

SNSの投稿を対象とした 鉄道における 改善点分析システム

学籍番号：1421102 氏名：齊藤 昭平

指導教員：鷹野 孝典 准教授

研究背景

鉄道を利用する人にとっては駅の設備，また環境がしっかりと整備されているかは重要である．

【駅の設備の例】

- エレベータやスロープなどのバリアフリー設備，乗り換え案内標識，トイレ，階段の数，通路の幅など

→ 改善点とともに，良いところがSNS上に投稿されることがある．

関連研究(1)

- **「Twitterに基づく都市鉄道の運転見合わせの発生状況の分析」** (堀江, 関谷, 金子, 土木学会論文集, 2015)

→ 首都圏で発生した運転見合わせを原因別に分析し, 鉄道会社のTwitterで発表されている運転再開見込み時間と実際に再開した時間との誤差を調査している.

- **「Twitter 上で共感を生み出すツイートの性質に関する考察」** (大川, 高間, 人工知能学会, 2012)

→ ツイートに対し多くのユーザが共感するケースに着目し, 発生メカニズムを解明を目指している.

- **「Linked Open Dataを用いた地域の防災支援システム」** (前田, 大場, 藤原, 情報処理学会, 2014)

→ 災害が発生したとき, 知りたい最新情報が取得できないことがあるため, 地図上に最新の被害状況などを集約し, 地図上に表示するシステムを提案している.

関連研究(2)

- **「Twitterにおける言及関係によるクラスタリングを利用したスパムアカウント判定手法の検討」** (菊池, 吉村, 情報処理学会, 2014)
 - Twitterのユーザの投稿を基に、クラスタ係数や普及数を指標したつながりの強弱を用いて、スパムアカウントの判定を行っている.
- **「Twitterにおけるテレビ番組の実況ツイートの可視化」** (菅野, 竹島, 加納, 東京工科大学, 2014)
 - 「#(番組名)」が付加されたツイートを取得し、番組前・中・後の感情を分析し、それを可視化する.
- **「Twitter上のあるユーザの意外な情報仲秋方法の提案」** (大原, 灘本, 甲南大学, 2014)
 - Twitterユーザの内容や属性データからTwitterユーザの近況の変化の情報を抽出する手法の提案を行う.

研究動機

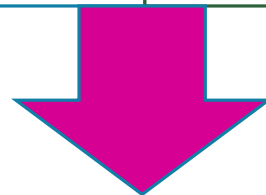
- **SNS上においては、改善点、良い点が投稿されることがある。**
- しかし、駅のレビューにおいては、駅ごとにしかなく、駅の特定の場所のレビューは存在しない。
- 特定の場所の投稿等を把握することによって、

Twitterの特徴:

- ・時系列順に内容が把握できる
- ・内容がシンプルなものが多い

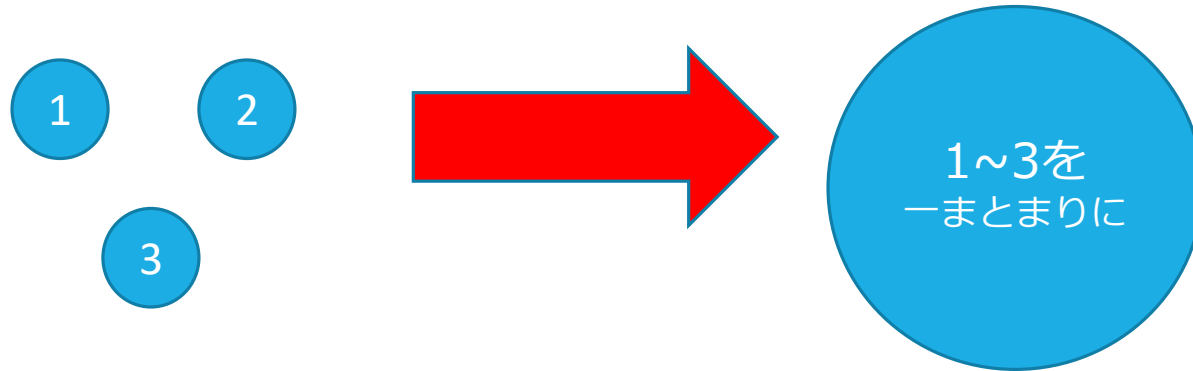
Facebookの特徴:

- ・外国人の投稿が多い
- ・駅のレビューなどで評価がされている



ある程度の利用者の投稿・レビューを把握することができ、今後の改善に向けていくことができる(改善点の場合).

研究動機(2)



このような，クラスタリングを行うことで，

本研究のアプローチ

- 同じ駅において、どういう改善を求めるツイートがあるか、種類ごとに提示しているサービスは無い(再掲).
- 同じ改善点を求めるツイートでも言い回しが異なることがある.
- **取得した投稿を基に、同じ部分での改善点や良い点の投稿数がどれほどあるかを同じような言い回しで分類分けを行う.**

