

社会センサデータ生成・共有基盤における ストリーミングデータ処理機構

中嶋 奎介 横山 正浩 義久 智樹 原 隆浩*

大阪大学情報科学研究科[†]

大阪大学サイバーメディアセンター[‡]

1 はじめに

近年普及している Twitter や Facebook といった SNS (Social Networking Service) は様々な人々や自治体に利用されており、それらの投稿を解析することで、実社会の情報を取得できる。本研究では、SNS の投稿を解析するプログラムを社会センサ、解析結果を社会センサデータと呼ぶ。例えば、地震というキーワードが含まれた Twitter のツイートを解析して地震を検知する手法が提案されており [1]、この手法における、ツイートの解析のためのプログラムが社会センサ、地震を検知したことと、その場所や時刻が社会センサデータとなる。

筆者らの研究グループでは、社会センサや社会センサデータをユーザ間で共有するために、社会センサデータ生成・共有基盤である S^3 システムを開発している [2]。 S^3 システムでは、ユーザは SNS の解析処理関数を記述した SSFD、社会センサの定義を記述した SSTD、社会センサデータの出力に関して記述した SSOC をシステムにアップロードする。 S^3 システムが三つのファイルから社会センサを作成し、社会センサデータを生成する。

SNS へ連続的に投稿されるデータを、データが投稿される (ストリーミングデータが発生する) 毎に解析することで、地震に関するツイートが投稿されてから短時間で地震を検知するといった、リアルタイムな社会センサデータの生成につながる [3]。しかし既存の S^3 システムでは、ストリーミングデータの発生頻度が高い場合、解析を行えず、社会センサデータをリアルタイムに生成できないことがあった。そこで本研究では、 S^3 システムのためのストリーミングデータ処理機構を提案する。提案する機構では、ストリーミングデータの受信モジュールを追加することで受信と解析を並列に実行し、社会センサデー

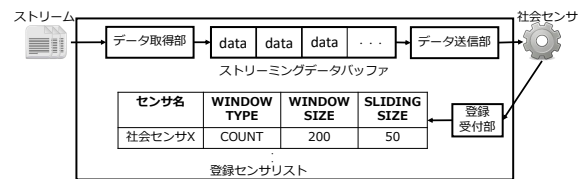


図1 ストリーミングデータ処理機構の構成

タ生成までの遅延を短縮させる。

2 ストリーミングデータ処理機構

2.1 機能要件

S^3 システムにおいてデータの受信と解析を並列に行うことで、社会センサが SNS への投稿を解析している間も SNS で発生するストリーミングデータを受信できる。そこで、本研究では、解析と並列に受信を実行できるストリーミングデータの受信モジュールを追加する。受信モジュールには、受信したデータを解析するために、データを社会センサへ配布する機能と一時的に蓄える機能が必要となる。また、受信したデータの配布先を提案機構に伝えるため、配布先となる社会センサを登録機能する機能が求められる。

2.1.1 システム構成

図1にストリーミングデータ処理機構のシステム構成を示す。各構成要素を以下に説明する。

データ取得部

SNS からストリーミングデータを常に受信し、ストリーミングデータバッファに格納する。

データ送信部

登録センサリストに従ってストリーミングデータバッファのデータを社会センサに配布する。

ストリーミングデータバッファ

データ取得部によって取得されたストリーミングデータを一時的に保持する。

登録受付部

ストリーミングデータの配布先である社会センサの登録を受け付ける。

登録センサリスト

社会センサの情報を保持する。センサ名には社会センサの名前を保持する。WINDOWTYPEにはウィンドウの種類を保持する。ウィンド

A Streaming Processing System for a Platform for Generate and Share Social Sensor data

*Keisuke Nakashima, Masahiro Yokoyama, Tomoki Yohishisa, Takahiro Hara

[†]Graduate School of Information Science and Technology, Osaka University

[‡]Cybermedia Center, Osaka University

```

<ELEMENT SSTD (NAME,UPDATE,KEYWORD,ITERATOR,VALUES,REFERENCES)>
<ELEMENT NAME (#PCDATA)>
<ELEMENT UPDATE (CRON|EVENT|ONCE|REUSE|MQTT)>
<ELEMENT EVENT EMPTY>
<ATTLIST EVENT F_TIME CDATA #REQUIRED>
<ELEMENT ITERATOR (EVENT_DATA)>
<ELEMENT EVENT_DATA EMPTY>
<ATTLIST EVENT_DATA WINDOW_TYPE CDATA #REQUIRED
WINDOW_SIZE CDATA #REQUIRED
SLIDING_SIZE CDATA #IMPLIED>
<ELEMENT VALUES (VALUE+)>
<ELEMENT VALUE EMPTY>
<ATTLIST VALUE V_NAME CDATA #REQUIRED
I_NAME_REF CDATA #REQUIRED
V_TYPE CDATA #REQUIRED
V_FUNCTION CDATA #REQUIRED>

