

社会センサデータ生成・共有基盤における社会センサの関連性提示機構

中嶋 奎介[†] 横山 正浩[†] 義久 智樹^{††} 原 隆浩[†]

[†] 大阪大学大学院情報科学研究科 〒565-0871 大阪府吹田市山田丘 1-5

^{††} 大阪大学サイバーメディアセンター 〒565-0047 大阪府茨木市美穂ヶ丘 5-1

E-mail: †{nakashima.keisuke,yokoyama.masahiro,yoshihisa,hara}@ist.osaka-u.ac.jp

あらまし 近年普及している SNS (Social Networking Service) への投稿を解析することで、実社会の情報 (社会センサデータ) を取得できる。筆者らの研究グループでは、社会センサデータや社会センサ (SNS への投稿を解析するプログラム) 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 システムを研究開発している。新しい社会センサを作成する際、既存の関連する社会センサのプログラムの記述等が参考になる。これまでの S^3 システムでは、社会センサの検索結果がリスト形式で提示され、社会センサ間の関連性がわからず、参考となる社会センサを発見するまでに多数の社会センサの記述を確認する必要があった。そこで、本研究では、検索結果を可視化し、社会センサ間の関連性を提示する機構の設計と実装を行った。提案する提示機構では、過去に新しい社会センサの作成のために参考にされた事実と記述の類似度を関連性と捉えて、社会センサの関連性を可視化する。この際、社会センサを節で示し、参考にされた社会センサと枝で接続して、社会センサ間の関連性を提示する。

キーワード 社会センサ, 可視化, ソーシャルマイニング

1. はじめに

近年, Twitter や Instagram といった SNS (Social Networking Service) が広く普及している。SNS は多くの人々や地方自治体など, 様々なユーザによって利用されており, それらのサービスに投稿された短文や写真から, 単語の出現頻度や写真の撮影位置を解析することで, 実社会の情報を取得できる。本研究では, SNS への投稿等を解析して実社会の情報を得ることを, 社会センサによって社会をセンシングすると捉え, 得られたデータを社会センサデータと呼ぶ。図 1 に社会センサデータ生成のイメージを示す。社会センサにより SNS への投稿等を解析して社会センサデータが生成され, 社会センサデータベースに蓄えられる。利用者はこれらのデータを閲覧して話題やイベントといった実社会の情報を取得できる。これまでに筆者らの研究グループでは, 社会センサデータや社会センサをユーザ間で共有することを目的として, 社会センサデータ生成・共有基盤である S^3 システムを研究開発してきた [1]。 S^3 システムでは, ユーザがアップロードした社会センサ (SNS への投稿を解析するプログラム) をデータベース内に保存しており, Web インタフェースを介してユーザ間で共有できる。

社会センサは, 解析するプログラムと, 入出力データに関する情報を記述して作成される。新しい社会センサを作成する際, 既存の社会センサを検索してその記述を閲覧し, 作成の参考にすることが考えられる。例えば, Twitter の投稿を解析して地震を検知するために, Twitter に投稿された短文を形態素解析して地震に関する単語の出現頻度を算出する場合がある。この場合, 同様に短文を形態素解析して話題抽出を行う社会センサの記述が参考になる。これまでの S^3 システムで社会センサを検索する際, キーワード検索の結果がリスト形式で提示されており, 既存の社会センサ間の関連性を確認しにくかった。

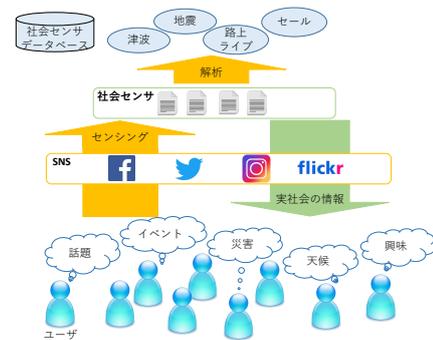


図 1 社会センサデータ生成のイメージ

このため, 参考になる社会センサを発見するまでに多数の社会センサの記述を確認する必要があった。既存の社会センサの作成時に参考にした社会センサ (作成履歴) や他の社会センサとの記述 (プログラムと入出力データ) の類似度を検索結果に提示することで, 社会センサ間の関連性を確認しながら記述内容を閲覧でき, 参考になる社会センサを発見するまでに確認する社会センサの数を削減できる。さらに, これまでの S^3 システムでは, 社会センサの作成時に付加されたキーワードを用いたキーワード検索しか利用できなかったため, 詳細な検索を行えなかった。社会センサに入力するデータや解析方法, 出力するデータの種類といった社会センサに関する情報をさらに付加することで, 詳細な検索を行える。

そこで本研究では, S^3 システムにおいて, 過去に新しい社会センサの作成のために参考にされた事実と記述の類似度を関連性と捉え, 社会センサ間の関連性を提示する機構の設計と実装を行った。提案する提示機構では, 付加情報によるキーワード検索に適合した社会センサ間の関連性をグラフ形式で提示する。社会センサを節で示し, 参考にした社会センサと枝で接続

し、計算した類似度を枝付近に表示する。提案する提示機構を用いることで、ユーザは社会センサ間の関連性を容易に把握でき、参考とする社会センサをより発見しやすくなる。

以降、2章で関連研究を紹介し、3章で提案する提示機構の設計について述べる。4章で提案する提示機構の実装について述べ、5章では提案する提示機構に関する評価実験について述べる。最後に6章で本論文をまとめる。

