The Customer Revisit Prediction
Using Macroscale Mobility Information

Beomyoung Kim (dglidgli@kaist.ac.kr)
Department of Industrial and Systems Engineering, KAIST

Introduction

Motivation
There is a research that successfully predicts the revisit of customers by using only movement data collected in a specific offline store through wireless fingerprinting technology [1]. Companies that have macroscale mobility information, such as Loplat X, also want application in a variety of areas, such as target marketing and customer segmentation through revisit prediction.

Purpose
1) To present model for predicting revisit time for specific place categories or brand rather than a single store.
2) To solve the data scarcity problem, which is not known whether the customer visited after that time when using data up to a certain point for training model.
3) To define features and embedding factors related to revisit in macroscale mobility information and to extend the WTTE-RNN methodology [2].

Method

Loss Function
When censored observations, maximize the probability over the right of that point. When uncensored observations, maximize the density function of that point [2].

\[
\max_y f(y|\theta) \\
\max_y \int_0^\infty f(s|\theta)ds
\]

Model

\[
\sum_{n=1}^N \sum_{t=0}^{t_0} u_i^n \log[Pr(\hat{Y}_t^n = y_1^n|x_{0,1}^n)] + (1 - u_i^n) \log[Pr(\hat{Y}_t^n > y_1^n|x_{0,1}^n)]
\]

X_id
X_category
X_day
X_hour
X_continuous
X_user
Emb.
Emb.
Emb.
Emb.
Emb.
LSTM
Dense
α
β
Assume the function of revisit time follows Weibull distribution.

Features

Embedding Features
1) Category 2) Day of the week 3) Hour 4) Customer Id
Continuous Feature
1) Time until next any log
Aggregated Features
* Also used in Cox Model
1) The number of interest category over the entire period
2) The number of all logs
3) The number of days with logs
4) The number of all logs divided by the days with logs
5) Time from last visit of interest category
6) The number of weekend over the entire period
7) Average interval between interest category
8) Average interval between all logs

Results of “Coffee Shop”

<table>
<thead>
<tr>
<th></th>
<th>Cox</th>
<th>WTTE-RNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-index</td>
<td>0.698</td>
<td>0.726</td>
</tr>
<tr>
<td>RMSE</td>
<td>205.24</td>
<td>199.49</td>
</tr>
<tr>
<td>Non-returning recall</td>
<td>0.488</td>
<td>0.438</td>
</tr>
<tr>
<td>Non-returning F1</td>
<td>0.543</td>
<td>0.555</td>
</tr>
</tbody>
</table>

Ours has less error in actual revisit time and can classify for customers who do not revisit compared with cox which is one of the existing survival analysis models.

Reference