```

図2 SSTDのDTD (一部抜粋)

ウの種類には“COUNT”と“TIME”がある。“COUNT”の場合、指定された数のストリーミングデータを受信する毎に処理を行い、“TIME”の場合、指定された時間の間に受信したストリーミングデータで処理を行う。WINDOW_SIZEには処理を行うストリーミングデータの数(ウィンドウサイズ)を保持する。SLIDING_SIZEにはウィンドウのずらし幅を保持する。

2.1.2 SSTDの設計の変更

ストリーミングデータ処理機構に社会センサを登録するために、SSTDに登録情報を記述できる必要がある。変更後のDTDを図2に示す。UPDATE要素内に“EVENT”を記述し、EVENT_DATA要素のWINDOW_TYPE属性、WINDOW_SIZE属性、およびSLIDING_SIZE属性内に登録情報を記述する。他の要素については文献[2]と変更はない。

2.2 評価実験

本節では評価実験の結果を示す。提案機構の有用性を確認するために、既存の S^3 システムのCRONを用いた実行方法と比較する。従来の S^3 システムでは、CRONを用いることで定期的にデータを取得できる。

2.2.1 実験環境

CentOS6.7, Intel Core i7 CPU@1.86GHz, 16GBのメモリを搭載しているワークステーション上に、Java1.8.0.131を用いて提案機構を実装した。 S^3 システムのその他の機能は先行研究ですでに実装済みである。SSFDはJavaで、SSTD, SSOCはXMLで記述している。

2.2.2 設定

S^3 システムへのストリーミングデータの入力として、TwitterのストリーミングAPIで取得するツイートをを用いた。データの取得には、主にTwitterのストリーミングサーバへの接続時間(グラフ中接続時間)と接続後から、データを取得して解析関数に渡す時間(グラフ中データ取得時間)がある。短時間にデータを取得して解析関数に渡すことでリアルタイムに解析できるため、評価項目としてデータ取得にかかる時間を用いる。実験用パラメータとして、地震の検出を想定してWINDOW_TYPEを“COUNT”, WINDOW_SIZEを“100”, SLIDING_SIZEを“100”とした。

2.2.3 実験結果

図3に実験結果を示す。CRONを用いた実行方法では、社会センサが実行されるごとにストリーミ

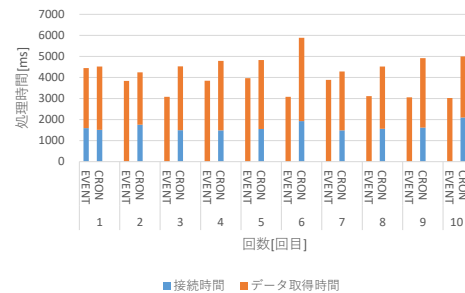


図3 CRONとEVENTのTweet取得にかかる時間

ングサーバに接続する必要があるため、提案機構を用いた処理の方がデータの取得にかかる時間が短くなっている。提案機構を用いることで、社会センサの登録時にストリーミングサーバに接続され、以降維持することにより解析結果を得るまでの遅延を短縮できる。

3 まとめ

本研究では、リアルタイムな社会センサデータの生成を目的として、 S^3 システムのためのストリーミングデータ処理機構を提案した。提案した機構を用いることで、受信と解析を並列に実行でき、データが発生してから解析結果を得るまでの遅延を短縮できることを確認した。今後の課題として、登録センサリストに登録される社会センサ数に対する提案機構の負荷の計測等が考えられる。

謝辞

本研究の一部は、文部科学省科学研究費補助金・基盤研究(A)(26240013), 基盤研究(B)(15H02702), JST国際科学技術共同研究推進事業(戦略的国際共同研究プログラム)の研究助成によるものである。ここに記して謝意を表す。

参考文献

- [1] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors, Proc. World Wide Web Conf., pp. 851–860 (2010).
- [2] Nakashima, K., Yokoyama, M., Taniyama, Y., Yohshihisa, T. and Hara, T.: S^3 System: A System for Sharing Social Sensor Data and Analytical Programs, Adjunct Proc. ACM Int'l Conf. on Mobile and Ubiquitous Systems: Computing Networking and Services, pp. 147–152 (2016).
- [3] Yadranjiaghdam, B., Yasrobi, S. and Tabrizi, N.: Developing a Real-time Data Analytics Framework For Twitter Streaming Data, learning, p. 14 (2016).