2. 関連研究

プログラムソースコードの検索に関する研究がいくつか行われている [2] [3] [4]。川満ら [5] は、高速に類似プログラムソースコードを検索するために LSH (Locality Sensitive Hashing) を用いた検索手法を提案している。川満らの手法では、プログラムソースコード間の類似度として、互いの共通要素の割合を示す Jaccard 係数を用いている。高速に動作して広く使われているハッシュ関数 (MinHash) を用いて、検索対象となるプログラムソースコードのハッシュ値をあらかじめ算出し、データベースに保存する。入力されたプログラムソースコードと類似したプログラムソースコードを検索する際、入力されたプログラムソースコードの MinHash を求める。データベースに保存された検索対象となるプログラムソースコードの MinHash と比較して、最尤推定によりプログラムソースコード間の類似度を算出し、類似するプログラムソースコードを検索している。プログラムソースコード間の類似度を高速かつ高精度に算出できるため、提案する提示機構では、川満らの手法と同様に、LSH を用いた類似度の算出を行う。

プログラムソースコードの可視化に関する研究がいくつか行われている [6]。Collberg ら [7] はプログラムソースコードの編集記録を可視化する GEVOL を提案している。GEVOL では、Java で記述されたプログラムに対して、編集時刻や編集内容を記録し、これらの編集履歴を時系列にして、グラフ形式で可視化して提示している。しかし、GEVOL では、プログラムソースコードの編集履歴を提示するのみであり、参考になる社会センサを検索するといった、類似したプログラムの検索を行えない。

kibana^(注1) は Elasticsearch^(注2) で検索した結果をグラフや表といった形式で可視化するサービスである。Elasticsearch の全文検索結果を kibana で可視化できるが、本研究で提案する提示機構のように、プログラムソースコード間の作成履歴や類似度を提示できない点で本研究とは異なる。また、インターネットを介して共有されているプログラムソースコードに対して、キーワード検索を行うサービスとして GitHub Code Search^(注3) がある。しかし、このサービスでプログラムソースコード間の関連性を可視化できない。

3. 社会センサの関連性提示機構の設計

本章では、 S^3 システムにおける社会センサの関連性提示機



センサ名	作成時刻	最終更新時刻	生成数	検索状況
extract_news_from_tweet	2017-07-13 11:41:29	-	-	終了
Twitter_Sensor	2017-07-13 11:45:20	-	-	終了
Instagram_Sensor	2017-07-13 11:46:37	-	-	終了
Flickr_Sensor	2017-07-13 11:52:05	-	-	終了

図2 これまでの S^3 システムの検索画面

構の設計について述べる。

3.1 S^3 システムについて

S^3 システムは、社会センサデータの生成および共有や社会センサの共有を目的として設計、実装したシステムである。まずはじめに、 S^3 システムの概要と、課題について述べる。

3.1.1 S^3 システムの概要

S^3 システムでは、ユーザは社会センサデータを生成するために、SNS への投稿等を解析するコードが記述されたファイル (SSFD: Social Sensor Function Description)、社会センサの型定義ファイル (SSTD: Social Sensor Type Definition)、出力される社会センサデータを社会センサデータベースに格納する際の設定を記述したファイル (SSOC: Social Sensor Output Configuration) の三つのファイルをアップロードする。 S^3 システムは、アップロードされたファイルをもとに、 S^3 システム上で動作する社会センサデータ生成プログラムを作成する。社会センサデータ生成プログラムは S^3 システム上で定期的に行われ、社会センサデータが生成される。生成された社会センサデータやアップロードされた三つのファイルは、Web インタフェースを介して検索、ダウンロードできる。

3.1.2 これまでの S^3 システムの課題と解決策

1章で説明した通り、新しい社会センサを作成する際、既存の社会センサを検索することが考えられる。図2にこれまでの S^3 システムの検索画面を示す。これまでの S^3 システムでは、SSTD に記述されたキーワードを利用して社会センサを検索し、検索結果はリスト形式で一覧表示される。そのため、社会センサ間の関連性がわからず、求める社会センサを発見するまでに多数の社会センサの記述を確認する必要があった。過去に参考にされた社会センサや他の社会センサとの記述の類似度を提示することで、社会センサ間の関連性を確認し、参考になる社会センサをより発見しやすくなると考えられる。

また、これまでの SSTD に記述されたキーワードのみでは社会センサに関する情報を十分に表現できておらず、参考とする社会センサを発見するまでに検索語を変更して複数回検索する必要があった。例えば、キーワードに“名詞”が記述されている場合、入力するデータが名詞であることを示すのか、出力するデータが名詞であることを示すのか不明確であった。SSTD のキーワードだけではなく、その他の情報に対応するタグ (属性と値) を付与することで、より詳しく検索できる。

3.2 社会センサの関連性提示機構の設計方針

提案する提示機構では、過去に新しい社会センサの作成のために参考にされた事実と記述の類似度を関連性と捉える。過去に参考にされた社会センサは、作成履歴を作成することで把握

(注1) : kibana: <https://www.elastic.co/jp/products/kibana>

(注2) : Elasticsearch: <https://www.elastic.co/jp/products/elasticsearch>

(注3) : <https://github.com/search>

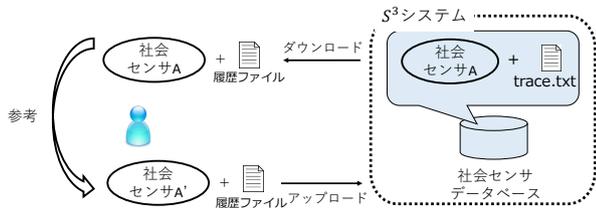


図 3 SSFD の作成履歴把握のイメージ

<p>[SSTD]</p> <ul style="list-style-type: none"> -社会センサデータ名 -分析関数名 -更新の情報 -メタデータ -入力元 -再利用した社会センサID 	<p>[SSOC]</p> <ul style="list-style-type: none"> -カラム名 -カラムの型 -データフロー制御部への出力許可 -再利用した社会センサID
---	--

