

加速度センサと心拍センサを用いたリアルタイムデータ処理による 広告配信

Digital Signage on Stream Data Processing Using Activity Monitor and Heart Rate Sensor

池田 政人*¹ 新美 礼彦*²
Masato Ikeda Ayahiko Niimi

*¹公立はこだて未来大学大学院 システム情報科学研究科 情報アーキテクチャ領域
Graduate School of Systems Information Science, Future University Hakodate

*²公立はこだて未来大学 システム情報科部 情報アーキテクチャ学科
Faculty of Systems Information Science, Future University Hakodate

In this study, we propose a digital signage system according to user attributes made by streaming data mining on storm using an activity monitor and heart rate sensor. We discuss Apache Storm and Apache Spark as streaming data process platform.

1. はじめに

TwitterのようなSNSデータや活動量計のような加速度や心拍センサからなるデータのように、大量のデータがリアルタイムで流れており、データマイニングの問題は、大量データ、多種データの他に、処理速度も取り上げる必要が出てきた[北川 10]。本研究では、このようなストリームデータの処理に注目した。

ストリームデータの代表として、Twitterデータがある。Twitterを用いたデータマイニングの事例は多いが、Tweet内容や位置情報を用いた分析が行われているだけであり、どのようなときにつぶやかれているのかをTweet以外の情報も用いて分析したものは少ない。ユーザの「どのようなとき」を分析する情報として、スマートフォンや活動量計の加速度センサを使う事が考えられる。加速度センサから移動しているのか止まっているのかを推定することは研究されている。しかし、活動量計の計測結果をTwitterを用いて報告する研究[西脇 13]はあるが、Twitterのつぶやきと活動量計の計測結果との関係に注目した研究はない。

そこで、本研究では、ユーザが身体に着用した加速度センサと心拍センサからのデータを元にStormを用いてストリームデータマイニングを行いユーザの属性に応じたデジタルサイネージによる広告配信を行うシステムを提案する。Tweet内容と活動量計、心拍センサのデータをリアルタイムで分析する処理をApache Storm, Apache Spark上に構築し、リアルタイムに属性情報をクラスタリングできるシステムを検討した。

2. ストリームデータ処理基盤

大量データへの問い合わせに対応したシステムとして、Google BigQuery[BigQuery] や Amazon Redshift[Redshift] などがある。クラウドサービスとして、大量データの蓄積、高速処理を容易に行えるシステムであり、データの傾向を対話的に分析するシステムに利用できる。しかし、機械学習のような処理は別に組む必要がある。

大量データを処理するプラットフォームとして、Apache Hadoopがある[Hadoop]。Apache Hadoopは処理をMapと

Reduceに分け、処理データは分割され、Mapで並列処理され、処理結果をReduceでまとめる処理を行う。MapとReduceをうまく組み合わせることにより、大量のデータに対して高速な処理を実現している。しかし、処理がバッチ処理であるため、データを全てためてからでないと処理できない点や、Map/Reduceの処理が独自で既存の分析アルゴリズムの適用が難しい点が問題となっている。特に、機械学習のような同じデータを繰り返し処理するアルゴリズムへの対応は難しい。

これに対し、Apache Stormは、ストリームデータを処理するように設計されている[Storm]。絶え間なく流れてくるデータに対し、データ変換処理が実行される。データのソースをSpout、変換処理を行う部分をBoltとよび、SpoutからBoltの組み合わせによって、処理を行うモデルである。

Apache Sparkも大量データを処理するプラットフォームである[Spark]。Map/Reduce処理を一般化しつつ、メモリ内に作業をキャッシュして処理を行い、繰り返し処理に使う共有データをメモリ上に保持することにより効率的な反復アルゴリズムが行えるように設計されている。また、機械学習やグラフアルゴリズムのライブラリも用意されており、ストリームデータマイニングを行う環境を容易に構築可能である。

本研究では、リアルタイムデータを処理できるApache Storm, Apache Sparkを用いてシステムを構築することとした。

3. 提案システム

提案システムは、Twitterのつぶやきデータと活動量計・心拍センサのデータをリアルタイムでクラスタリングを行い、クラスタに応じた広告をデジタルサイネージに表示するシステムである。Twitterのつぶやきデータと活動量計・心拍センサのデータはあらかじめどのように分類したら良いかがわからないため、教師なし学習が行えるクラスタリングを行うことにした。クラスタがユーザの属性を表すため、ユーザの属性にあわせた広告を提示可能なシステムが構築可能となる。

提案システムでは、まず、Twitter Streaming API[Twitter Streaming APIs]を用いて、データを取り込む。活動量計や心拍センサのデータはスマートフォン経由で取り込むか、Webインタフェースが用意されているなら、Webインタフェース経由で取得する。取得したデータをそれ

