SCIENCE ▪ PASSION ▪ TECHNOLOGY

Understanding the Impact of Weather for POI Recommendations

Christoph Trattner, Alex Oberegger, Lukas Eberhard, Denis Parra, Leandro Marinho, Know-Center@Graz University of Technology

September 15, 2016

Content

1. Motivation
2. Dataset
3. Empirical Analysis
4. Weather Aware POI Recommender (WPOI)
5. Conclusions & Future Work

Motivation

- Geolocation services on portable devices
- Location-based social networks (LBSN)
  - Yelp
  - Foursquare
- Vast amount of check-in data
  - Comments on places
  - Recommendations on places
  - Ratings on places
- Use data to assist users

- Context aware recommender systems (CARS) concentrated on
  - Social context
  - Geographical context
  - Time context
- No research on the impact of weather
  - Pool not popular at rainy conditions
  - Ice cream shop not popular at cold temperatures

Problem Statement

Definition

Given a user u, the user's check-in history L^u, i.e., the POIs that the user has visited in the past, and the current weather context c the aim is to predict the POIs $\hat{L}^u = \{l_1, \ldots, l_{|L|}\}$ that the user will likely visit in the future that are not in L^u.

Research Questions

- **RQ1** Are the users' mobility patterns influenced by weather?
- **RQ2** How can weather context information be incorporated into existent recommender systems?
- **RQ3** To what extent can weather information be used to increase recommender accuracy?
- **RQ4** Which weather feature provides the highest impact on the recommender accuracy?

Dataset

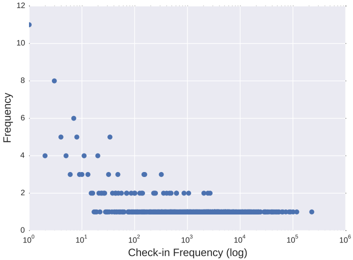
- Foursquare Check-in Dataset from Dingqi Yang [3][4]
- Worldwide check-ins from April 2012 to September 2013
- Filtered by U.S. cities
  - ~3,000,000 Check-ins
  - ~500,000 Venues
  - ~50,000 Users
  - 60 Cities

Dataset

- Weather Data from forecast.io weather API [1]
  - One API call per $<$venue, city, time$>$ triple
  - $\sim$27,000 API calls
  - Eight weather attributes
    - Visibility, Precipitation intensity, Humidity, Cloud cover, Pressure, Windspeed, Temperature, Moonphase
- Outdated category information needed recrawling