図 4 作成履歴把握のために SSTD および SSOC に記述する内容

できる。作成履歴は、社会センサが参考にした社会センサを示す履歴である。社会センサを検索するユーザは、提示された作成履歴を確認し、社会センサの作成のために参考にした社会センサを確認できる。さらに、社会センサ間の関連性を示す指標として、SSFD, SSTD, SSOC の記述内容の類似度を用いる。社会センサ間の類似度を提示することで、社会センサ間の類似の度合いを確認しながら検索できる。関連性の提示にはグラフ形式の可視化がよく用いられており、本研究においても関連性をグラフ形式で可視化する。また、社会センサの内容についてより詳しい情報を付加するために、三つのファイルで構成される社会センサに対して、属性と値が対になったタグを用いる。

以下に、各機能の詳細を述べる。

3.2.1 社会センサの作成履歴の把握機能

提案する提示機構では、作成履歴を提示するために、作成履歴を把握する機能が求められる。記述方式の違いから、SSFD と、SSTD および SSOC に分けて説明する。

図 3 に SSFD の作成履歴の把握機能の概要を示す。ユーザが S^3 システム内にある SSFD を閲覧し、参考にするためにダウンロードボタンをクリックすると、 S^3 システムは SSFD, SSTD, SSOC のファイルと共に履歴ファイルを zip 形式等でアーカイブしたファイルを作成する。ユーザはこのファイルをダウンロードする。履歴ファイルには、ダウンロードされた社会センサの識別子となる暗号化された社会センサ ID が記述されている。暗号化された社会センサ ID を用いるのは、ユーザによる改ざんを防ぐためである。例えば、作成した社会センサを多くの人に参考にしてもらうために自身が作成した社会センサを参考元と偽って記述するといった改ざんが考えられる。社会センサのダウンロード後、ユーザはダウンロードした SSFD を参考にして、新たな SSFD を作成してファイルをアップロードする。アップロードする際に、履歴ファイルを含めることで、 S^3 システムは参考にした社会センサを把握できる。履歴ファイルを含めずにアップロードして作成履歴を把握させないことも可能である。

SSFD と異なり、SSTD や SSOC は XML のようなマークアップ言語で記述するため、タグ形式で情報を付加できる。提案する提示機構では、SSTD と SSOC の作成履歴を把握するために、作成履歴に関する情報を付加するようにこれらを変更した。SSTD および SSOC の変更内容を図 4 に示す。太字の部分が、提案する提示機構のために追加した項目である。SSTD や SSOC に、参考にした社会センサの ID を記述することでこれらを把握できる。

3.2.2 社会センサ間の類似度の算出機能

類似度の算出方法の違いから SSFD と、SSTD および SSOC に分けて説明する。

SSFD はプログラム言語で記述されているため、プログラムソースコード間の類似度を算出する既存手法を用いる。ユーザが作成した新たな SSFD を S^3 システムにアップロードする際、履歴ファイルが含まれていれば、参考にした社会センサが明らかであるため、 S^3 システムはアップロードされた SSFD と履歴ファイルに記述された社会センサの SSFD 間の類似度を算出する。履歴ファイルが含まれていなかった場合、 S^3 システムは新しくアップロードされた SSFD と S^3 システム内のすべての SSFD との類似度を算出する。

ユーザが作成した新たな SSTD および SSOC を S^3 システムにアップロードする際、参考にした社会センサ ID が記述されていれば、 S^3 システムはその社会センサの SSTD 間および SSOC 間の類似度を算出する。2 章で説明した通り、類似度の計算方法として様々な方法が考えられるが、実装した提示機構では、4.2.2 で説明する計算方法を用いている。

3.2.3 社会センサへのタグの付与およびタグ検索の機能

提案する提示機構では、社会センサに対して属性と値の組でタグを付与する。属性と値の組で社会センサにタグを付与することで、単純なキーワードよりも、詳細に社会センサの記述内容を表現できる。また、これらのタグを用いて社会センサを検索するために、タグを用いた検索機能が必要である。社会センサの検索時に、タグを指定して検索することを可能にする。

3.2.4 社会センサの可視化に関する機能

3.2 で述べた通り、社会センサの関連性をグラフ形式で可視化するため、提案する提示機構では、社会センサを節で表し、参考にした社会センサと枝で接続する。ユーザが参考となる社会センサを発見しやすくするために、さらに下記の機能を開ける。

類似度による作成履歴の表示制限

ユーザが検索結果内のある社会センサに着目し、その社会センサとの類似度が高いあるいは低い社会センサを発見するとき、不要な作成履歴を表示し、不要な社会センサも表示されるため、グラフ形式で表示されている社会センサの関連性の視認性が低くなる場合がある。例えば、ある社会センサを閲覧し、それと全く異なる社会センサを発見するとき、類似度が高い社会センサとの間に枝があると視認性が低くなる。そこで、提案する提示機構では、ユーザが表示する類似度を設定すると、設定範囲内の作成履歴のみ表示する機能を提供する。

