

Twitter Streaming Location

60151039 Higashi Yunosuke

Twitter Streaming Locationとは

- LocationTrainerApp.java(学習)
 - まずジオタグ(位置情報)付のtweetをストリーミングから抽出し、分類器に学習させる。
 - Twitter Streaming Location.jsonで分類する。
- LocationClassifierApp.java(推定)
 - 学習したデータで、任意に入力した文章がどこで発信されたものであるかを推定する。

LocationTrainerApp.java

- 2点からできる四角形の中にジオタグが入っていたら,そのtweetをクラス分けをする

B:経度(138.946381), 緯度(35.52328)

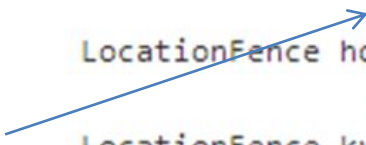


A:経度(138.946381), 緯度(35.523285)

LocationTrainerApp.java

```
public void trainTweets() throws Exception {
    LocationFence tokyo = new LocationFence("Tokyo", //
        138.946381, 35.523285, 139.953232, 35.906849);
    LocationFence hokkaido = new LocationFence("Hokkaido", //
        139.546509, 41.393294, 145.742798, 45.729191);
    LocationFence kyusyu = new LocationFence("Kyusyu", //
        129.538879, 31.147006, 131.856995, 33.934245);

    LocationFence[] locations = new LocationFence[] { tokyo, hokkaido,
        kyusyu };
    List<double[]> requestCoordinates = new ArrayList<double[]>();
    for (LocationFence l : locations) {
        requestCoordinates.addAll(Arrays.asList(l.getCoordinates()));
    }
}
```

A 

https://github.com/jubatus/jubatus-example/blob/master/twitter_streaming_location/java/src/main/java/example/twitterstreaminglocation/LocationTrainerApp.java

LocationTrainerApp.java

- 学習の実行 指定したフェンス内で呟かれたtweetだけを抽出し学習する。



```
root@localhost:/home/oss-wakayama/jubatus-example/twitter_streaming_ _ □ ×
ファイル(E) 編集(E) 表示(V) 検索 (S) 端末(T) ヘルプ(H)
Tokyo @RORO_RUK0 中国のお正月はいつ？
Kyusyu フロタのアルバム欲しいなー。お金無いから買えん(;_;)
Hokkaido @chanebi43 ん2753
Hokkaido @tb09095177574 いくか
Tokyo @yuuuusukeeeeeee この顔でいったんだけどなあ。 http://t.co/0HwxdcSLYc
Tokyo まって、、TシャツMってめっちゃ大きいやん
Tokyo 麻衣子とイブランチ♡(@ Chambres D'hotels HANA (シャンプルドット ハナ)) [pic]: http://t.co/v9Fc4hEVfp
Tokyo るく
Kyusyu @ahan208 冬休み北部ばかり～(T_T)
Tokyo @odoru002 よかったね
Tokyo おつドリ頂いて帰ります。

BARは新人なので、チャイラテ。全牛で注文(^_^) @ Starbucks Coffee 竹橋パレスサイド店 http://t.co/3uA64RckKr
Tokyo @ko_chin1115 http://t.co/forF06IaGj
Tokyo 来年のクリスマスは彼氏と過ごせるように頑張ろ～～～～～～～～。
Tokyo 会いたいんだが(笑)
Kyusyu 急行追い越し待ち @ 大橋駅にタッチ！ http://t.co/Tkotu0w9Yn
Kyusyu 本城 ZONEオワッテルから誰か入って。わら

もうハローワークいかんでよくなった♥FE0E
働いてこよっ！ http://t.co/khxETFicvF
```

