

# POI Atmosphere Categorization Using Web Search Session Behavior

---



Kota Tsubouchi,  
Yahoo Japan Corporation



Hayato Kobayashi,  
Yahoo Japan Corporation



Toru Shimizu  
Yahoo Japan Corporation



# AGENDA

---

01

## Background

About POI atmosphere categorization

02

## Proposed Method

LSTM with DSSM training

03

## Experiment

POI categorization,  
POI atmosphere categorization

04

## Conclusion

Take home messages





# Background

About POI atmosphere categorization



“Park”

(Park Image 1)

(Park Image 2)

(Park Image 3)



# POI Atmosphere Categorization

- POI categorization that can distinguish "atmospheres"

How to realize it?



## Existing Research on POI categorization



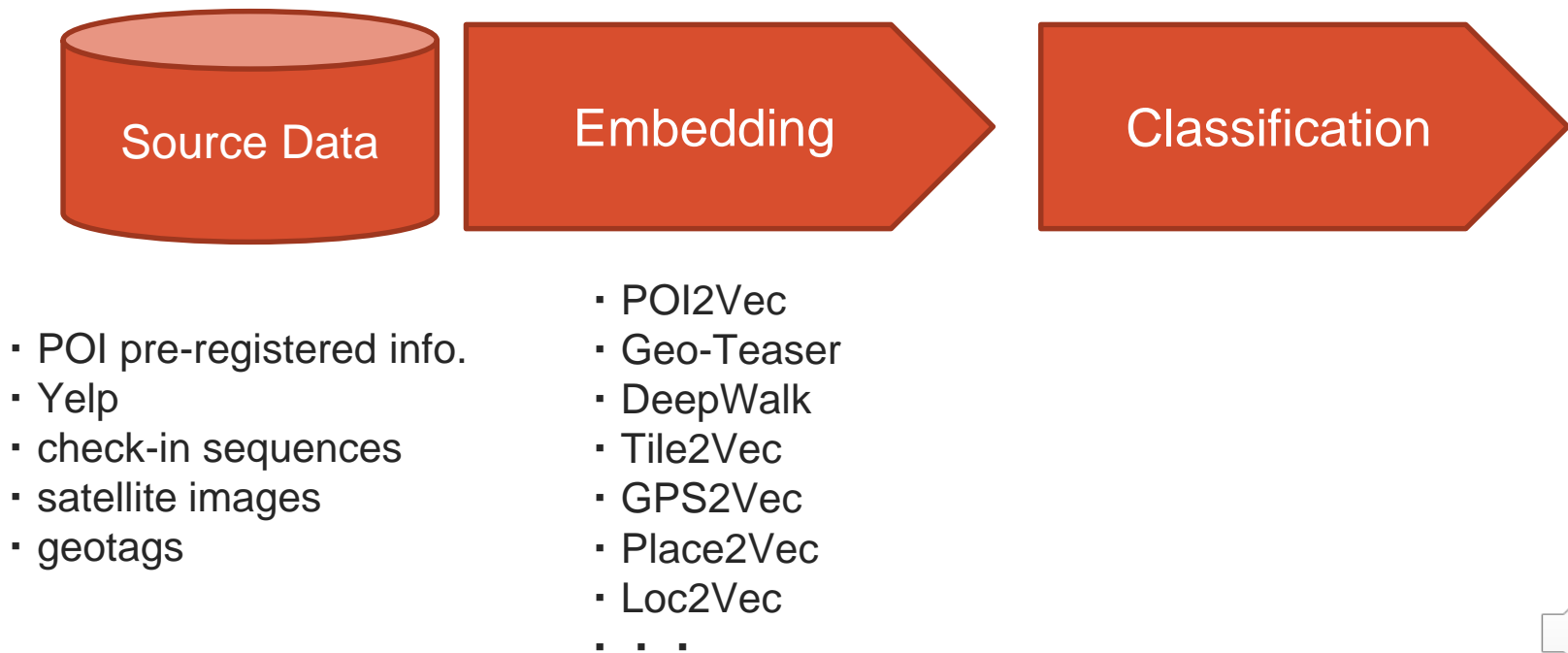
## Existing Research on POI categorization



- POI pre-registered info.
- Yelp
- check-in sequences
- satellite images
- geotags



## Existing Research on POI categorization





## Existing Research on POI categorization

Source Data

- POI pre-registered info.
- Yelp
- check-in sequences
- satellite images
- geotags

Embedding

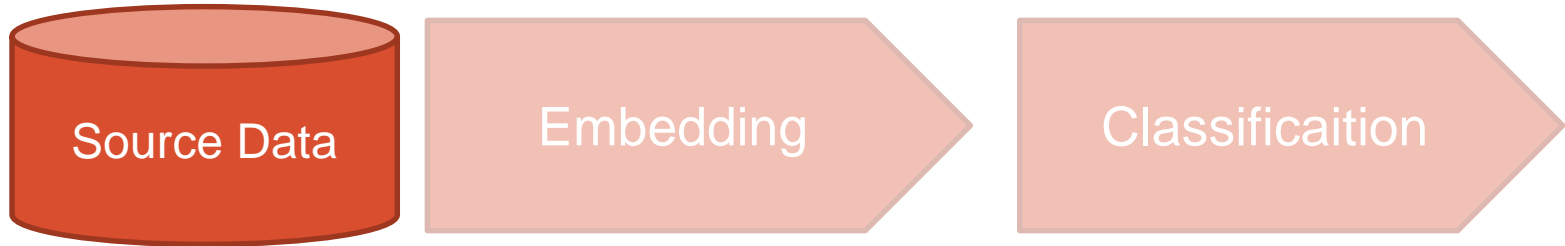
- POI2Vec
- Geo-Teaser
- DeepWalk
- Tile2Vec
- GPS2Vec
- Place2Vec
- Loc2Vec
- . . .

Classification

- k-means
- t-SNE
- DBSCAN
- . . .



# Existing Research on POI categorization



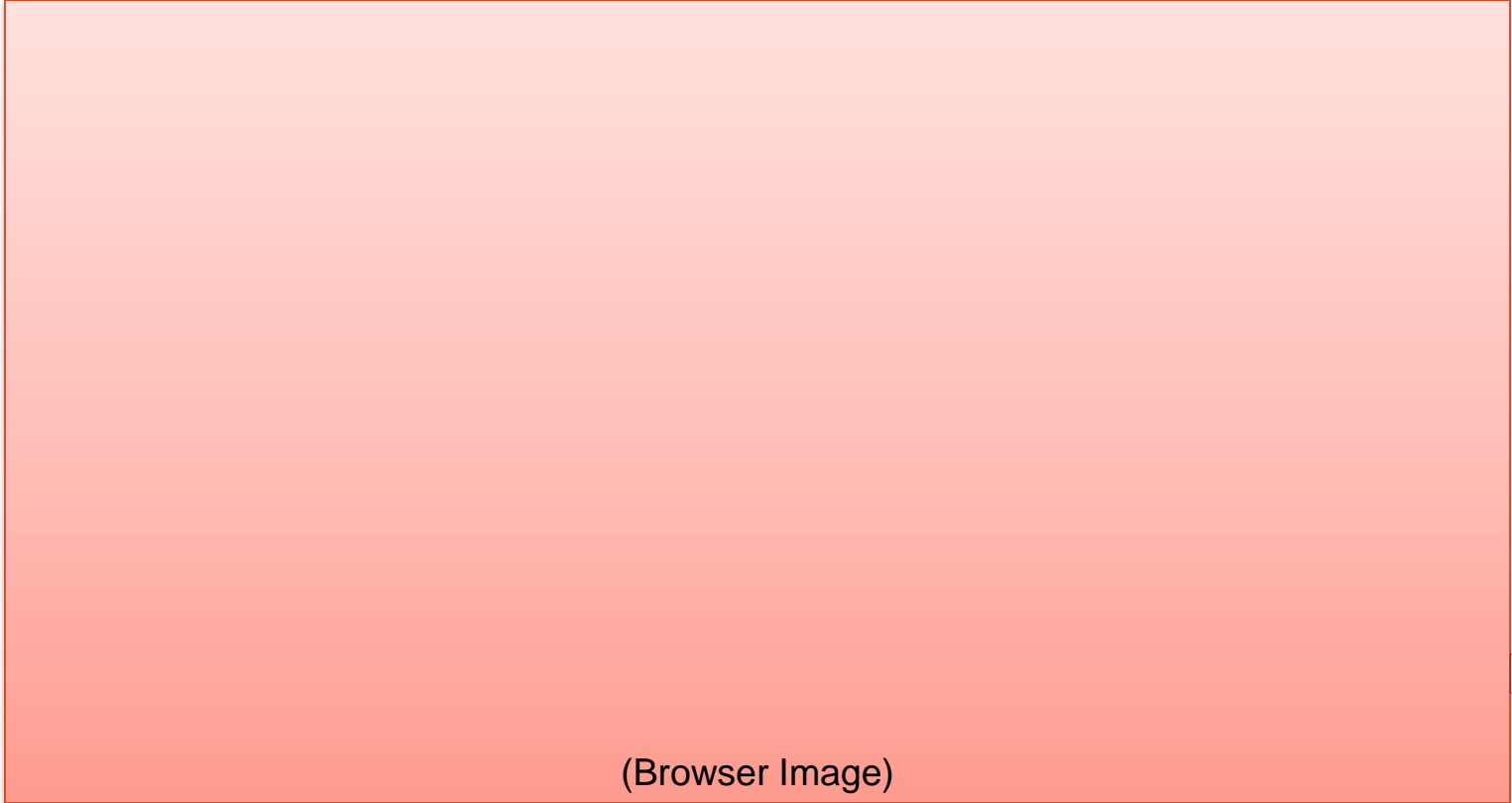
- POI pre-registered info.
- Yelp
- check-in sequences
- satellite images
- geotags

not relevant to the atmosphere of POI

- POI2Vec
- Geo-Teaser
- k-means
- t-SNE
- DBSCAN
- . . .
- Place2Vec
- Loc2Vec
- . . .