検索結果外の関連性の提示



図 5 参考にした SSFD の識別子のアップロード画面

検索した社会センサを可視化する際、ユーザが入力した検索条件に適合した社会センサが参考にした社会センサも参考になる可能性があるこれらも表示する機能を提供する。

ホップ数による社会センサの表示制限

ある社会センサに着目した時、その社会センサを始点として出ている枝に接続している社会センサまでを1ホップと定義する。ホップ数が大きくなるほど、始点となる社会センサと関連の低い社会センサが提示される可能性がある。そこで、提案する提示機構では、ユーザが入力した検索条件に適合した社会センサに着目し、その社会センサから指定されたホップ数以上で接続している社会センサを表示しない機能を提供する。

履歴ファイルの有無による作成履歴の表示制限

ユーザが作成した新たな SSFD を S^3 システムにアップロードする際、履歴ファイルを含めなかった場合、 S^3 システムは新しくアップロードされた SSFD と S^3 システム内のすべての SSFD との類似度を算出する。しかし、全ての社会センサとの類似度を表示すると、社会センサの作成履歴の把握が困難になる。そこで、提案する提示機構では、作成履歴の表示の有無を選択する機能を提供する。

4. 社会センサの関連性提示機構の実装

3章で述べた設計に基づき、 S^3 システム上に社会センサの関連性提示機構を実装した。本章では、実装した機能の詳細について述べる。

4.1 実装環境

社会センサの作成履歴の把握機能や社会センサ間の類似度の算出機能、社会センサへのタグの付与機能やタグ検索の機能は Java1.8.0.131 で実装した。また、社会センサの可視化には JavaScript のライブラリである d3.js [8] を用いた。履歴ファイルには、trace.txt という名前のテキストファイルを用いた。

社会センサの新規登録、社会センサの確認、社会センサデータの取得などの機能は、先行研究で実装済みである。ユーザがシステムに入力するファイルも先行研究と同様に、Java ファイルもしくは JAR ファイル、SSTD および SSOC は XML ファイルとした。

4.2 実装した機能

4.2.1 社会センサの作成履歴の把握機能

図 5 に SSFD の識別子をアップロードする画面を示す。trace.txt に記述する社会センサ ID の暗号化および復号化に用いる秘密鍵は S^3 システムが保持する。

SSTD および SSOC の DTD を図 6、図 7 にそれぞれ示す。今回の提示機構の実装に伴って、REFERENCE 要素の ID 属性を記述できるように修正している。 S^3 システムに登録されてい

```
<ELEMENT SSTD (NAME|UPDATE|KEYWORD|ITERATOR|VALUES|REFERENCES)>
<ELEMENT NAME (#PCDATA)>
<ELEMENT UPDATE (CRON|EVENT|ONCE|REUSE|MQTT)>
<ELEMENT CRON EMPTY>
<IATTLIST CRON U_TIME CDATA #REQUIRED
F_TIME CDATA #REQUIRED>
<ELEMENT EVENT EMPTY>
<IATTLIST EVENT F_TIME CDATA #REQUIRED>
<ELEMENT ONCE EMPTY>
<ELEMENT REUSE EMPTY>
<IATTLIST REUSE F_TIME CDATA #REQUIRED>
<ELEMENT MQTT EMPTY>
<IATTLIST MQTT F_TIME CDATA #REQUIRED>
<ELEMENT KEYWORD (#PCDATA)>
<ELEMENT ITERATOR (EVERY+|EVENT_DATA|MQTT_DATA)>
<IATTLIST EVERY I_NAME CDATA #REQUIRED
SIZE CDATA #REQUIRED
OVERLAP CDATA #REQUIRED
I_TYPE CDATA #REQUIRED
I_FUNCTION CDATA #REQUIRED>
<ELEMENT EVENT_DATA EMPTY>
<IATTLIST EVENT_DATA WINDOW_TYPE CDATA #REQUIRED
WINDOW_SIZE CDATA #REQUIRED
SLIDING_SIZE CDATA #IMPLIED>
<ELEMENT MQTT_DATA EMPTY>
<IATTLIST MQTT_DATA TOPIC CDATA #REQUIRED
QOS CDATA #REQUIRED>
<ELEMENT VALUES (VALUE+)>
<ELEMENT VALUE EMPTY>
<IATTLIST VALUE V_NAME CDATA #REQUIRED
I_NAME_REF CDATA #REQUIRED
V_TYPE CDATA #REQUIRED
V_FUNCTION CDATA #REQUIRED>
<ELEMENT REFERENCES (REFERENCE*)>
<ELEMENT REFERENCE EMPTY>
<IATTLIST REFERENCE ID CDATA #REQUIRED>
```

図 6 SSTD の DTD

```
<ELEMENT SSOC (FILE_ACCESS|DATA_ACCESS|TRANSFER2FC|COLUMN|INDEXKEY|REFERENCES)>
<ELEMENT DATA_ACCESS EMPTY>
<IATTLIST DATA_ACCESS CLASS CDATA #REQUIRED>
<ELEMENT FILE_ACCESS EMPTY>
<IATTLIST FILE_ACCESS CLASS CDATA #REQUIRED>
<ELEMENT TRANSFER2FC EMPTY>
<IATTLIST TRANSFER2FC PERMISSION CDATA #REQUIRED>
<ELEMENT COLUMN (COLUMN_DATA+)>
<ELEMENT COLUMN_DATA EMPTY>
<IATTLIST COLUMN_DATA C_NAME CDATA #REQUIRED
C_TYPE CDATA #REQUIRED>
<ELEMENT INDEXKEY (#PCDATA)>
<ELEMENT REFERENCES (REFERENCE*)>
<ELEMENT REFERENCE EMPTY>
<IATTLIST REFERENCE ID CDATA #REQUIRED>
```