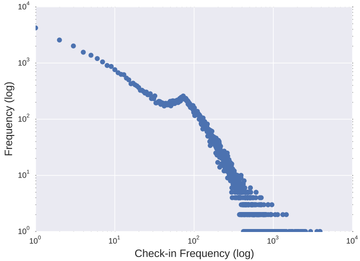
Check-in Distributions

Categories



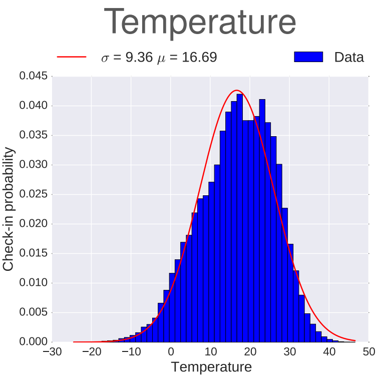
$\sim$200 main categories

User

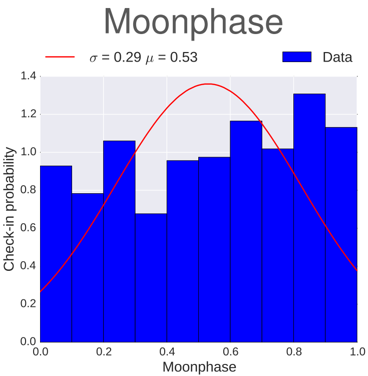


$\sim$11,000 user with $>$100 check-ins

Weather Distributions

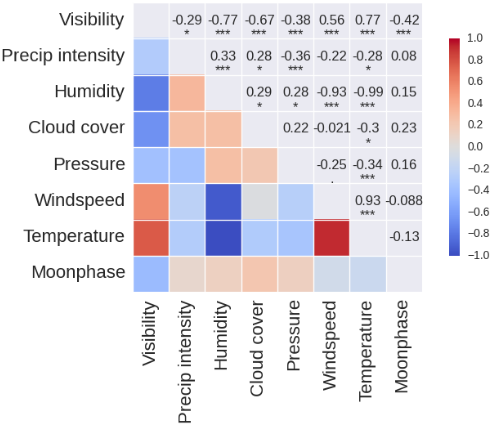


Normal distributed



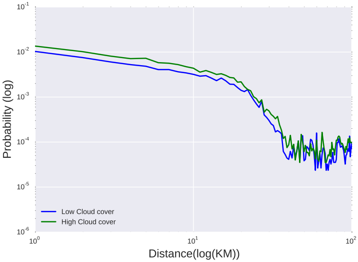
Uniform distributed

Correlation between weather features

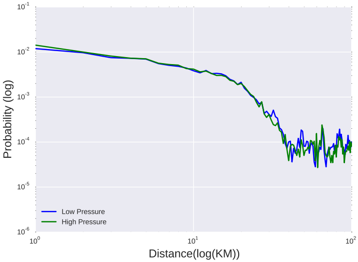


Impact on Travel Distance

Cloud Cover

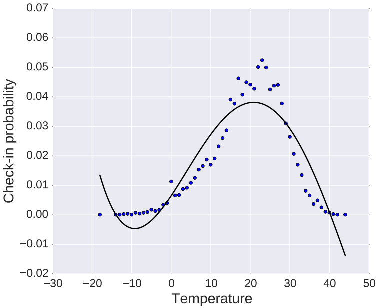


Pressure

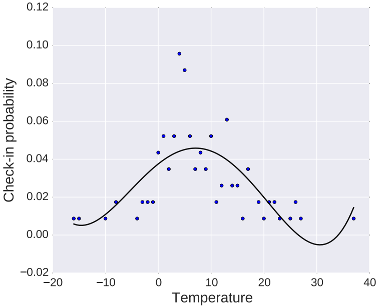


Impact on Category Popularity

Ice Cream Shop

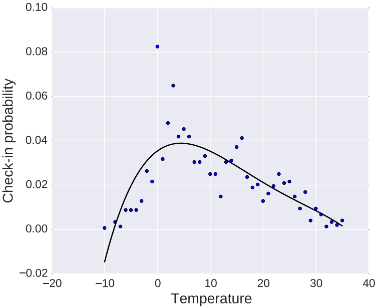


Ski Area

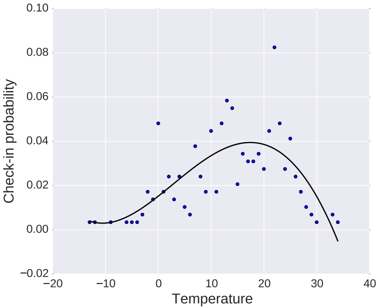


User Mobility Patterns

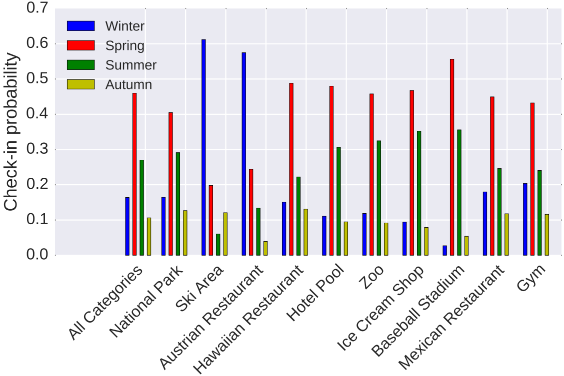
Mobility pattern of a "frosty" user



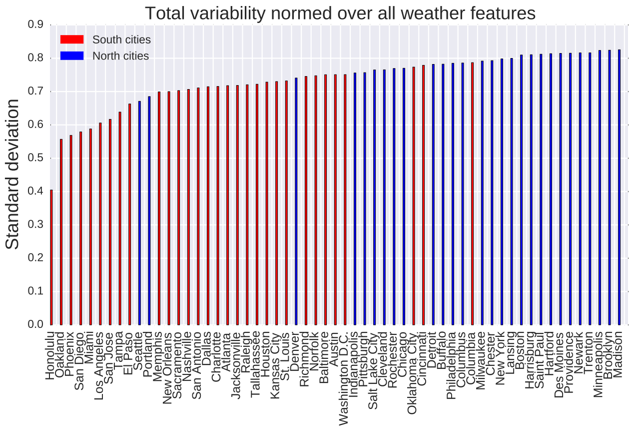
Mobility pattern of a "heated" user



Seasonal Impact on Categories



Regional Weather Variability



Weather Aware POI Recommender (WPOI)