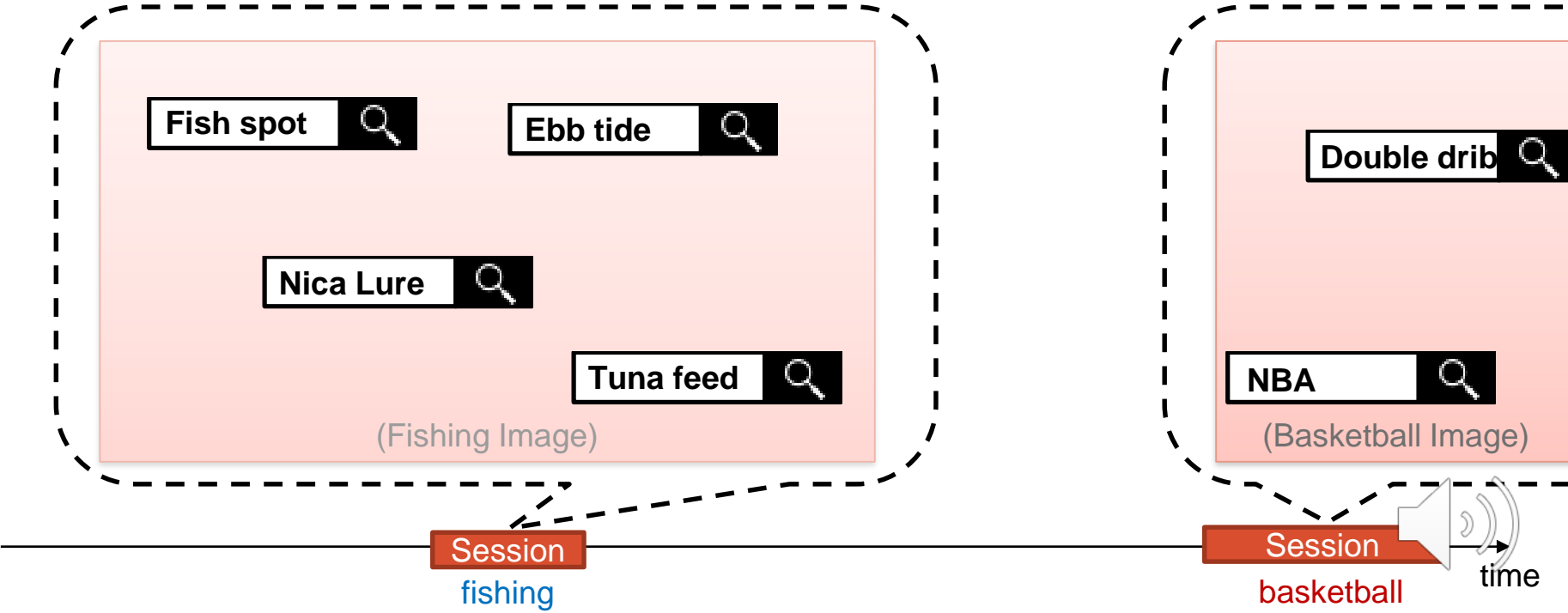
# Web Search Session Behavior



(Browser Image)



# Web Search Session Behavior



# Web Search Session Behavior

Session	Original Query	English Translation
1	代々木公園	Yoyogi Park
1	代々木公園 ランチ	Yoyogi Park, lunch
1	代々木公園 ランチ 子連れ	Yoyogi Park, lunch, with children
2	城山公園	Shiroyama Park
2	城山公園 子供	Shiroyama Park, children
2	メルヘン館 鹿児島	Fairy Tale Museum, Kagoshima (prefecture)
3	城南島海浜公園	Jonanjima Kaihin Park
3	城南島海浜公園 釣り	Jonanjima Kaihin Park, fishing
4	くじら運動公園	Kujira Undou Park
4	くじら運動公園 川遊び	Kujira Undou Park, swimming in river
4	くじら運動公園 バーベキュー	Kujira Undou Park, barbecue
5	光が丘公園	Hikarigaoka Park
5	東京 紅葉スポット	Tokyo, autumn leaves spots
6	豊洲ぐるり公園	Toyosu Gururi Park
6	豊洲ぐるり公園 夜景	Toyosu Gururi Park, night view



# POI Atmosphere Categorization

- POI categorization that can distinguish "atmospheres"
- To capture the POI atmosphere is to leverage embedded vector from user behavior on a web search engine