連絡先: 新美礼彦, 公立はこだて未来大学, 北海道函館市亀田中野町 116 番地 2, niimi@fun.ac.jp

ぞれ Apache Storm か Apache Spark を用いてクラスタリングする。Apache Storm を使うか、Apache Spark を使うかは、システムの構築のしやすさやスループットなどの結果を見て判断する。Twitter は Twitter データのみでクラスタリングを行い、活動量計・心拍計センサのデータは別にクラスタリングを行う。クラスタリングした結果、データにクラスタのラベルを付与し、ラベルに応じて、広告を配信する。Twitter データのクラスタ数 × 活動量計・心拍計センサのデータのクラスタ数の組み合わせができる。どのクラスタにどのような広告を配信すれば良いのかは、システムの利用場面によって検討する必要がある。

Apache Storm と Apache Spark にデータを渡すために Elasticsearch[Elasticsearch] を利用する。Elasticsearch は分散型 RESTful な検索サーバであり、スキーマフリーでデータを格納できる。Twitter と活動量計・心拍計センサデータという属性の違うデータを格納するのに適している。Elasticsearch には機能を拡張するためのプラグインを用意することができる。Twitter Streaming APIs からデータを取得するプラグインが公開されており、活動量計・心拍計センサを取得するための処理もプラグインとして実装可能である。入力データの種類が増えた際もプラグインを追加するだけで良く、Elasticsearch の利用によりシステムの拡張性を確保できる。また、Apache Storm と Apache Spark もサポートされている [Apache Storm support, Apache Spark support]。このような理由により、Elasticsearch を用いると提案システムの構築が容易になるだけでなく、今後の拡張性も確保できる。

提案システムの適用サービスとして、マラソン大会や、サイクリング大会を想定している。この場合、選手に活動量計や心拍計を使ってもらい、観客が Tweet することを想定している。提案システムを利用することによって、Tweet や活動量計・心拍計センサの分析から、大会の盛り上がりや会場付近の話題のスポットなどの自動抽出やクラスタリング結果にあわせた、選手や観客への情報提供が行える。

実験で使用する活動量計として fitbit[Fitbit] や Polar M400[Polar] を想定している。どちらも安価であり、Web インタフェースが用意されているため、実験に利用しやすい。

4. おわりに

そこで、本研究では、ユーザが身体に着用した加速度センサと心拍センサからのデータを元に Storm を用いてストリームデータマイニングを行いユーザの属性に応じたデジタルサイネージによる広告配信を行うシステムを提案し、Tweet 内容と活動量計、心拍計センサのデータをリアルタイムで分析する処理を Apache Storm, Apache Spark 上に構築し、リアルタイムに属性情報をクラスタリングできるシステムを主に実装面を中心に検討した。今後は、実際にシステムを実装し、実装容易性やデータ処理のスループットの検討を行うと共に、提案システムに向いている広告サービスを検討する。

参考文献

- [北川 10] 北川 博之, 川島 英之, 天笠, 俊之: 2 センシングデータ処理基盤技術: ストリームデータ処理 (<特集> センシングネットワーク). 情報処理, Vol.51, No.9, pp.1119-1126, 情報処理学会, 2010.
- [西脇 13] 西脇 雅人, 中嶋 名菜, 池上 由美, 川上 諒子, 黒部 一道, 松本 直幸: 活動量計と Twitter を併用した生活介

入が身体活動量に与える影響: 無作為割り付け介入試験. 体力科学, Vol.62, No.4, pp.293-302, 2013.

- [BigQuery] What is BigQuery? - Google BigQuery - Google Cloud Platform <https://cloud.google.com/bigquery/what-is-bigquery>.
- [Redshift] Amazon Redshift (クラウドデータウェアハウスソリューション) - アマゾン ウェブ サービス (AWS 日本語) <https://aws.amazon.com/jp/redshift/>.
- [Hadoop] Welcome to Apache Hadoop! <https://hadoop.apache.org/>.
- [Storm] Storm, distributed and fault-tolerant realtime computation <https://storm.apache.org/>.
- [Spark] Apache Spark - Lightning-Fast Cluster Computing <https://spark.apache.org/>.
- [Twitter Streaming APIs] The Streaming APIs - Twitter Developers <https://dev.twitter.com/streaming/overview>.
- [Elasticsearch] Elastic Revealing Insights from Data (Formerly Elasticsearch) <https://www.elastic.co/>.
- [Apache Storm support] Apache Storm support <http://www.elastic.co/guide/en/elasticsearch/hadoop/master/storm.html>.
- [Apache Spark support] Apache Spark support <http://www.elastic.co/guide/en/elasticsearch/hadoop/current/spark.html>.
- [Fitbit] Fitbit 公式サイト <http://www.fitbit.com/jp>.
- [Polar] 心拍計と GPS 対応スポーツウォッチ - Polar Japan <http://www.polar.com/ja>.