

派遣会議：19th International Conference on Database and Expert Systems Applications (DEXA '08)

開催期間：2008年9月1～5日

開催場所：トリノ（イタリア）

#### 参加した国際会議(DEXA '08)について

<参加者>約600人(併合カンファレンスを含む)

<並列するカンファレンス>DaWak, NBis, EGOV, EC-Web, TrustBus, Globe, DEXA-Workshop

#### 発表の様子

<発表内容> "Detecting Current Outliers: Continuous Outlier Detection Over Time-Series Data Streams" (時系列データストリームに対する連続的外れ値検出手法) で発表25分, 質疑応答5分を行った。

#### <質疑応答>

質問1: 実験に使用したデータはfloat型か, int型か?

回答1: float型。

質問2: あるデータ分布があり, 外れ値を見つけ出すようにアプローチしたいときは, どのようにすればよいか?

回答2: パラメタDの値を小さくする。

#### 研究について

<データストリームに対する連続的外れ値検出に関する研究の概要>

データ集合の中から他とは大きく異なる特徴や値を持つ異常値や例外を検出する外れ値検出の技術は, ネットワークアクセスログやビジネスデータからの不正行為の検出, 医療分野での異常値の検出などの様々な分野に応用され, 注目を集めている。一方で, センサ等のデバイスの普及やネットワーク技術の進歩により, 時々刻々と変化する大量の時系列データを取得することが可能となり, このような時系列データに対する外れ値検出では, 異常の検出などのリアルタイム性を求められることが多く, 処理の高速化は大変重要である。そこで本研究では時系列データストリームの特徴を利用した効率的な外れ値検出手法を提案する。

#### <現在検討中の内容>

・提案手法では, 時系列でデータ分布に大きな変化が生じた際に, 既存手法よりも処理時間が長くなってしまふことがある。このようなことが生じた際に, 既存手法へと切り換えるハイブリッドな手法を検討中である。

・データ構造としてキャッシュの使用の有無による処理時間の評価  
提案手法で扱うデータ構造として, キャッシュを使用する方法と, 使用しない方法の2つの方法が考えられる。この違いによる処理時間の評価を試みている。