# AGENDA

---

01

## Background

About POI atmosphere categorization

02

## Proposed Method

LSTM with DSSM training

03

## Experiment

POI categorization,  
POI atmosphere categorization

04

## Conclusion

Take home messages





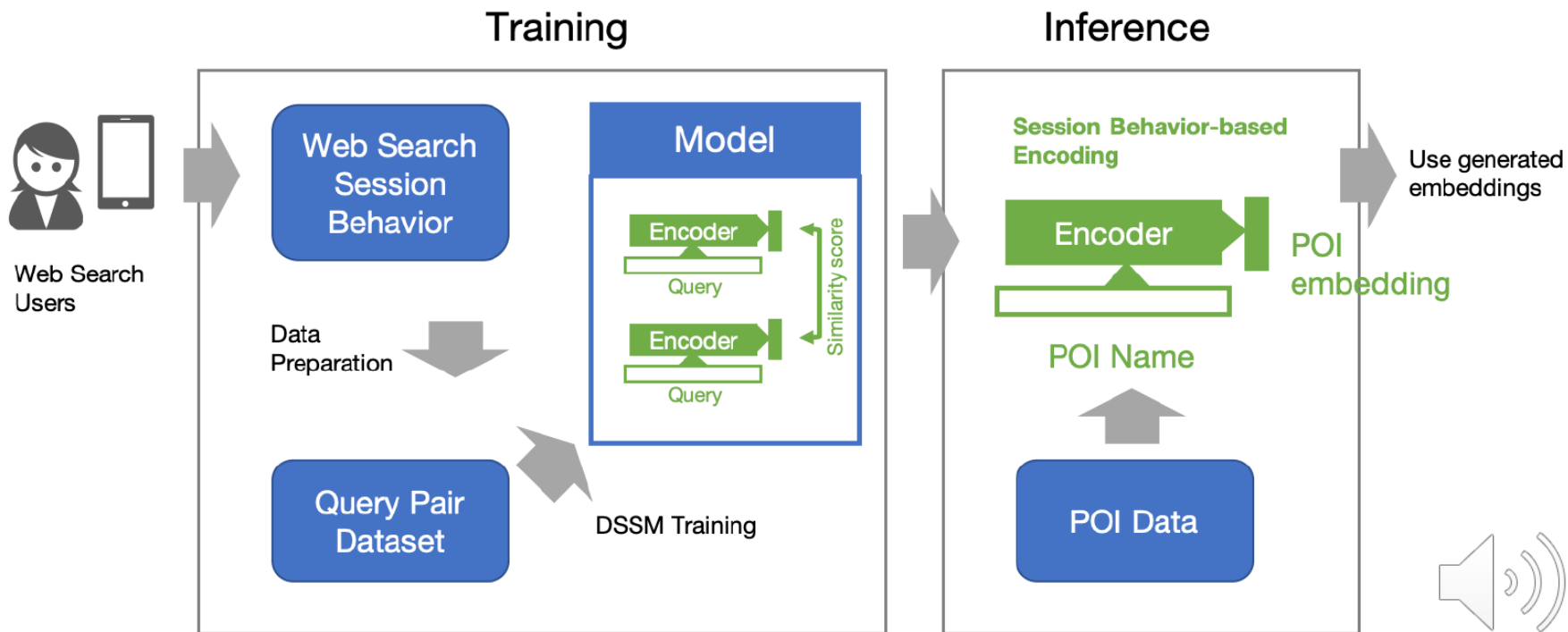
# Proposed Method

LSTM with DSSM training

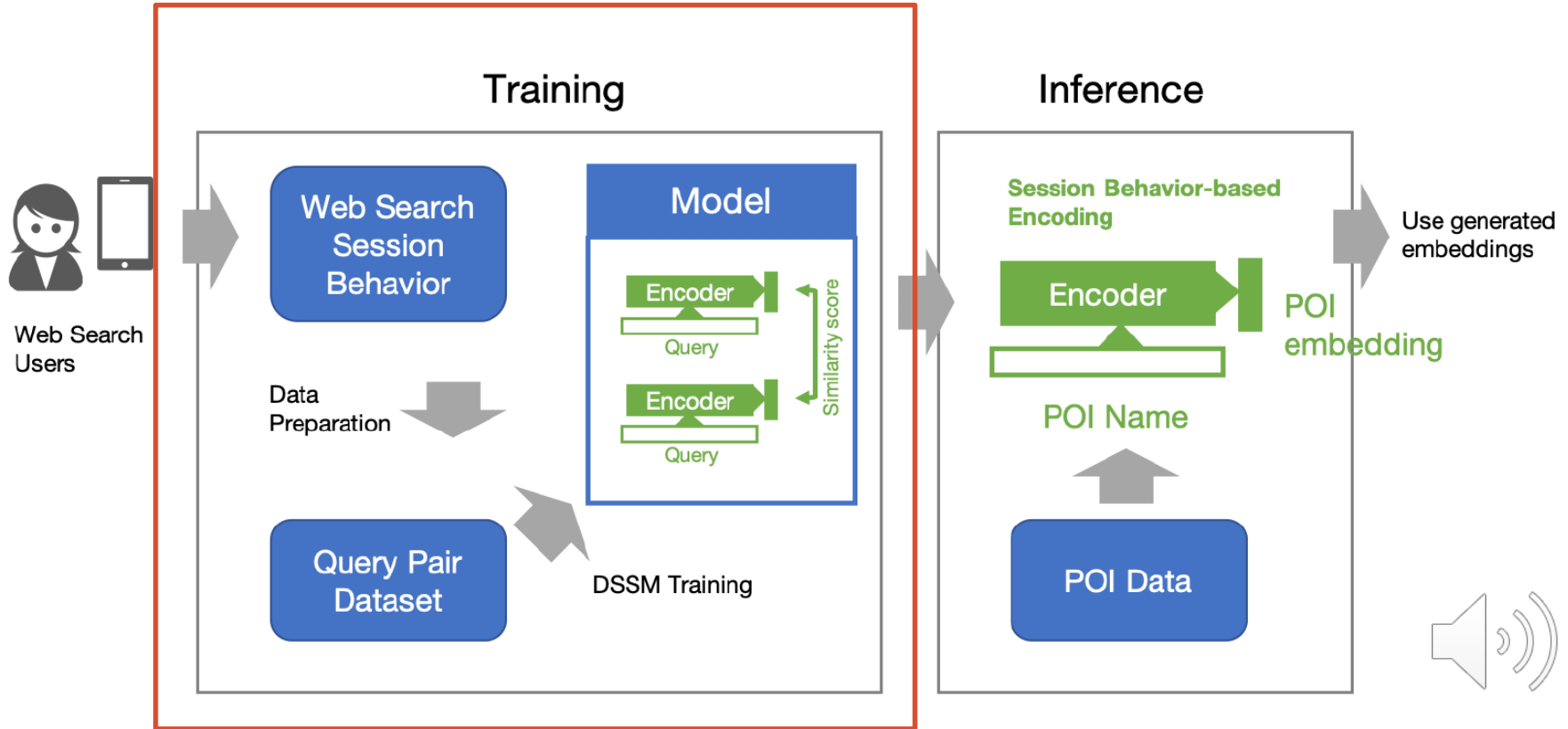




# POI atmosphere categorization system



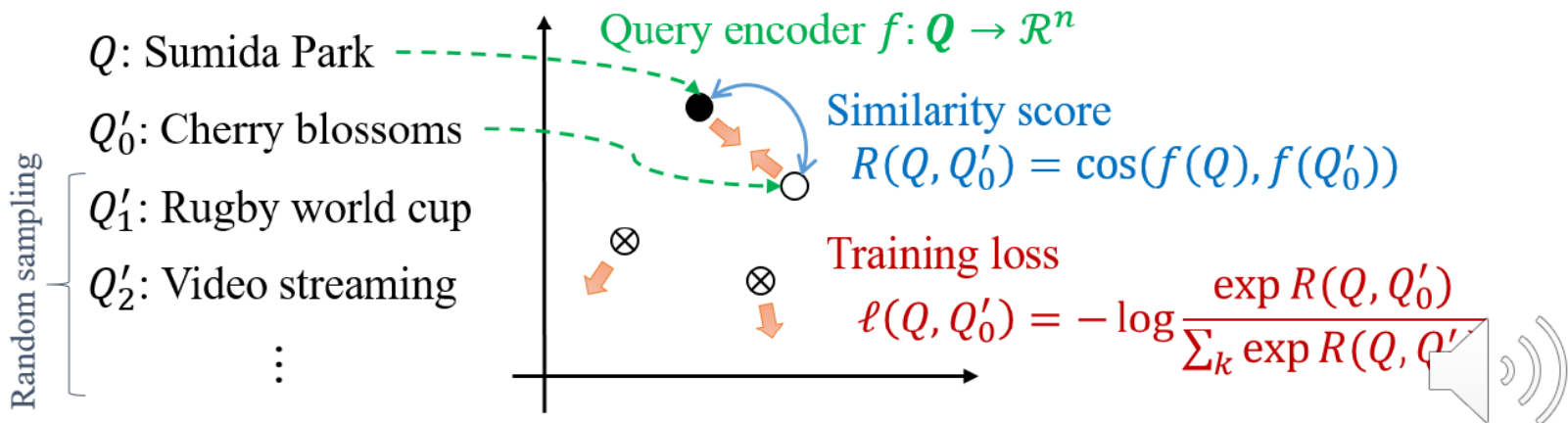
# POI atmosphere categorization system



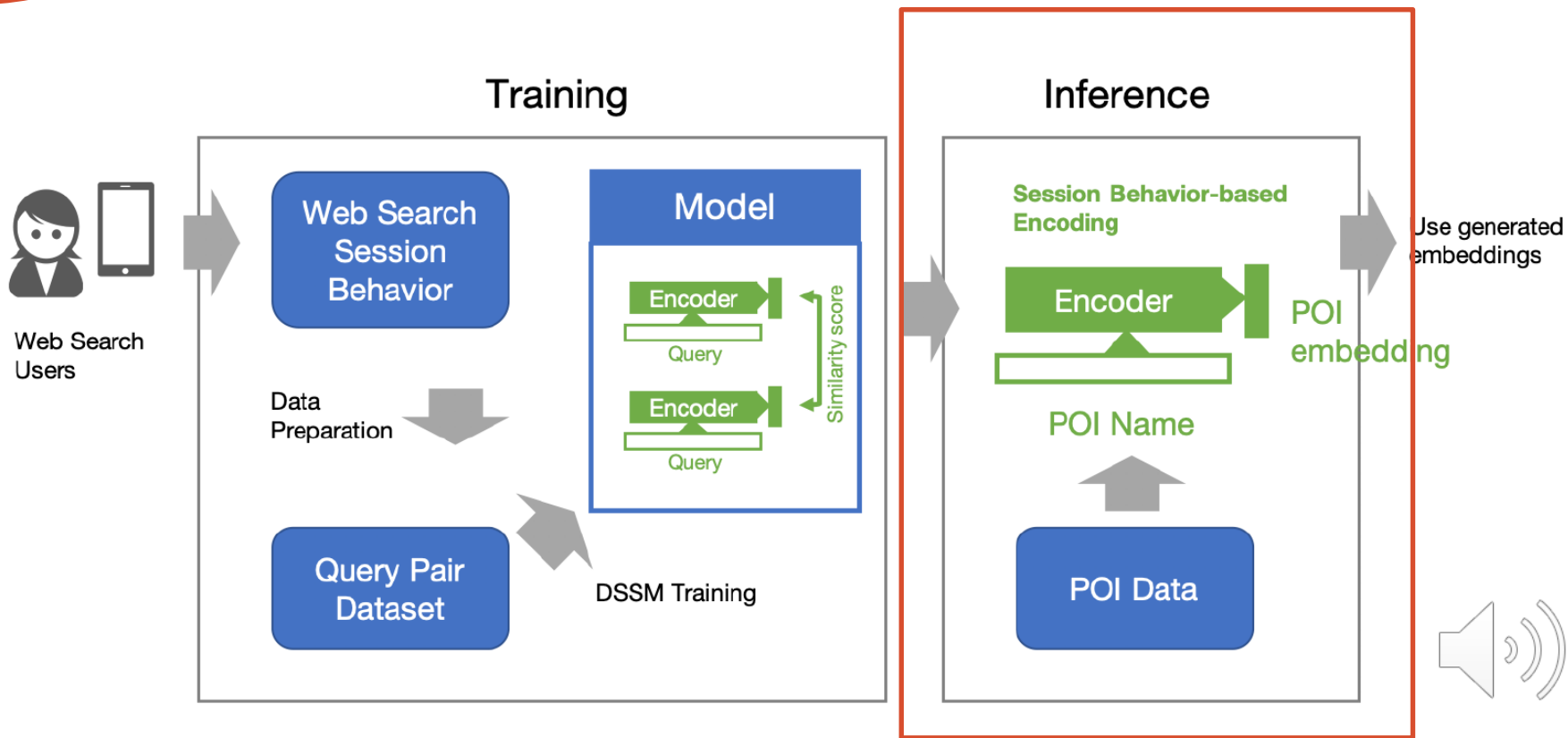
# DSSM training for next query prediction

- Next-query prediction with a deep structured semantic model (DSSM) can be useful for representing “nuance” of each queries.