- Many basic recommender algorithms available
  - Most Popular (MP)
  - KNN-Algorithms (User, Item)
  - Matrix Factorization (MF)
- Enrich existing recommender system with weather context
- Test recommender accuracy in four cities representing variety of climate (cardinal directions)
  - Minneapolis
  - Boston
  - Miami
  - Honolulu

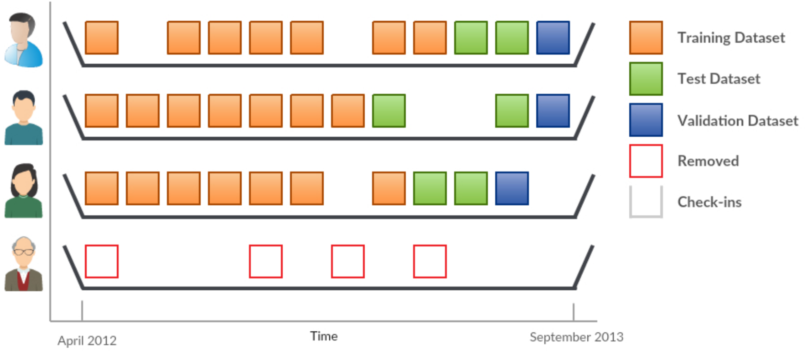
- Rank-GeoFM by Li et al. (2015) [2]
    - Based on MF
    - Context aware (Geo and Time)
    - Easy to extend
    - High accuracy
- Based on iterative learning of latent model parameters (weights)
- Learns model parameters with stochastic gradient descent (SGD)

$$y_{ult} = \underbrace{u_u^{(1)} \cdot l_l^{(1)}}_{\text{user preference score}} + \underbrace{t_t^{(1)} \cdot l_l^{(2)}}_{\text{time preference score}} + \underbrace{u_u^{(2)} \cdot \sum_{l* \in N_k(l)} w_{ll*} l_{l*}^{(1)}}_{\text{geographical influence score}} + \underbrace{l_l^{(3)} \cdot \sum_{t* \in \mathcal{T}} m_{tt*} t_{t*}^{(1)}}_{\text{temporal influence score}}$$

- $U^{(1,2)}, L^{(1,\dots,3)}, T^{(1)} \rightarrow$ latent model parameters to learn
- wll is the probability that l is visited given that $l*$ has been visited in terms of geographical position and mtt is the probability that the popularity of a POI in time slot t is influenced by those in time slot $t*$.

- Replace time with weather feature
- Start algorithm for each weather feature separately
- Measure accuracy and compare with RankGeoFM and base algorithms

$$y_{ulc} = \underbrace{u_u^{(1)} \cdot l_l^{(1)}}_{\text{user preference score}} + \underbrace{f_c^{(1)} \cdot l_l^{(2)}}_{\text{weather preference score}} + \underbrace{u_u^{(2)} \cdot \sum_{l^* \in N_k(l)} w_{ll^*} l_{l^*}^{(1)}}_{\text{geographical influence score}} + \underbrace{l_l^{(3)} \cdot \sum_{c^* \in FC} wi_{cc^*} f_{c^*}^{(1)}}_{\text{weather influence score}}$$


Training Dataset
Test Dataset
Validation Dataset
Removed
Check-ins
April 2012
Time
September 2013

Algorithm 1: WPOI

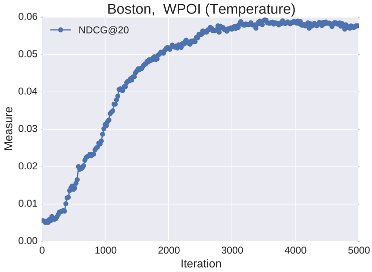
Input: check-in data D, hyperparameters that steer context influence and learning rate

Init: Initialize Θ with $\mathcal{N}(0, 0.01)$; Shuffle D

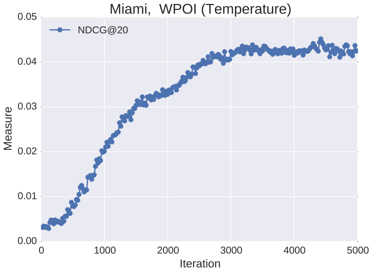
```
repeat
    for (u, l, c) ∈ D do
        Compute y_ulc and n = 0
        repeat
            Sample a POI l′ and feature class c′, Compute y_ul′c′ and set n++
        until (u,l’,c’) ranked inccorrectly || everything ranked correct
        if (u,l’,c’) ranked inccorrectly then
            update latent model parameters Θ
            according to the gradient of the error function.
        end
    end
until convergence
return Θ = {L^(1), L^(2), L^(3), U^(1), U^(2), F^(1)}
```