LocationClassifierApp.java

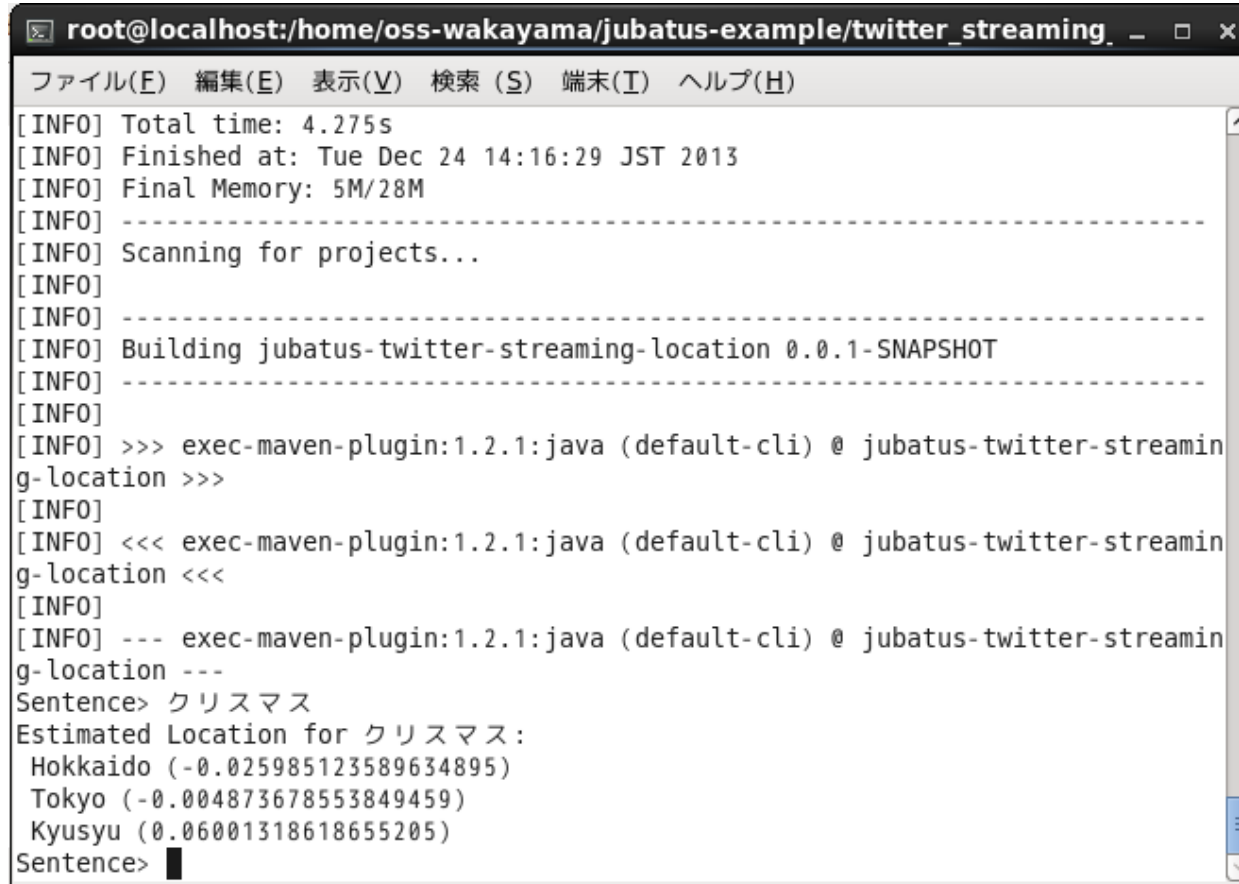
- 基本的に入力された文章に対して推定値を表示するだけ。

```
if (result.get(0).size() > 0) {  
    List<EstimateResult> est = result.get(0);  
    Collections.sort(est,  
        Collections.reverseOrder(new Comparator<EstimateResult>() {  
            public int compare(EstimateResult o1, EstimateResult o2) {  
                return (int) (o1.score - o2.score);  
            }  
        }));  
    System.out.println("Estimated Location for " + sentence + ":");  
    for (EstimateResult e : est) {  
        System.out.println("  " + e.label + " (" + e.score + ")");  
    }  
} else {  
    // No estimation results; maybe we haven't trained enough  
    System.out.println("No estimation results available.");  
    System.out.println("Train more tweets or try using another text.");  
}
```

https://github.com/jubatus/jubatus-example/blob/master/twitter_streaming_location/java/src/main/java/example/twittestreaminglocation/LocationClassifierApp.java

学習の実行

- 推定の実行



```
root@localhost:/home/oss-wakayama/jubatus-example/twitter_streaming_ _ □ ×
ファイル(E) 編集(E) 表示(V) 検索 (S) 端末(T) ヘルプ(H)
[INFO] Total time: 4.275s
[INFO] Finished at: Tue Dec 24 14:16:29 JST 2013
[INFO] Final Memory: 5M/28M
[INFO] -----
[INFO] Scanning for projects...
[INFO] -----
[INFO] Building jubatus-twitter-streaming-location 0.0.1-SNAPSHOT
[INFO] -----
[INFO] >>> exec-maven-plugin:1.2.1:java (default-cli) @ jubatus-twitter-streamin
g-location >>>
[INFO]
[INFO] <<< exec-maven-plugin:1.2.1:java (default-cli) @ jubatus-twitter-streamin
g-location <<<
[INFO]
[INFO] --- exec-maven-plugin:1.2.1:java (default-cli) @ jubatus-twitter-streamin
g-location ---
Sentence> クリスマス
Estimated Location for クリスマス:
Hokkaido (-0.025985123589634895)
Tokyo (-0.004873678553849459)
Kyusyu (0.06001318618655205)
Sentence> █
```

- クリスマスという文章は九州で発せられた可能性が高いと推定された

twitter_streaming_location.json

- "method": "NHERD"
- 重みづけ: bigram(2文字ずつ重みづけ)
- 正規化(idf)

https://github.com/jubatus/jubatus-example/blob/master/twitter_streaming_location/twitter_streaming_location.json