Search session: (... , “Sumida Park”, “Cherry blossoms”, ... )



# POI atmosphere categorization system



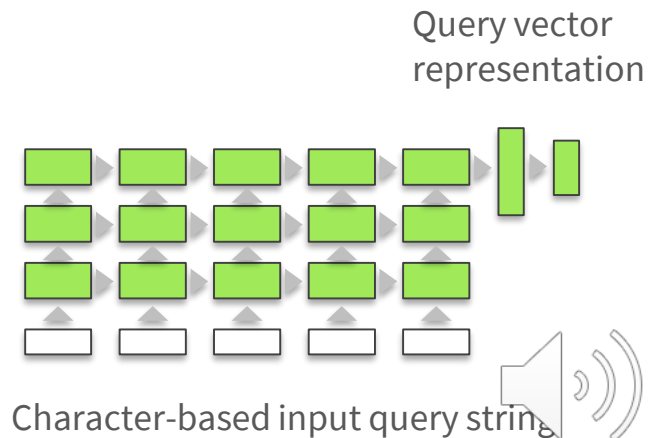
# Data / training conditions

## Data

- 480M single queries
- The average length of queries in the dataset was 9.14 Japanese characters
- The average number of terms in each query was 1.74
- The number of unique characters was 26,383

## Training Condition

- Maximum iteration : 5M
- 3-layer 1024-dimensional LSTM
- 256-dimensional vector representation
- 50 days training with one GPU (Tesla V100)



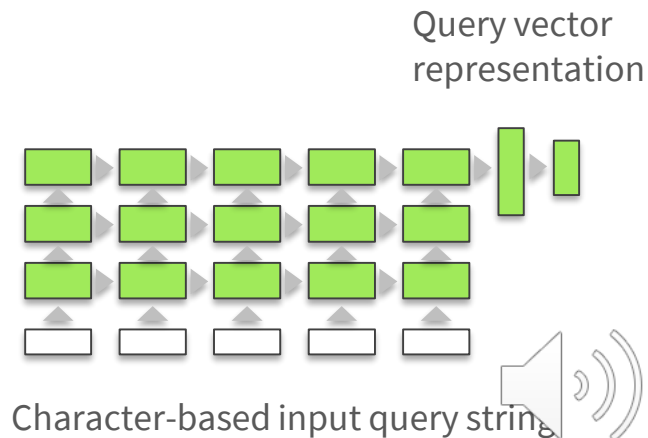
# Data / training conditions

## Data

- 480M single queries
- The average length of queries in the dataset was 9.14 Japanese characters
- The average number of terms in each query was 1.74
- The number of unique characters was 26,383

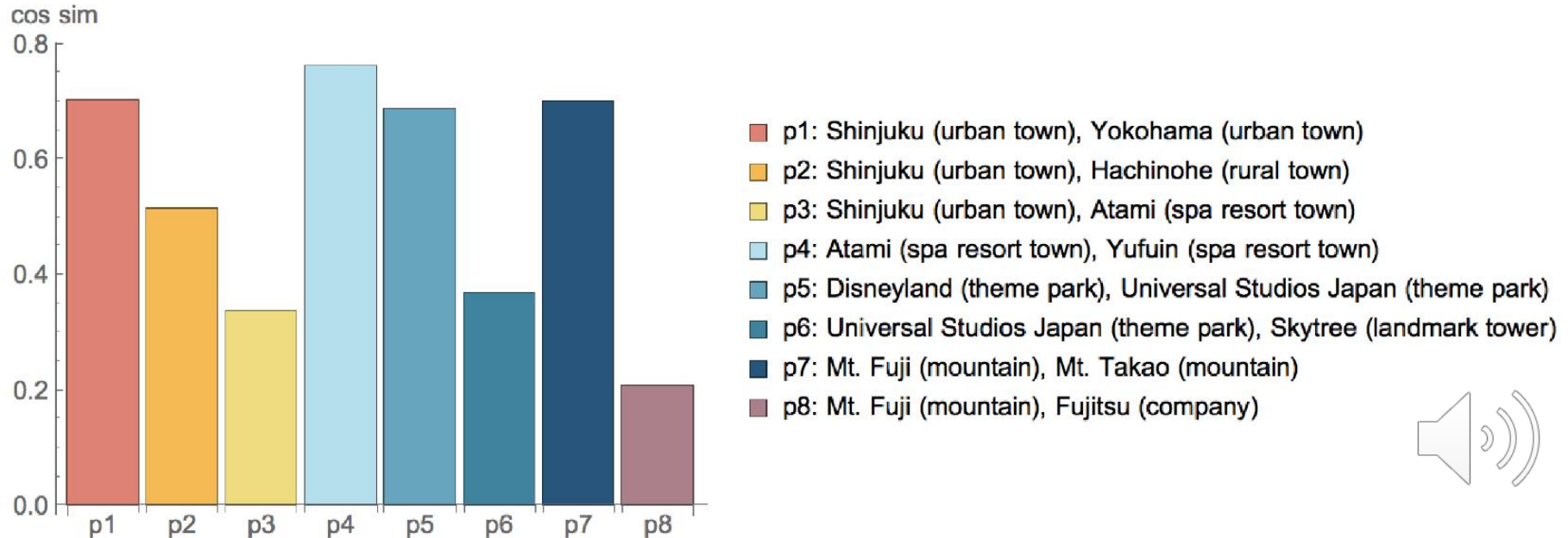
## Training Condition

- Maximum iteration : 5M
- 3-layer 1024-dimensional LSTM
- 256-dimensional vector representation
- 50 days training with one GPU (Tesla V100)



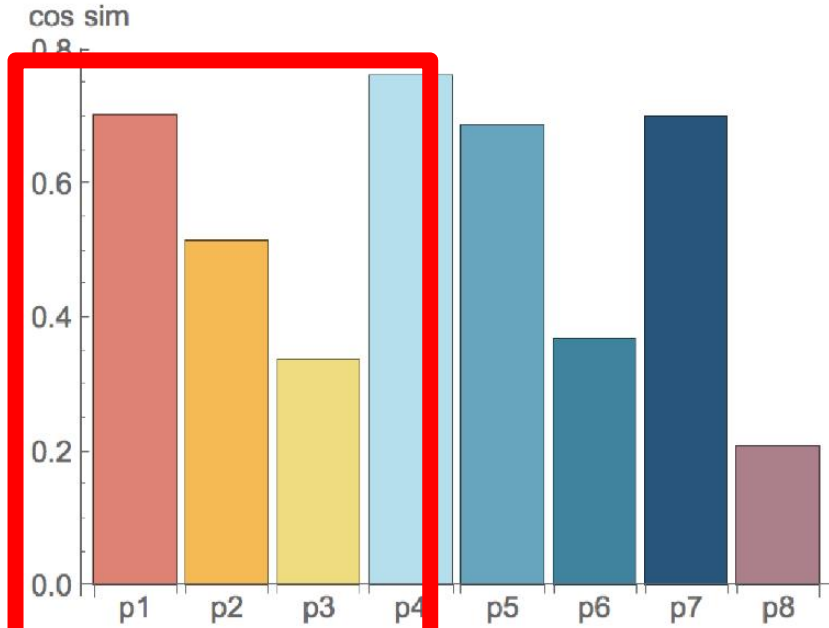
# Preliminary evaluation

- Query pairs which have the same nuance got high cosine similarity scores.



# Preliminary evaluation

- Query pairs which have the same nuance got high cosine similarity scores.



