

類似区間抽出による多次元時系列データからの 特徴的系列パターンの発見

野口 哲哉

要旨

コンピュータ技術の発達に伴い、膨大な量のデータの取得、蓄積が可能となった。そのため、様々な分野で蓄積したデータを分析し、新しい知識や有益なパターンを発見するデータマイニングの研究が盛んに行われている。その中で時系列データを対象としたデータマイニングを時系列データマイニングと呼ぶ。時系列データとは、時間順に並べられ時間の経過と共に変化するデータで、例として株価やヘルスケアデバイスのログデータなどが挙げられる。時系列データマイニングの主な目的は、時系列データからの規則性や傾向、特徴の抽出である。多次元時系列データからのパターンマイニング研究では、主に各時系列データの特徴的な部分の組み合わせを求める。それに対し本研究では、多次元時系列を類似区間に着目して区間系列に変換することで一つの区間イベント系列とする。その区間イベント系列に対し、ある一定の時間幅内に現れるパターンを抽出することで、従来のパターンマイニングでは捕らえられないパターンを見つける。この時の頻度計算には、系列データの頻度計算アルゴリズムを基に区間イベントに適用する。検証実験には、2007年から2015年までの株取引が行われた約2200日の株価データ100社分を使用し、効率的にパターンを発見できるか、またパラメータの変化によるパターンの違いを確かめる。