図 7 SSOC の DTD

る SSTD や SSOC を参考にした場合、ユーザは REFERENCE 要素の ID 属性に参考にした社会センサ ID を記述する。

4.2.2 社会センサ間の類似度を計算する機能

SSFD は Java で記述されたプログラムソースコードであり、2章で説明した通り、プログラムソースコードの類似度を算出する様々な方法が提案されている。高精度かつ高速に類似度を算出する手法として、LSH を用いた手法が文献 [5] において提案されている。提案する提示機構において、高精度かつ高速に類似度を算出するために、本研究でも LSH を用いた類似度の算出を行う。LSH を用いた類似度の算出には、プログラムソースコードの n-gram (n 文字ごとの文字列) が必要であり、本研究では、文献 [5] でも用いられている 3-gram を用いて実装した。

SSTD はマークアップ言語で記述されており、記述内容のほとんどが要素と属性の宣言であり、ユーザによる自由記述部分が限られている。ユーザが自由に記述できる部分だけで類似度を算出するため、要素の値の一致数の割合を類似度として利用する。SSTD には、SSFD で利用する変数名などの SSFD の記述内容に関連する要素 (ITERATOR 要素, VALUES 要素) や、参考にした社会センサに関連する要素 (REFERENCES 要素), 社会センサの実行に関連する要素 (UPDATE 要素など) が含まれる。SSTD に記述されている社会センサに関する定義が類似する SSTD の類似度を大きくするため、これらのうち、SSFD の記述内容に関連する要素、社会センサの実行に関連する要素の一致数の割合を類似度とした。

SSOC もマークアップ言語で記述されているため、要素の一致数の割合を類似度として利用する。SSTD と同様に、SSOC



図 8 タグの付与画面

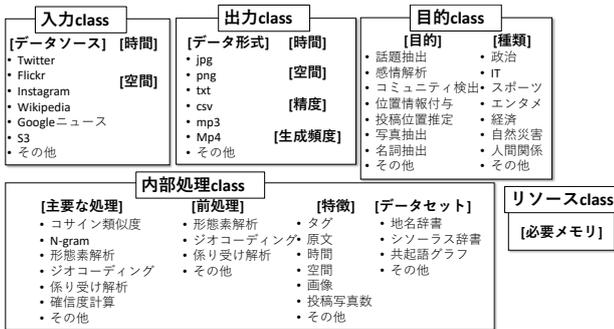


図 9 タグセット



図 10 社会センサの検索画面



図 11 検索結果の表示の例

には SSFD で利用する変数名などの SSFD の記述内容に関連する要素 (COLUMN 要素) や、参考にした社会センサに関する要素 (REFERENCES 要素), 社会センサや社会センサデータへのアクセスに関連する要素 (DATA_ACCESS 要素, FILE_ACCESS 要素など) が含まれる. SSOC に記述された社会センサデータの出力方法が類似する SSOC の類似度を大きくするため, これらのうち, SSFD の記述内容に関連する要素, 社会センサや社会センサデータへのアクセスに関連する要素の一致数の割合を類似度とした.

4.2.3 社会センサへのタグの付与およびタグ検索の機能

3.2.3 で述べた通り, 提案する提示機構には, 社会センサにタグを付与する機能が求められる. タグを付与するために, 社会センサのアップロード時に属性と値を入力する画面を設けた. 図 8 にタグの付与画面を示す. 属性欄に属性を, 値欄にその社会センサを表す値を入力し, 追加ボタンをクリックすると社会センサにタグを付与できる. 属性および値は図 9 に示すタグセットを参考に付与する. 本タグセットは, これまでに作成した社会センサを参考にして作成した. 不足する場合には属性を追加できる.

図 10 に社会センサの検索画面を示す. 属性欄に適した値を値欄に入力し, “検索” ボタンをクリックすることで社会センサを検索できる.

4.2.4 社会センサの可視化に関する機能

図 11 に検索結果の表示画面を示す. 4.2.3 で述べた社会センサの検索機能を用いてユーザが検索すると, 図 11 に示す検索結果の表示の例のように社会センサをリスト形式で提示する. ユーザは “RELEVANCE の可視化” ボタンをクリックすると, 検索結果を可視化できる. 図 12 に提案する提示機構の可視化

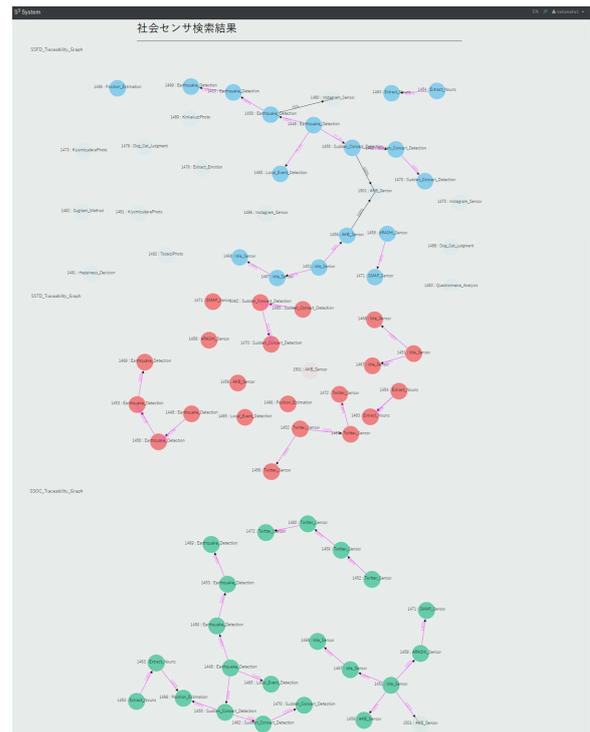


図 12 社会センサ可視化の例

の例を示す. 円が社会センサを表す節を示し, 視認性を高めるために色を付けている. 色はランダムに決定した. 作成履歴を表す枝において, ユーザが指定して把握した作成履歴は紫色で, ユーザが指定して把握した作成履歴以外は黒色で示している. 節と同様に, 色はランダムに決定した.