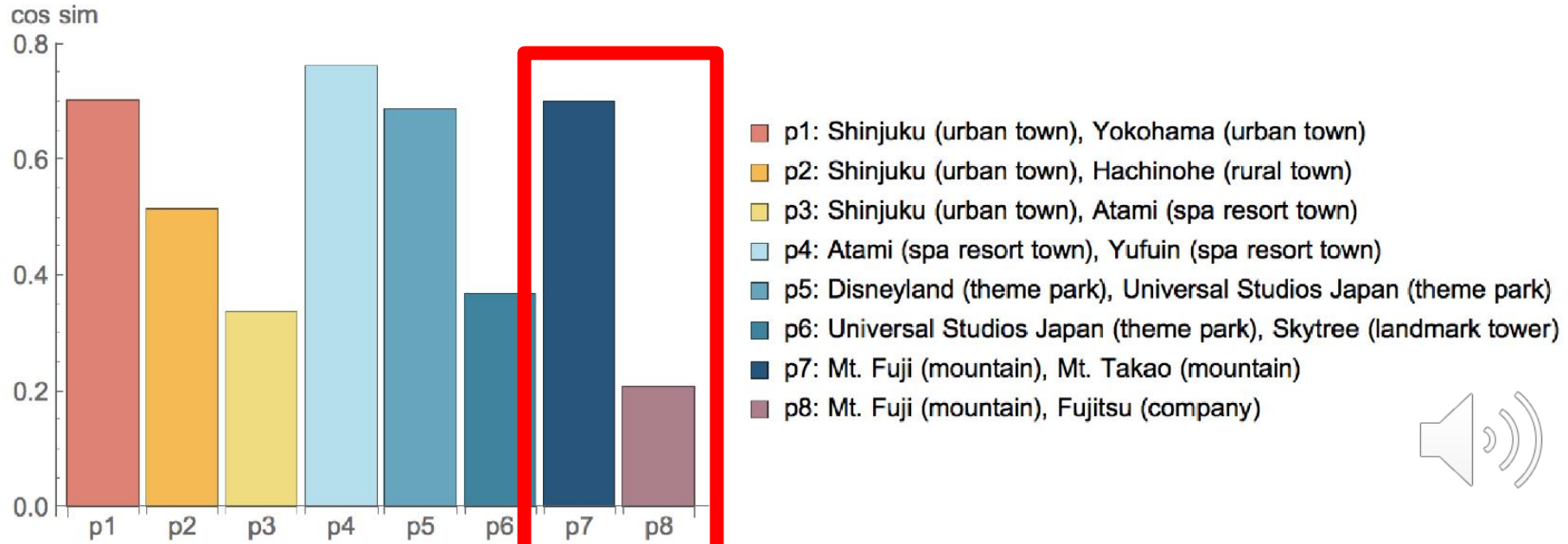
- p1: Shinjuku (urban town), Yokohama (urban town)
- p2: Shinjuku (urban town), Hachinohe (rural town)
- p3: Shinjuku (urban town), Atami (spa resort town)
- p4: Atami (spa resort town), Yufuin (spa resort town)
- p5: Disneyland (theme park), Universal Studios Japan (theme park)
- p6: Universal Studios Japan (theme park), Skytree (landmark tower)
- p7: Mt. Fuji (mountain), Mt. Takao (mountain)
- p8: Mt. Fuji (mountain), Fujitsu (company)





# Preliminary evaluation

- Query pairs which have the same nuance got high cosine similarity scores.



# AGENDA

---

01

## Background

About POI atmosphere categorization

02

## Proposed Method

LSTM with DSSM training

03

## Experiment

POI categorization,  
POI atmosphere categorization

04

## Conclusion

Take home messages





# Experiment

Exp.1: POI categorization for basic performance evaluation



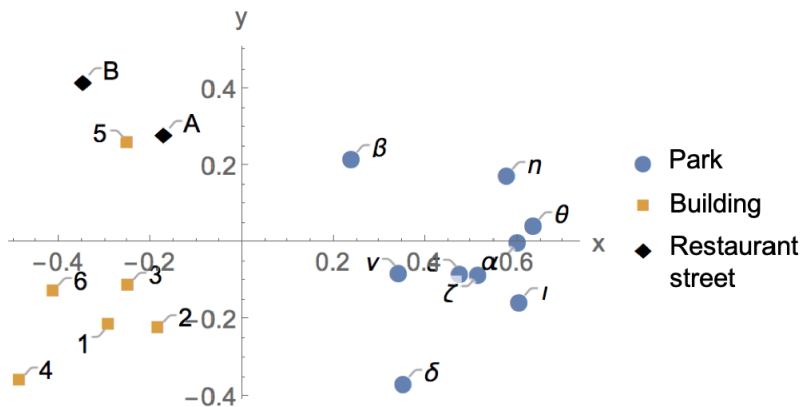
## Task1: POI classification

- Classification for POI of mixed categories located in Tokyo metropolitan area.
  - 9 parks
  - 6 famous buildings
  - 2 restaurant streets

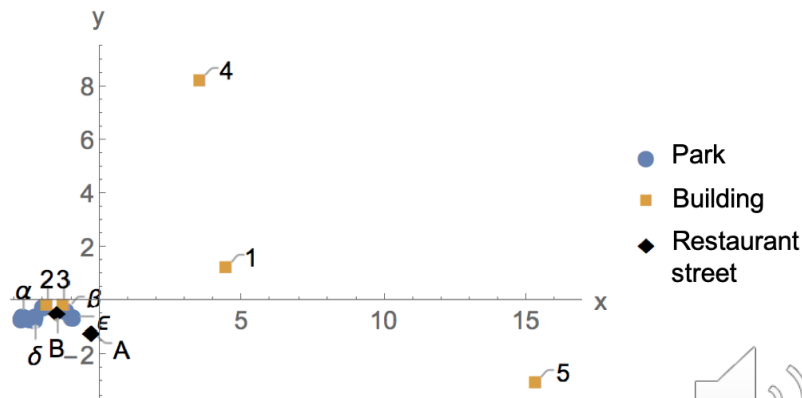


# Two-dimensional PCA-based visualizations

- Proposed method can clearly separate POI well.



(a) Proposed method.

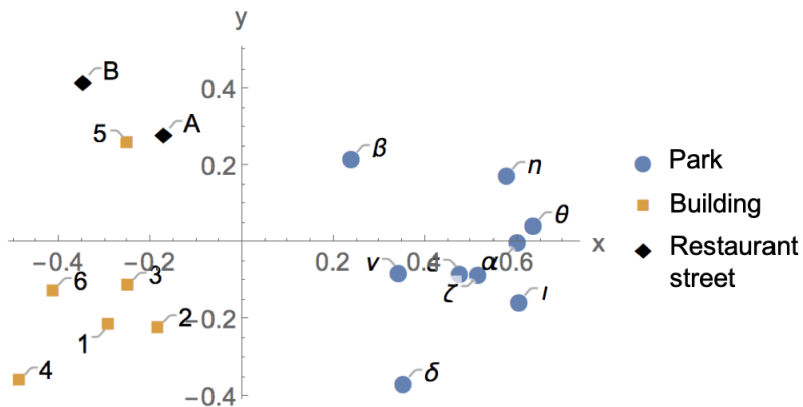


(b) Skip-gram.

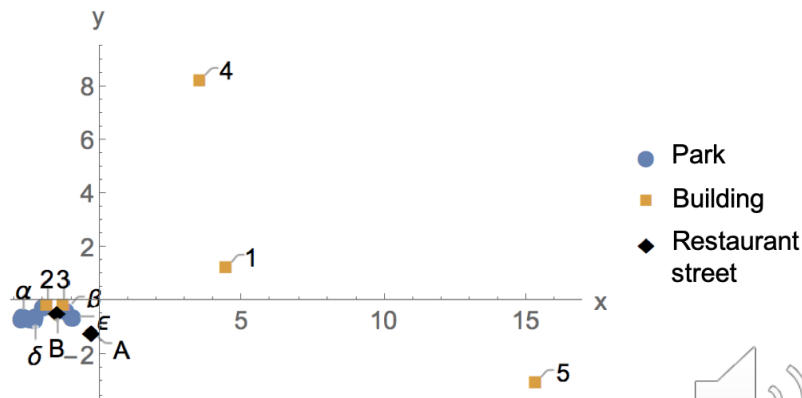


# Two-dimensional PCA-based visualizations

- Proposed method can clearly separate POI well.



(a) Proposed method.

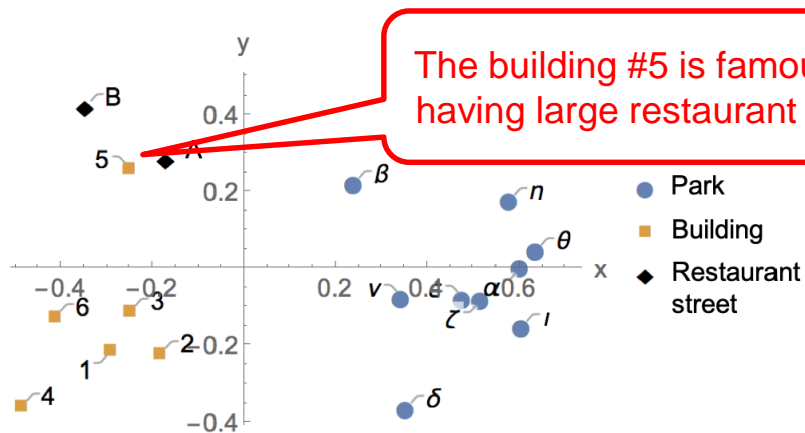


(b) Skip-gram.