Weather Aware POI Recommender (WPOI)

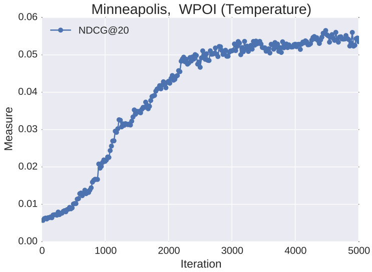
Boston



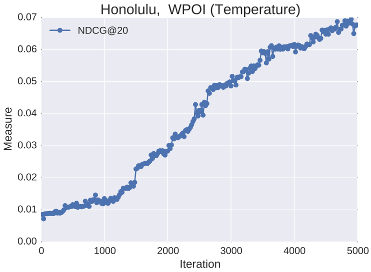
Miami



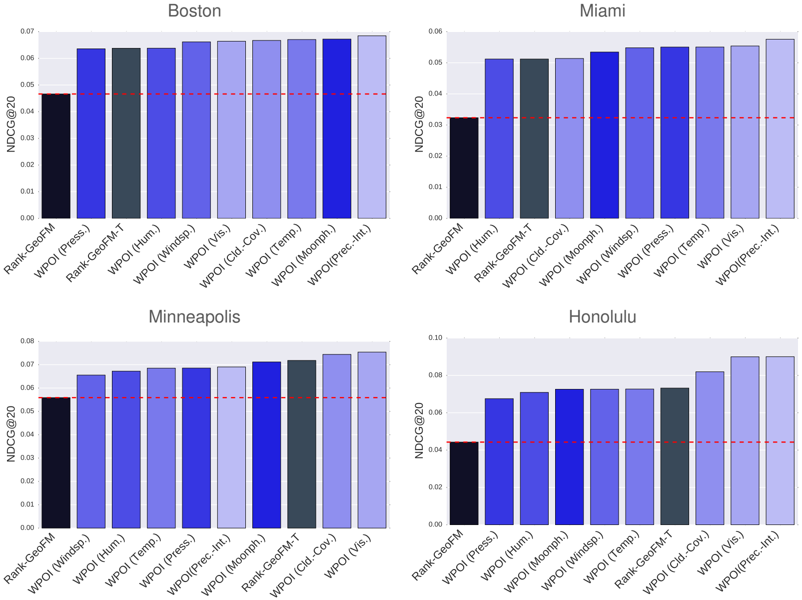
Minneapolis



Honolulu



Weather Aware POI Recommender (WPOI)



Conclusions

Empirical Analysis

- Weather features have little impact on travel distance
- Seasonality has an impact on category popularity
- Category popularity is dependent on the region
- Higher weather variability in the north

WPOI

- Weather context more useful than time
- WPOI better in regions closer to tropical zone
- Precipitation intensity and visibility improve accuracy best
- Weather context is indeed a useful contextual information in POI recommender systems

Future Work

- Only one weather feature at a time
- Incorporate travel distance probabilities under different weather conditions
- Incorporate user sensitivity to weather conditions

Questions?

References I

[1] FORECAST.IO.

forecast.io weather api, 2015.

[2] LI, X., CONG, G., LI, X.-L., PHAM, T.-A. N., AND KRISHNASWAMY, S.

Rank-geofm: A ranking based geographical factorization method for point of interest recommendation.

In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2015), SIGIR '15, ACM, pp. 433–442.

References II

[3] YANG, D., ZHANG, D., CHEN, L., AND QU, B.

Nationtelescope: Monitoring and visualizing large-scale collective behavior in lbsns.

Journal of Network and Computer Applications 55 (2015), 170–180.

[4] YANG, D., ZHANG, D., AND QU, B.

Participatory cultural mapping based on collective behavior in location based social networks.

ACM Transactions on Intelligent Systems and Technology (2015).

in press.