ユーザが類似度による作成履歴の表示制限機能を用いる場合,

“類似度を選択” ボタンをクリックし類似度を選択して、値と“以上”あるいは“以下”を選択することで指定された類似度の社会センサのみ表示される。

ユーザが検索結果の社会センサの作成時に参考にした社会センサの表示機能を用いる場合、“検索結果外の REFERENCE 表示：なし” ボタンをクリックし、“あり”あるいは“なし”を選択する。すると、検索結果に含まれる社会センサを濃い色で、検索結果外の REFERENCE 表示機能を用いて可視化された社会センサを薄い色で表示される。

ユーザが履歴ファイルの有無による作成履歴の表示制限機能を用いる場合、“申告：あり” ボタンをクリックし、“あり”、“なし”、“どちらも”のいずれかを選択することで、ユーザが指定して把握された作成履歴を表示するか、ユーザが指定して把握された作成履歴以外を表示するか、そのどちらも表示するかを選択できる。

5. 評価実験

提案する提示機構の有用性を検証するために、ユーザ実験を行う。本実験では、4章で実装したグラフ形式の社会センサの可視化を用いて、従来のリスト形式の表示よりも参考となる社会センサを発見しやすいか評価することを目的とする。

5.1 実験に用いた社会センサ

本節では、実験の実施に際し、S³システムで共有されている社会センサを説明する。まず、筆者らの研究グループで開発した表1に示す三つの社会センサを説明する。

社会センサ ID1

あらかじめ Twitter の REST API を用いて取得したツイートをデータベースに保存しておき、その保存したツイートを入力とする。ツイートの本文を Kuromoji^(注4)で形態素解析し、名詞を抽出して出力する。

社会センサ ID2

入力は社会センサ ID1 と同様である。文献[9]の実験で用いたクラスタリング手法の重心を出力する。

社会センサ ID3

あらかじめ Flickr の REST API を用いて取得した写真をデータベースに保存しておき、その保存した写真を入力とする。文献[10]で用いられた確信度算出手法の一つである積型撮影確信スコアを用いて写真の確信度を算出し、写真に付与し出力する。

実験のために、これらの社会センサを参考にして新たに11個の社会センサを作成して用いる。データソースの変更や解析方法の変更などにより新たな社会センサを作成した。作成した社会センサと参考にした社会センサ間の関連性を表2に示す。これらの詳細については本論文最後の付録に記述する。

5.2 実験内容

被験者は、20代の大学生および大学院生の男性10名である。実験の手順は以下のとおりである。

- (1) 提案する提示機構の説明および操作説明
- (2) 1分間の自由操作

表1 既存の社会センサ

社会センサID	入力	出力
1	DBから取得したTweet	形態素解析して出てきた名詞
2	DBから取得したTweet	クラスタリングして出てきたクラスタの重心
3	DBから取得したFlickerの写真	確信度が付与された写真

表2 新しく作成した社会センサの関連性

社会センサID	参考にした社会センサID	SSFD間の類似度 (%)
4	1 (SSFDを参考)	36.8
5	4 (SSFDを参考)	97.5
6	4 (SSFDを参考)	93.1
7	4 (SSTDを参考)	97.5
8	1 (SSFDを参考)	32.1
9	8 (SSFDを参考)	27.0
10	5 (SSFDを参考)	68.6
11	2 (SSFDを参考)	88.1
12	3 (SSFDを参考)	100
13	12 (SSTDを参考)	100
14	11 (SSTDを参考)	100

(3) 課題の実行

(4) アンケートの回答

提案する提示機構の説明および操作説明では、文章資料を用いて説明を行い、全ての被験者に同じ説明を行う。実験中適宜説明資料の閲覧を可能とする。その後、被験者に提案する提示機構の操作に慣れさせるために、1分間自由に操作させる。自由操作終了後、被験者にリスト形式とグラフ形式の検索結果を提示する。

新たな社会センサを作成する際、参考にした社会センサの数が多い社会センサが参考になりやすいと考えられる。参考にした社会センサとは、作成履歴に直接記述している社会センサだけでなく、作成履歴でつながった（作成履歴に直接記述している社会センサをいくつか経由して間接的に参考にした）社会センサも含む。ただし、類似度が低い場合には誤って作成履歴に記述している可能性がある。そこで、提案する提示機構を用いることで参考となる社会センサを発見しやすいか評価するために、参考にした社会センサの数が多い社会センサを発見する以下の課題を実施した。

課題1

リスト形式を用いて、検索結果の社会センサの中で、作成履歴（ただし、類似度が40%以下の社会センサ間の作成履歴を除く。）でつながった社会センサ数が最も多い社会センサIDを答えよ。