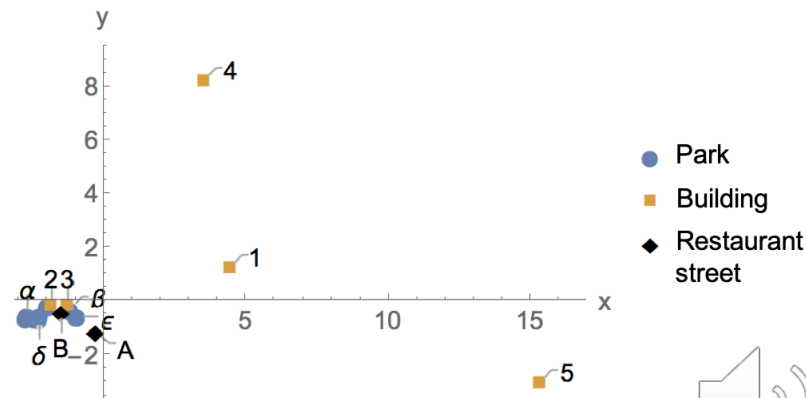


# Two-dimensional PCA-based visualizations

- For embeddings of parks, buildings, and restaurant streets
- Proposed method can clearly separate POI well.



(a) Proposed method.



(b) Skip-gram.



## Task2: park classification

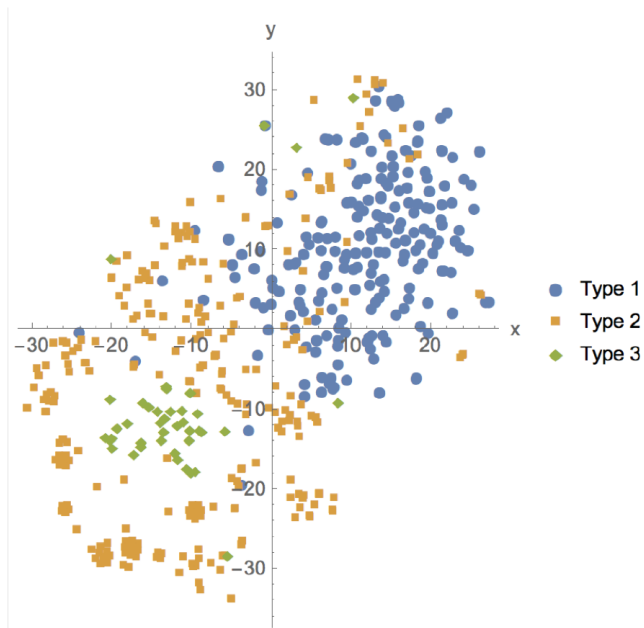
- Classification for 532 parks in Tokyo metropolitan area
- Japanese gov. established classification scheme of parks:

- Type 1: Parks that are usually in the suburbs and utilized for sports activities and recreation.
- Type 2: Green belts in urban spaces for improving the cityscape.
- Type 3: Greenways and forests for improving the safety and comfort of city life.

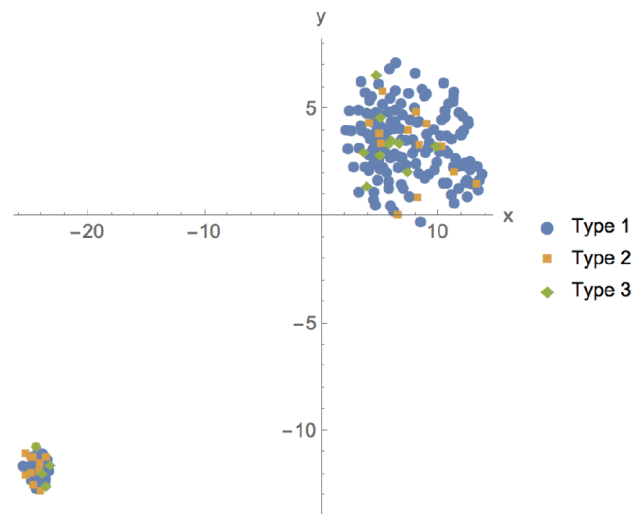




# Qualitative Evaluation



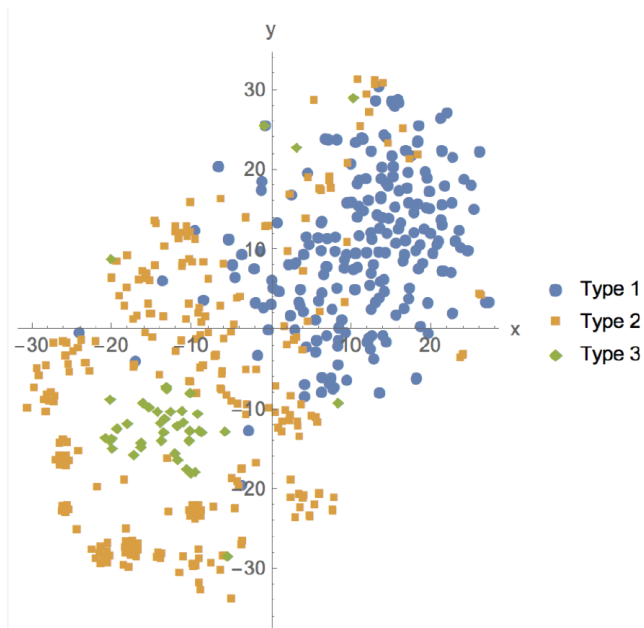
(a) Proposed method.



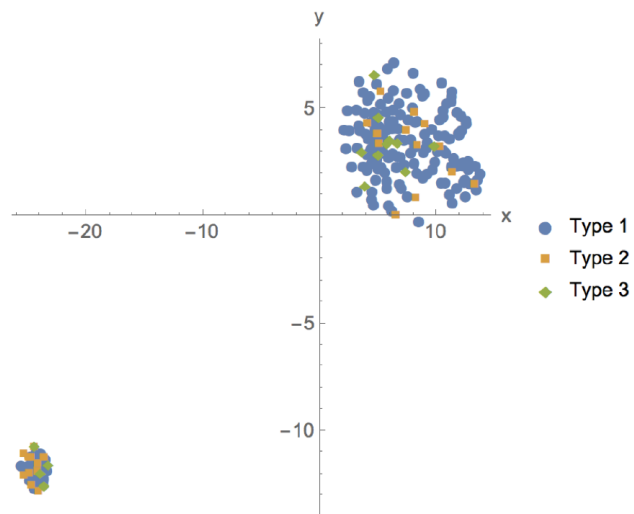
(b) Skip-gram.



# Qualitative Evaluation



(a) Proposed method.



(b) Skip-gram.



# Quantitative Evaluation

## Separability Score

- the average accuracy of 10 runs of a simple classifier

## Results

- Proposed method clearly perform well compared with skip-gram classification

	One-layer	Two-layer
Proposed method	$0.929 \pm 0.002$	$0.962 \pm 0.003$
Skip-gram	$0.822 \pm 0.000$	$0.861 \pm 0.008$





# Experiment

Exp.2: Connecting Park Atmosphere to the Real World



## Selected parks for detail analysis

Id	Park Name	Type	Area	$x$	$y$
$\alpha$	Komaba	Nei.	40,396	0.089	0.271
$\beta$	Jingudoori	Nei.	3,128	-0.473	0.057
$\gamma$	Minami Ikebukuro	Urb.	7,818	-0.047	-0.109
$\delta$	Futako Tamagawa	Dis.	63,000	0.420	-0.191
$\epsilon$	Denen Chofu Seseragi	Spo.	30,300	0.218	0.102
$\zeta$	Setagaya	Dis.	78,957	0.158	0.002
$\eta$	Senzoku Ike	Gen.	40,000	0.076	0.386
$\theta$	Utsukushi Ga Oka	Nei.	21,832	0.263	0.351
$\iota$	Aobadai	Nei.	38,000	0.173	0.240