Weather feature	Properties	Range in dataset
Precipitation intensity	Precipitation intensity measured in milimeters of liquid water/hour.	$0mm/h - 34,29mm/h$
Temperature	Temperature measured in degree Celsius	$-24,48^{\circ} - 46,58^{\circ}$
Wind speed	Wind speed measured in meters/second	$0m/s - 19,13m/s$
Cloud cover	Value between 0 and 1 displaying the percentage of the sky covered by clouds.	$0 - 1$

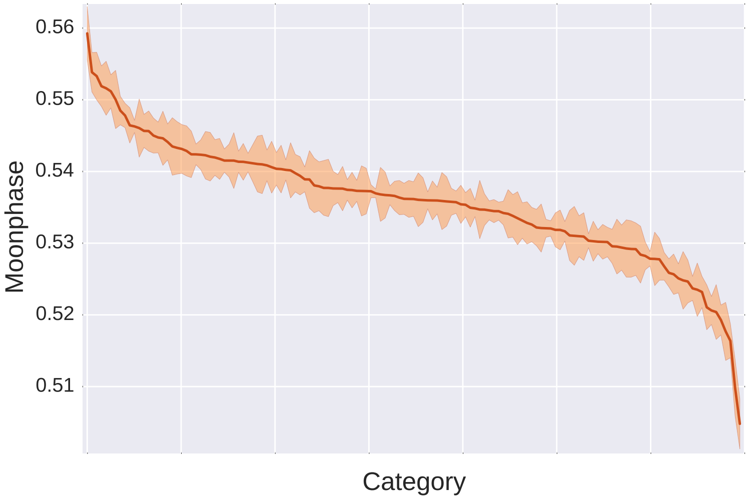
Humidity	Value between 0 and 1 representing the "Percentage relative humidity" is defined as the partial pressure of water vapor in air divided by the vapor pressure of water at the given temperature."	$0,02\phi - 1,00\phi$
Pressure	Atmospheric pressure measured in hectopascals.	$957,11hPa - 1046,05hPa$
Visibility	Value representing the average visibility in kilometers capped at 16,09	$0km - 16,09km$
Moonphase	Value from 0 to 1 representing the range between new moon and full moon	$0 - 1$

Correlations

Humidity	Temperature	-0.99	Relative humidity represents the saturation of moisture in the air and cold air does not need that much moisture to be saturated.
Windspeed	Temperature	0.93	Foehn-effect, Windchill factor not included
Humidity	Windspeed	-0.93	See positive correlation between windspeed and temperature and negative between humidity and temperature.

Correlations

Temperature	Visibility	0.77	At colder temperatures saturation is reached earlier and therefore fog occurs more often that blurs visibility.
Humidity	Visibility	-0.77	High humidity blurs visibility
Cloud cover	Visibility	-0.67	Clouds cover the sun → visibility is low


0.56
0.55
0.54
0.53
0.52
0.51
Moonphase
Category

Algorithm 2: WPOI

Input: check-in data D, geographical influence matrix W, weather influence matrix WI, hyperparameters $\epsilon, C, \alpha, \beta$ and learning rate γ_g and γ_w

Output: parameters of the model $\Theta = \{L^{(1)}, L^{(2)}, L^{(3)}, U^{(1)}, U^{(2)}, F^{(1)}\}$

init: Initialize Θ with $\mathcal{N}(0, 0.01)$; Shuffle D
repeat
    for $(u, l, c) \in D$ **do**
        Compute y_{ulc} and n = 0
        repeat
            Sample a POI l' and feature class c', Compute $y_{ul'c'}$ and set n++
        until $I(x_{ulc} > x_{ul'c'})I(y_{ulc} < y_{ul'c'} + \epsilon) = 1$ or $n > |L|$
        if $I(x_{ulc} > x_{ul'c'})I(y_{ulc} < y_{ul'c'} + \epsilon) = 1$ **then**
            $\eta = E\left(\left\lfloor \frac{|L|}{n} \right\rfloor\right) \delta_{ull'}$
            $g = \left(\sum_{c^* \in FC_f} wi_{c'c^*} f^{(1)}_{c*} - \sum_{c^+ \in FC_f} wi_{cc^+} f^{(1)}_{c+}\right)$
            $f^{(1)}_c \leftarrow f^{(1)}_c - \gamma_w \eta (l^{(2)}_{l'} - l^{(2)}_l)$
            $l^{(3)}_l \leftarrow l^{(3)}_l - \gamma_w \eta g$
            $l^{(2)}_{l'} \leftarrow l^{(2)}_{l'} - \gamma_w \eta f_c$
            $l^{(2)}_l \leftarrow l^{(2)}_l + \gamma_w \eta f_c$
        end
    end
    Project updated factors to accomplish constraints
until convergence
return $\Theta = \{L^{(1)}, L^{(2)}, L^{(3)}, U^{(1)}, U^{(2)}, F^{(1)}\}$

