

売り上げの予測する Web アプリケーション開発

東京情報大学、総合情報学部、総合情報学科、
情報システム学系、4年 J19317 ドアン ホアイ ナム

目次

- 1. 内容の振り替え
- 2. 類似の研究
- 3. 進捗の状況
- 4. 現在の成果
- 5. 相関係数の分析
- 6. 問題点と取り組む予定
- 7. 論文提出の予定

研究内容

動機&目的: 賢いな売り上げを予測システムを作り出すことにより、シフト管理ができて、人件費を削減できる。

システムの構造: web and 予測システム。

人工知能部分: NUMPY, PANDAS, SCIKIT-LEARN (教師あり—重回帰)。

参考文献

- 外食産業の売上予測

https://www.jstage.jst.go.jp/article/jasmin/2012s/0/2012s_128/_pdf/-char/ja

- この論文から分かったポイント

1. 常に天気情報と売り上げの関係があるとは言えない。

2. 平日、休前日、休日の3区分を分けて、推論する。

- **平日**: 天気情報と売り上げの相関はみられず(いずれも有意水準が 10%以上)、前日の売上など過去の売上との相関が高いことがわかった。
- **休前日**: 夜の売り上げが高い、気温と売上に相関がある。
- **休日**: 休日は気温に加えて、降水量も売上との相関が高く、雨が強ければ強いほど、売上が低下する傾向にある。雨天の場合は外出を控える人の存在が影響していると考えられる。

進捗の状況

研究計画



プログラミング学習、データ収集

Web 設計プログラム作成、単体テスト

結合、結合テスト

論文作成

1年間のデータを修得済 ☒
Web 設計プログラム作成単体
テスト: 80%,
後サーバに上げる。

(WEB)現状

- FRONTEND: 完了。



BACKEND:80%完了



予測システム(相関係数一覧)

```
In [50]: #相関係数昇順
abs_cor = train_cor.map(abs)
abs_cor.sort_values(ascending = False)
```

```
Out[50]: customers      1.000000
Saturday    0.416916
Sunday      0.374563
Tuesday     0.299331
Thursday    0.282620
Monday      0.239051
arg_temperature 0.153688
rain_10      0.096630
rain_total   0.056573
Wednesday   0.027621
avg_wind     0.023284
corona_chiba 0.009439
corona_all   0.004104
Name: customers, dtype: float64
```

相関係数の絶対値	解釈
0.0~0.2	ほとんど相関関係がない
0.2~0.4	やや相関関係がある
0.4~0.7	かなり相関関係がある
0.7~1.0	強い相関関係がある

```
In [51]: #相関係数は外れ値に影響しやすい。
#そのため 散布図→外れ値→相関係数
```

予測精度=決定係数 = 0.32

```
avg_temperature    1.002424  
rain_total         1.002424  
Wednesday          1.002424  
dtype: float64
```

```
In [70]: #正解データも標準化します  
sc_model_y = StandardScaler()  
sc_model_y.fit(y_train)  
sc_y = sc_model_y.transform(y_train)
```

```
In [74]: model = LinearRegression()  
model.fit(sc_x, sc_y)
```

```
Out[74]: LinearRegression()
```

```
In [72]:
```

```
Out[72]: -29.75324183995554
```

```
In [75]: #検証データも標準化する必要があります。  
sc_x_val = sc_model_x.transform(x_val)  
sc_y_val = sc_model_y.transform(y_val)  
# 標準化した検証データで決定係数を計算する  
model.score(sc_x_val, sc_y_val)
```

```
Out[75]: 0.3202149750307055
```


問題点と取り組む予定

- 問題点:

精度がひくい

人工知能の部分をウェブアプリに結びつく方法、相関係数が0.3以下、天気情報、コロナ感染者数との関係が期待したが、低かった。他の前処理がうまくいかないかと考えている。

これから取り組み:

前処理をやり直す。

データをもう1年間増やす。

緊急事態宣言期間の有無データを導入。

ウェブサイトをサーバーに上げる。

平日、休前日、休日の3区分に分けて、予測する。

卒業論文提出の予定

- 現在:他の論文を読み、論文の形式を考察しています。
- 来週から試行錯誤をしながら論文を書き始める予定です。
- 1月の中旬に完成する予定です。

ご清聴ありがとうございました。