$\alpha$ : Komaba



$\beta$ : Jinguodori



$\gamma$ : Minami Ikebukuro



$\delta$ : Futako Tamagawa



$\epsilon$ : Denen Chofu Seseragi



$\zeta$ : Setagaya



$\eta$ : Senzoku Ike



$\theta$ : Utsukushi Ga Oka

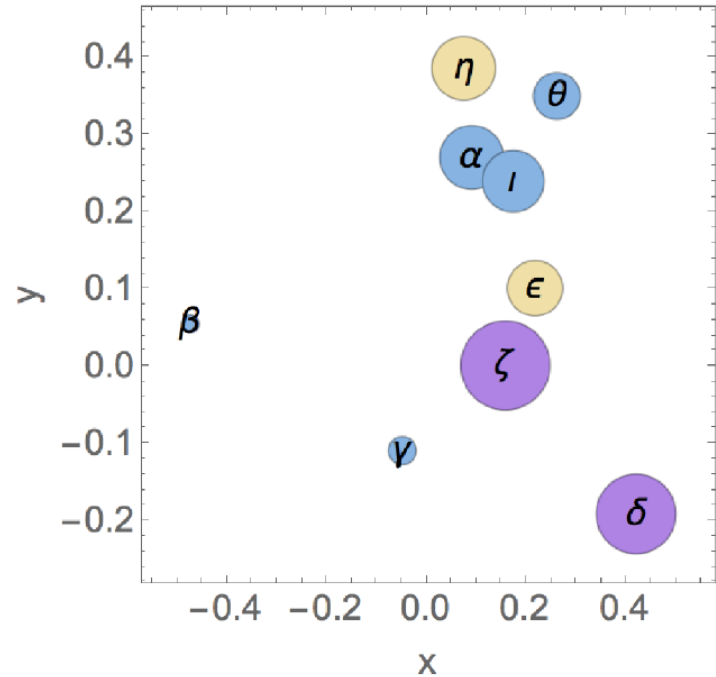


$\iota$ : Aobadai

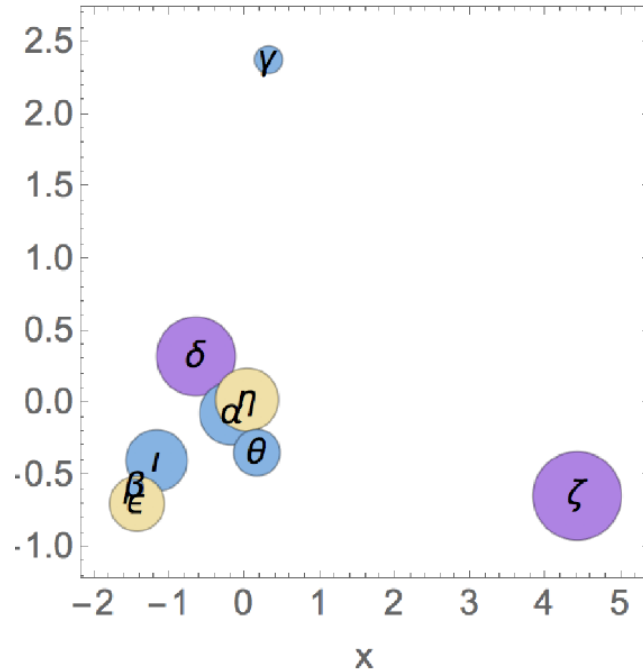


# Two-dimensional PCA visualizations for embeddings of nine parks

Proposed method



Skip gram

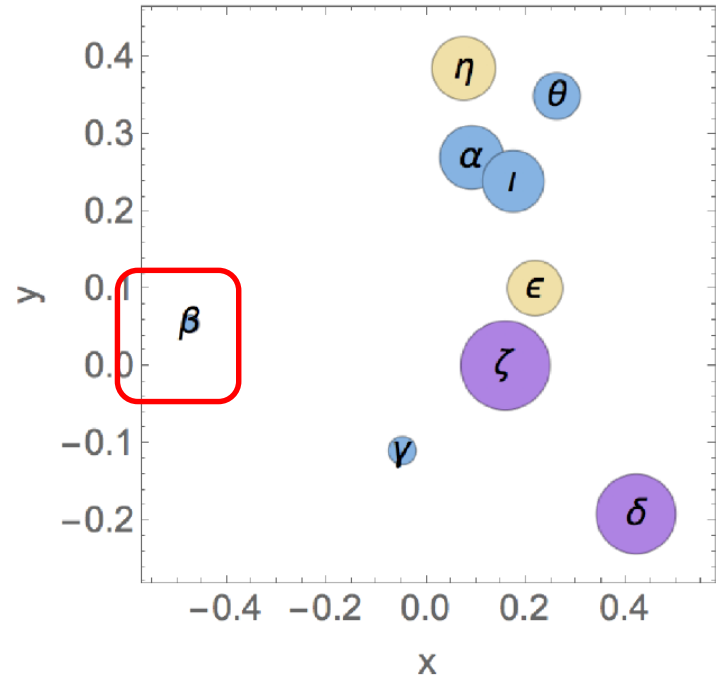


- District
- Sports and general
- Neighboring and urban

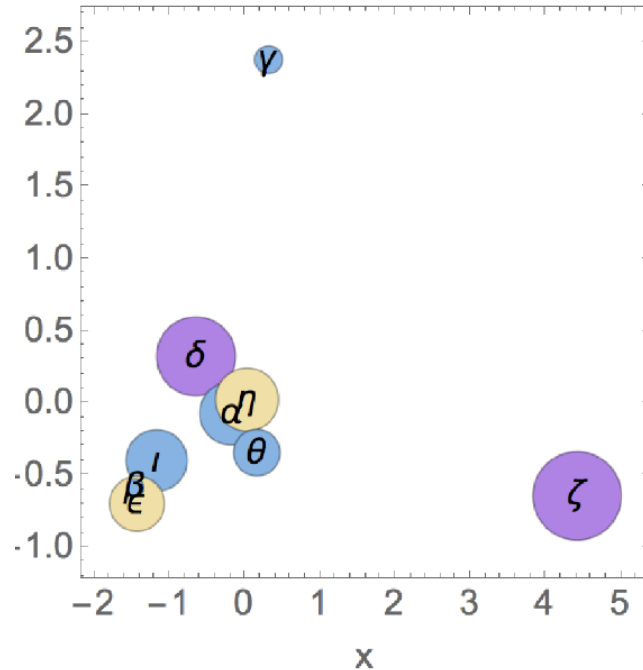


# Two-dimensional PCA visualizations for embeddings of nine parks

Proposed method



Skip gram



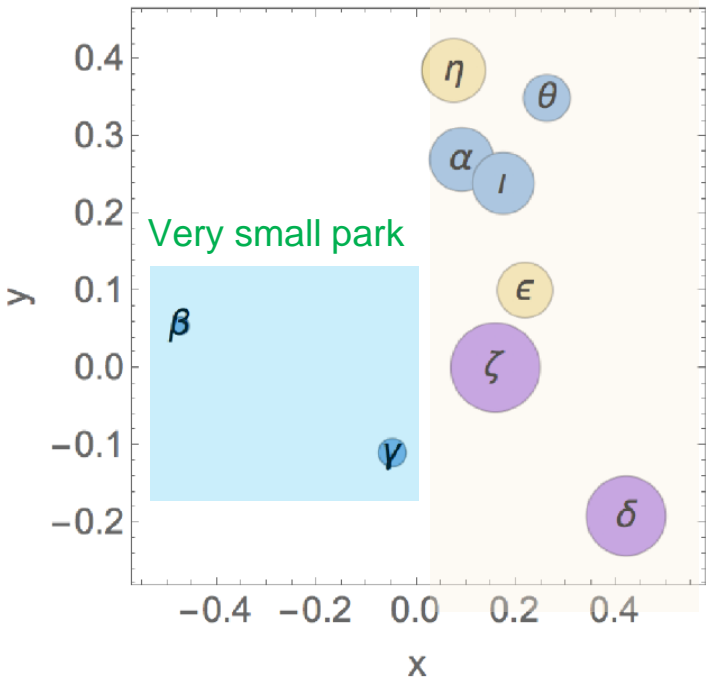
- District
- Sports and general
- Neighboring and urban