Symbol	Description
$\mathcal{U}$	set of users $u_1, u_2, ..., u_{\|U\|}$
$\mathcal{L}$	set of POIs $l_1, l_2, ..., l_{\|L\|}$
FC_f	set of classes for feature f
F	set of weather feature classes $f_1, f_2, ..., f_{\|FC_f\|}$
Θ	latent model parameters containing the learned weights $\{L^{(1)}, L^{(2)}, L^{(3)}, U^{(1)}, U^{(2)}, F^{(1)}\}$ for locations, users and weather features
X_{ul}	$\|U\|x\|L\|$ matrix containing the check-ins of users at POIs
X_{ulc}	$\|U\|x\|L\|x\|FC_f\|$ matrix containing the check-ins of users at POIs at a specific feature class c
D_1	user-POI pairs: $(u, l)\|x_{ul} > 0$
D_2	user-POI-feature class triples: $(u, l, c)\|x_{ulc} > 0$
$d(l, l')$	geo distance between the latitude and longitude of l and l'

W	geographical probability matrix of size $\|L\|x\|L\|$ where $w_{ll'}$ contains the probability of l' being visited after l has been visited according to their geographical distance. $w_{ll'} = (0.5 + d(l, l'))^{-}1)$
WI	probability that a weather feature class c is influenced by feature class c'. $wi_{cc'} = \frac{\sum_{u \in U} \sum_{l \in L} x_{ulc} x_{ulc'}}{\sqrt{\sum_{u \in U} \sum_{l \in L} x_{ulc}^2} \sqrt{\sum_{u \in U} \sum_{l \in L} x_{ulc'}^2}}$
$N_k(l)$	set of k nearest neighbors of POI l
y_{ul}	the recommendation score of user u and POI l
y_{ulc}	the recommendation score of user u, POI l and weather feature class c
$I(\cdot)$	indicator function returning $I(a) = 1$ when a is true and 0 otherwise
ϵ	margin to soften ranking incompatibility
$Incomp(y_{ulc}, \epsilon)$	a function that counts the number of locations $l' \in \mathcal{L}$ that should be ranked lower than l at the current weather context c and user u but are ranked higher by the model.
γ_w	learning rate for updates on weather latent parameters.

γ_g	learning rate for updates on latent parameters from base approach.
$E(\cdot)$	a function that turns the rating incompatibility $Incomp(y_{ulc}, \epsilon)$ into a loss. $E(r) = \sum_{i=1}^{r} \frac{1}{i}$
$\mathcal{O}$	objective function to minimize during the iterative learning. $\mathcal{O} = \sum_{(u,l,c) \in D2} E(Incomp(y_{ulc}, \epsilon))$
$s(a)$	sigmoid function $s(a) = \frac{1}{1+exp(-a)}$
$\delta_{ucll'}$	function to approximate the indicator function with a continuous sigmoid function. $\delta_{ucll'} = s(y_{ul'c} + \epsilon - y_{ulc})(1 - s(y_{ul'c} + \epsilon - y_{ulc}))$
$\lfloor \frac{\lvert L \rvert}{n} \rfloor$	if the n^{th} location l' was ranked incorrect by the model the expactation is that overall $\lfloor \frac{\lvert L \rvert}{n} \rfloor$ locations are ranked incorrect.
$\frac{\partial \overline{E}}{\partial \Theta}$	calculation of stochastic gradient $= E\left(\lfloor \frac{\lvert L \rvert}{n} \rfloor\right) \delta_{ucll'} \frac{\partial(y_{ucl'} + \epsilon - y_{ucl})}{\partial \Theta}$
$\Theta \leftarrow \Theta - \frac{\partial \overline{E}}{\partial \Theta}$	SGD based optimization of the latent model parameters

Dataset Statistics

City	#Users	#Venues	#Check-ins	Sparsity
Minneapolis	436	797	37,737	89.1%
Boston	637	1141	42,956	94.3%
Miami	410	796	29,222	91.0%
Honolulu	173	410	16,042	77.4%

Formulas

$$SE_{\overline{x}} = \frac{SD}{\sqrt{n}} \qquad (1)$$