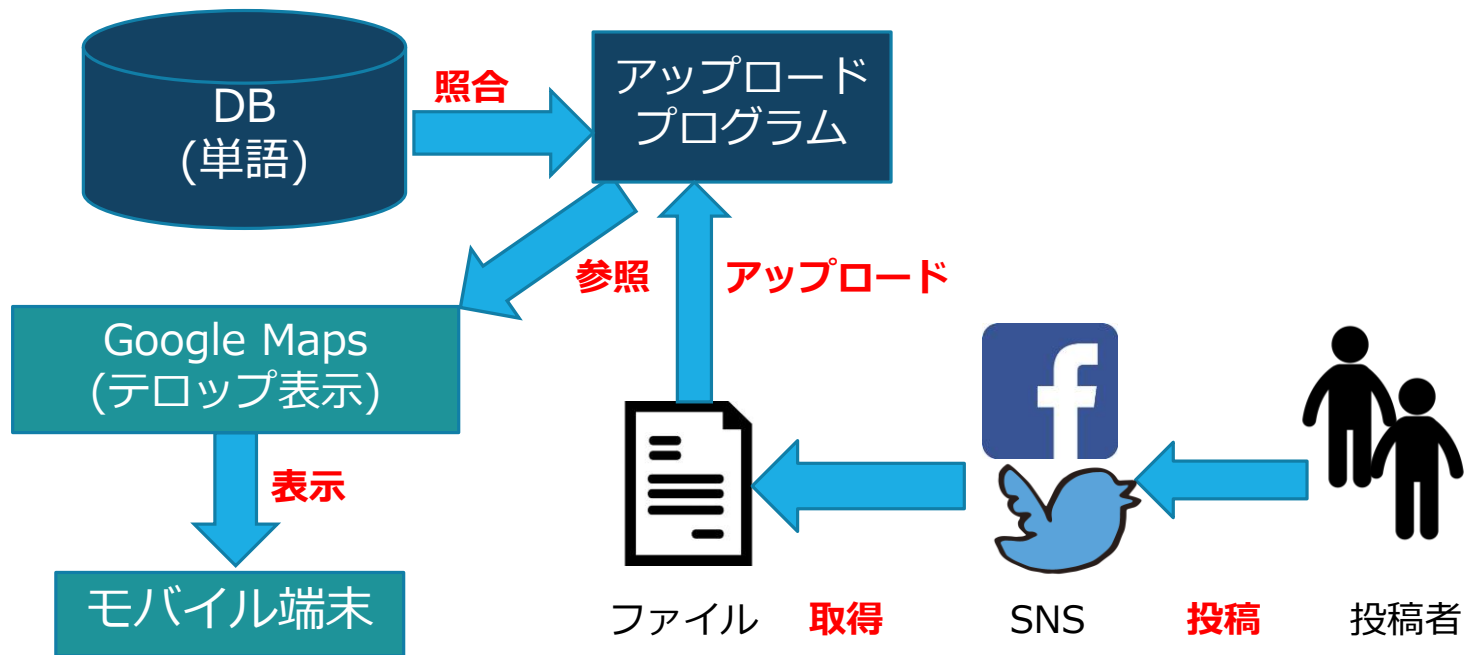


駅の良い点、改善すべき点が駅ごとに把握できるようになる.

言い回しが類似している言葉

- 分かりづらい、不明瞭、分からない
- 複雑、煩雑、ややこしい
- 狭い、狭苦しい、きつい
- きれい、清潔

提案システム



実装

- アップロードされたファイルとDBに格納した単語があるが照合させる.
- 取得した投稿を基に, どの箇所に改善の声, また良い点の投稿が集中しているかをクラスタリングを用いて分類を行う.
- Google Maps APIを用いてマッピングを行う.

→ 主にターミナル駅において, どのような部分を改善してほしいというツイート、また良いと思ったところのツイートがどれだけあったかをGoogleマップ上に吹き出しで表示していく.

テロップの一例)

〇〇駅の改善点
通路が狭い: 70件
案内が分かりにくい: 54件
...

〇〇駅の良い点
清潔: 60件
...

実験

- 最初にSNSの投稿を取得する.
- その投稿文の内容を,
-

単語の一例

改善点

狭い

混んでる

わかりづらい

複雑

場所

エレベータ

エスカレータ

改札

ホーム

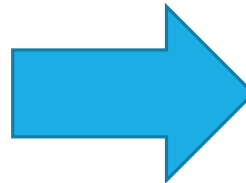
進捗(1)

まず、Twitter APIを用いたTwitter投稿検索サービスの「ついぽーと」で**205件のツイートを取得し**、テキストファイルで保存した。今回の検索キーワードは「駅 混んで」とした。そのなかで「混んでない」が含まれるツイートを抽出した(その他の検索キーワードでもデータを取得した)。

・ ここのトイレ混んでない
・ ここの乗り換え通路は混んでないからいい。

・ 今日凄く混んでない?
・ いつも空いているのに、今日に限って混んでない?

...



混んでないが含まれていたツイート…
38件(約18.5%)

空いている意味で用いられているツイート…
33件(38件のうちの86.8%)

混んでいる意味で用いられているツイート…
5件(38件のうちの13.2%)

↑取得したツイート(一部)

進捗(2)

「東京駅 わからない」と検索し、取得した投稿(抜粋)

- 東京駅ついたらいっつもどこで化粧したらいいかわからないどころも混んでいる(憤怒)
- ♪(^o^)♪ 東京駅着いたぞ (ゝ⊖ゝ)今から京葉線に乗り換えたいんだがどこにあるかわからない |乗り換え案内|

今後の予定

- SNSの投稿を随時取得する。また、今後取得した投稿の内容次第で、単語リストに単語を追加する。
- ファイルをアップロードできるプログラムを作成する(DBに格納した単語が含まれているツイートの件数をカウントさせる)。
- Google maps APIでマッピングをするプログラムを作成する。
- 関連度の算出方法
- 論文の執筆