課題2

グラフ形式を用いて、検索結果の社会センサの中で、作成履歴（ただし、類似度が40%以下の社会センサ間の作成履歴を除く。）でつながった社会センサ数が最も多い社会センサIDを答えよ。

例えば、図11や図12に示すリスト形式およびグラフ形式の検索結果を提示する。実験では、順序効果を検証するために、被験者を2グループに分けて、課題に取り組む順序を変更した。具体的には、被験者1~5は課題1、課題2の順番で取り組ま

(注4) : <https://github.com/atilika/kuromoji/downloads>

表 3 実験アンケート

設問1	グラフ形式での可視化は社会センサの関連の把握の役に立ちましたか。 [+2(とても役に立った)~-2(全く役に立たなかった)]
設問2	グラフ形式とリスト形式のどちらが社会センサ間の関連の把握がしやすかったですか。 [+2(グラフ形式がとても発見しやすかった)~-2(リスト形式がとても発見しやすかった)]
設問3	設問3について理由を教えてください
設問4	その他、グラフ形式の可視化について意見をお願いします。

表 4 実験の課題の成否

被験者	リスト形式	グラフ形式
1	成功	成功
2	失敗	成功
3	成功	成功
4	失敗	成功
5	成功	成功
6	成功	失敗
7	成功	成功
8	成功	成功
9	成功	成功
10	失敗	成功

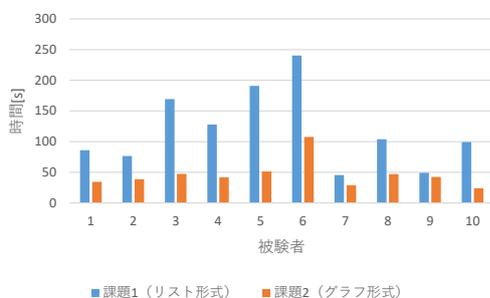


図 13 課題を完了するまでにかかった時間

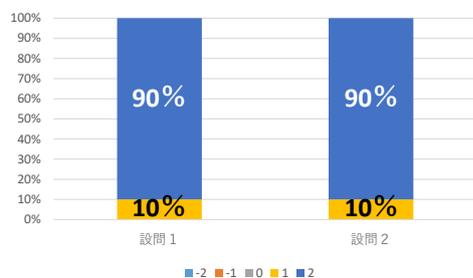


図 14 アンケート結果

せ、被験者 6~10 は課題 2、課題 1 の順番で取り組ませた。実験では、正しい社会センサ ID を回答した場合は成功、正しくない社会センサ ID を回答した場合は失敗とした。被験者が課題を完了するまでの時間を計測し、被験者には表 3 に示すアンケートに答えさせた。

5.3 実験結果

表 4 に被験者の実験課題の成否を示す。また、図 13 に被験者が課題を完了するまでにかかった時間を示す。図 13 より、全ての被験者において課題 1 に回答するまでにかかる時間は、課題 2 に回答するまでにかかる時間よりも長い。また、課題 2 に失敗した被験者が被験者全体の 1 割であったのに対して、課題

1 では、4 割の被験者が失敗しており、リスト形式では課題で求める社会センサを発見しにくいことがわかる。これらの結果から、本実験において、検索結果の社会センサを可視化し、グラフ形式で提示することにより、従来のリスト形式で提示することと比べて、課題で求める社会センサをより早く、正確に発見できることを確認した。

図 14 に実験で行ったアンケートの結果を示す。まず設問 1 において、全ての被験者がグラフ形式の可視化が役に立ったと選択している。次に設問 2 において、全ての被験者がリスト形式よりもグラフ形式の方が役に立ったと選択している。設問 2 の回答の理由を下記に示す。

- 関連性が図示されており、直感的に把握できたため。(9 人)
- リスト形式では、結局自分でグラフを作っているようであったため。(1 人)

多くの被験者が関連性が図示されており、直感的に関連性を把握できたためと回答した。また、リスト形式で関連性を把握する際に、結局自分で社会センサ間の関連についてのグラフを作成した被験者もいた。これらの結果より、本実験では、社会センサの提示方法はリスト形式よりもグラフ形式の方が役に立ったといえる。また、設問 4 の被験者の意見を下記に示す。

- 関連性を把握するだけならば、階層的に表示してほしい。
 - 類似度がかぶって見えなくなるところを改善してほしい。
 - ポップアップで詳細な情報が見れたらよりよい。
 - どの社会センサを閲覧したのかわかりにくいと思う。
- インタフェース関連の意見が多く、今後改善が必要である。

5.4 考察

評価実験により、提案する提示機構を用いて検索結果をグラフ形式で表示することで、従来のリスト形式で提示することと比べて、参考となる社会センサをより早く、正確に発見できることを確認した。しかし、「ポップアップで詳細な情報が見れたらよりよい。」とアンケート結果にあった通り、グラフ形式では関連性を提示するために、リスト形式ほどその他の詳細な情報を提示できない。さらに、「どの社会センサを閲覧したのかわかりにくいと思う。」とアンケート結果にある通り、リスト形式では検索結果が順番に並んでいるのに対して、グラフ形式では画面全体にバラバラに表示されているため、すでに確認した社会センサの把握が難しくなっている。また、社会センサ名や社会センサデータ数などの要素によってソートできず、全ての社会センサを確認したい場合などはリスト形式の方が使いやすいと考えられる。

