターンの列挙手法を2つ提案する。第一の方法では、各系列を頂点、系列同士の類似区間を辺、辺ラベルを相対的な開始時刻と終了時刻とするグラフを構成し、その上で頻出部分グラフパターンを抽出する。これに対し第二の方法では、各系列を頂点、系列同士の類似区間を辺、辺ラベルを相対的な開始時刻とするグラフを構成し、その上で頻出部分グラフパターンをパターン候補として列挙し、終了時刻を含めた各イベントの順序関係を考慮したパターンを抽出する。株価データを用いた実験を通じ、提案した各手法が実データに対して有効であることを検証する。