6. まとめ

本研究では、社会センサの作成時に参考になる社会センサを発見しやすくするため、社会センサの関連性の提示機構の設計と実装を行った。提案する提示機構では、過去に新しい社会センサの作成のために参考にされた事実と、記述の類似度を社会センサの関連性と捉え、社会センサの関連性を可視化する。社会センサを節で示し、参考にした社会センサと枝で接続し、計算した類似度を枝付近に表示する。ユーザは社会センサの関連

性を確認することで、参考とする社会センサをより発見しやすくなる。また、社会センサの内容についてより詳しい情報を付加するためにタグを付与した。ユーザはタグを用いて、社会センサをより詳細に検索できる。

今後、一度閲覧した社会センサがわかる表示方法にしたり、類似度を強調したりなど、可視化機能の改善を考えている。

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) (JP26240013), 基盤研究 (B) (JP15H02702), 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。

文 献

- [1] K. Nakashima, M. Yokoyama, Y. Taniyama, T. Yohshihisa, and T. Hara, "S3 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.
- [2] 石原知也, 堀田圭佑, 肥後芳樹, 楠本真二, "再利用実績に基づいたコード片検索手法の提案," 電子情報通信学会技術研究報告 (SS2014-6), pp.61–66, 2013.
- [3] N. Kawamitsu, T. Ishio, T. Kanda, R.G. Kula, C. De Roover, and K. Inoue, "Identifying source code reuse across repositories using lcs-based source code similarity," Proc. IEEE Int'l Conf. Source Code Analysis and Manipulation (SCAM), pp.305–314, 2014.
- [4] 曾和寛史, 尾花将輝, 深海悟, "ソフトウェアメトリクスを用いたソースコード検索手法の提案," 第 77 回全国大会講演論文集, pp.309–310, 2015.
- [5] 川満直弘, 石尾隆, 井上克郎他, "Lsh アルゴリズムを利用した類似ソースコードの検索," 情報処理学会研究報告 (2016-SE-191), vol.31, no.3, pp.1–8, 2016.
- [6] K. Inoue, R. Yokomori, T. Yamamoto, M. Matsushita, and S. Kusumoto, "Ranking significance of software components based on use relations," IEEE Transactions on Software Engineering, pp.213–225, 2005.
- [7] C. Collberg, S. Kobourov, J. Nagra, J. Pitts, and K. Wampler, "A system for graph-based visualization of the evolution of software," Proc. 2003 ACM symposium on Software visualization, pp.77–ff, 2003.
- [8] S. Teller, "Data Visualization with D3.js", Packt Publishing Ltd, 2013.
- [9] 愛甲善之助, 谷山雄基, 横山正浩, 中嶋奎介, 義久智樹, 原 隆浩, "社会センサデータ生成・共有基盤におけるデータフロー制御機構," DICOMO2017, pp.1026–1033, 2017.
- [10] 谷山雄基, 義久智樹, 原 隆浩, 西尾章治郎, "指定位置を撮影した写真を撮影条件と地図情報により検索するシステム," 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO2015) シンポジウム論文集, pp.1301–1308, 2015.

付 録

社会センサ ID4

社会センサ ID4 は社会センサ ID1 を参考にして作成した。入力をデータベースではなく、Twitter の REST API から直接取得した新宿市役所から半径 10km 以内でツイートされたツイートとした。出力は社会センサ ID1 と同じである。

社会センサ ID5

社会センサ ID5 は、社会センサ ID4 を参考にして作成した。入力を Twitter の REST API から直接取得した大阪大学から半径 10km 以内でツイートされたツイートとした。出力は社会センサ ID4 と同じである。

社会センサ ID6

社会センサ ID6 は、社会センサ ID4 を参考にして作成した。入力を Twitter の REST API から直接取得した#センター試験というハッシュタグが付与されたツイートとした。出力は社会センサ ID4 と同じである。

社会センサ ID7

社会センサ ID7 は、社会センサ ID4 を参考にして作成した。入力と出力は社会センサ ID4 と同じだが、社会センサの実行間隔を変更した。

社会センサ ID8

社会センサ ID8 は、社会センサ ID1 を参考にして作成した。入力を Twitter の Streaming API から直接取得したツイートとした。出力は社会センサ ID1 と同じである。

社会センサ ID9

社会センサ ID9 は、社会センサ ID8 を参考にして作成した。入力をストリーミングデータ処理機構が取得したツイートとした。出力は社会センサ ID8 と同じである。

社会センサ ID10

社会センサ ID10 は、社会センサ ID5 を参考にして作成した。入力は社会センサ ID5 と同じである。ホットトピックを、取得したツイートから抽出される名詞をカウントし、カウント数が最大の名詞と定義し、出力をホットトピックとした。

社会センサ ID11

社会センサ ID11 は、社会センサ ID2 を参考にして作成した。入力を Twitter の REST API から直接取得した新宿市役所から半径 10km 以内でツイートされたツイートとした。出力を社会センサ ID2 と同様に、クラスタリングして出てきたクラスタの重心とした。

社会センサ ID12

社会センサ ID12 は、社会センサ ID3 を参考にして作成した。入力は社会センサ ID3 と同じである。出力として、文献 [10] で用いられた確信度算出手法のひとつである線形結合型撮影確信スコアを用いて写真の確信度を算出し、写真に付与した、確信度付写真とした。

社会センサ ID13

社会センサ ID13 は、社会センサ ID12 を参考にして作成した。入力および出力は社会センサ ID13 と同じであるが、社会センサの実行間隔を変更した。

社会センサ ID14

社会センサ ID14 は、社会センサ ID11 を参考にして作成した。入力および出力は社会センサ ID11 と同じであるが、社会センサの実行間隔を変更した。