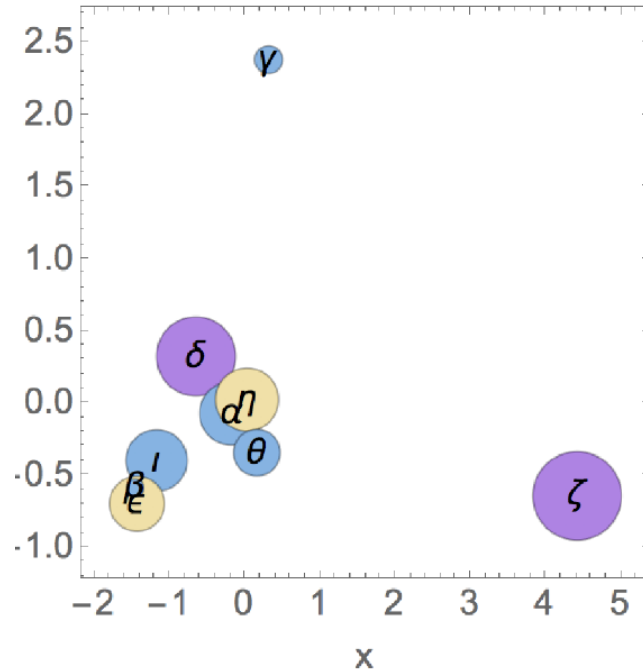


# Two-dimensional PCA visualizations for embeddings of nine parks

Proposed method



Skip gram

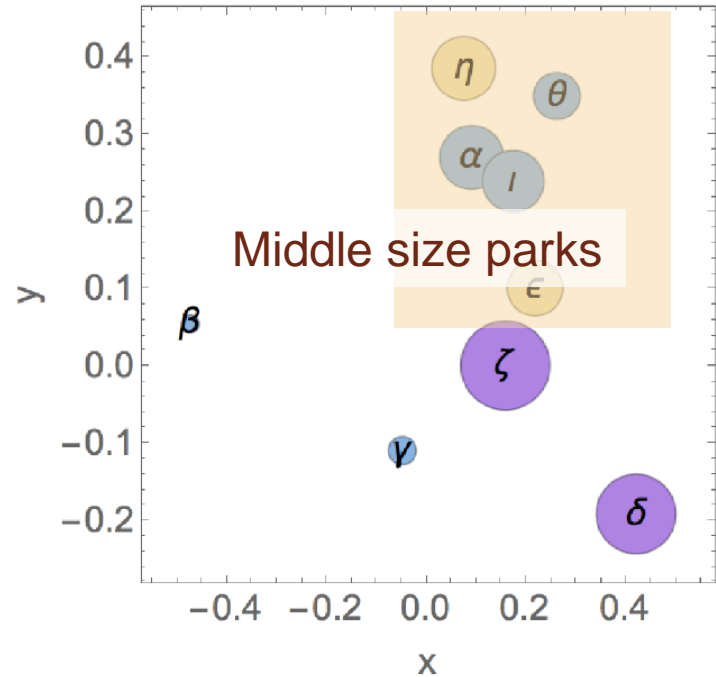


- District
- Sports and general
- Neighboring and urban

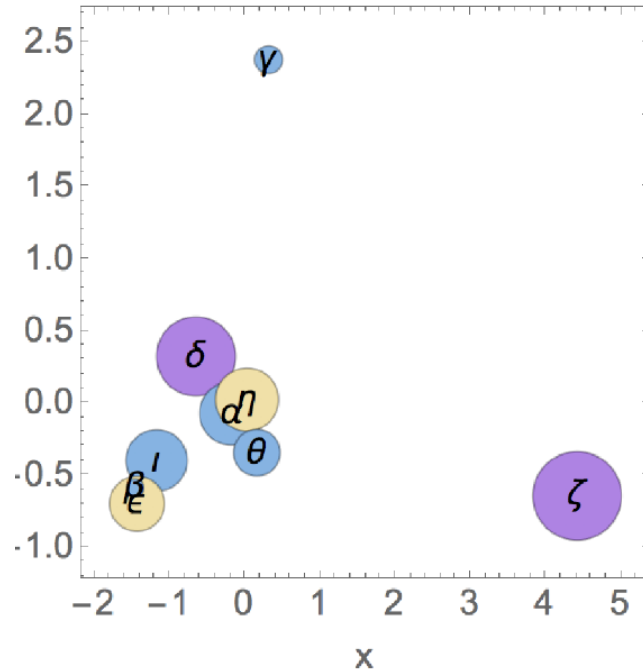


# Two-dimensional PCA visualizations for embeddings of nine parks

Proposed method

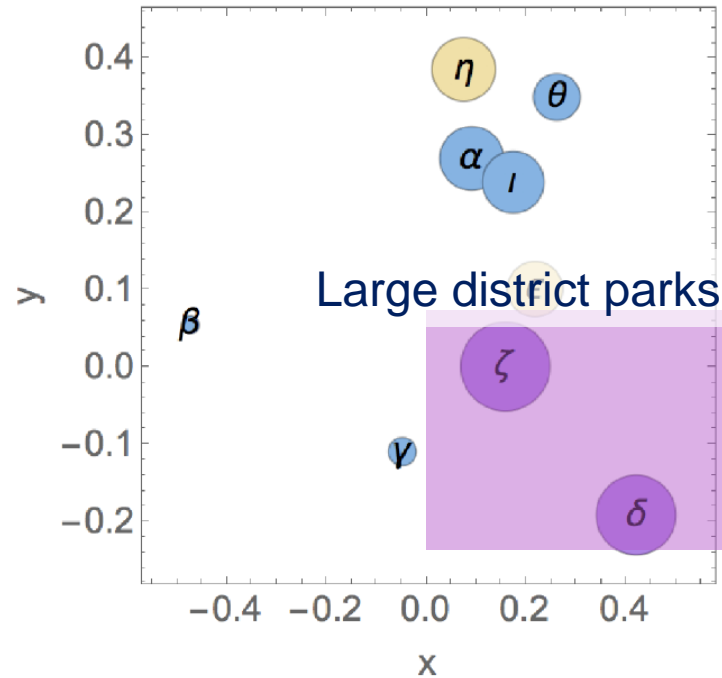


Skip gram

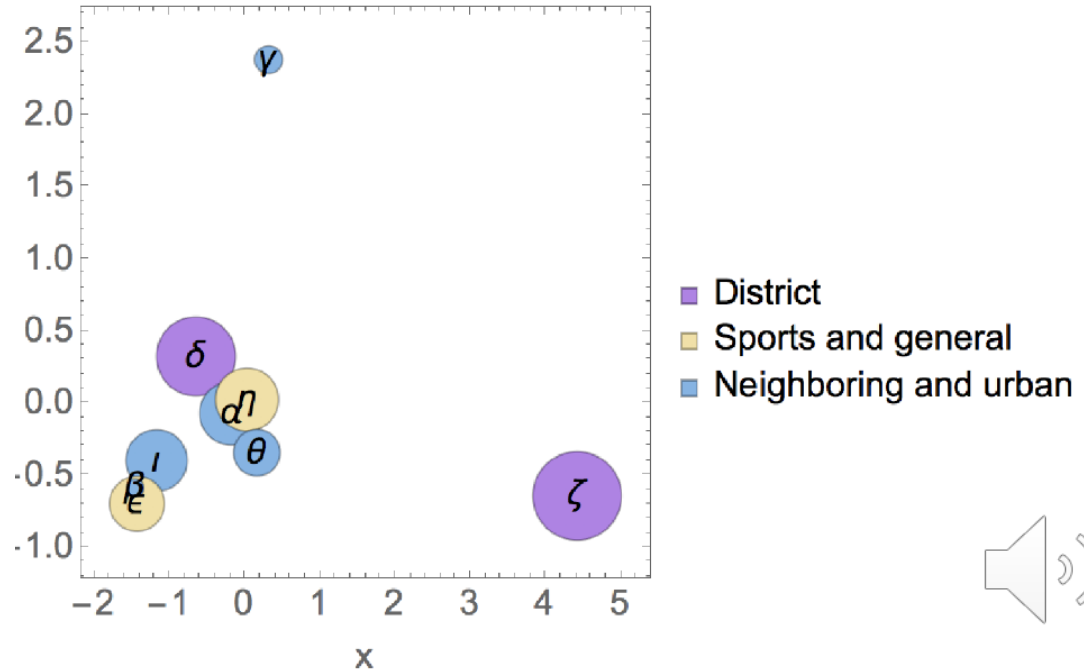


# Two-dimensional PCA visualizations for embeddings of nine parks

Proposed method

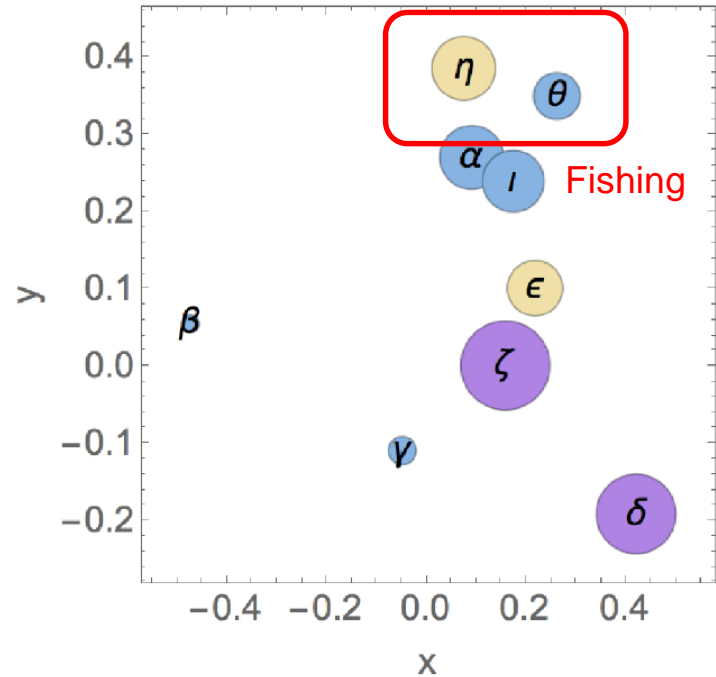


Skip gram

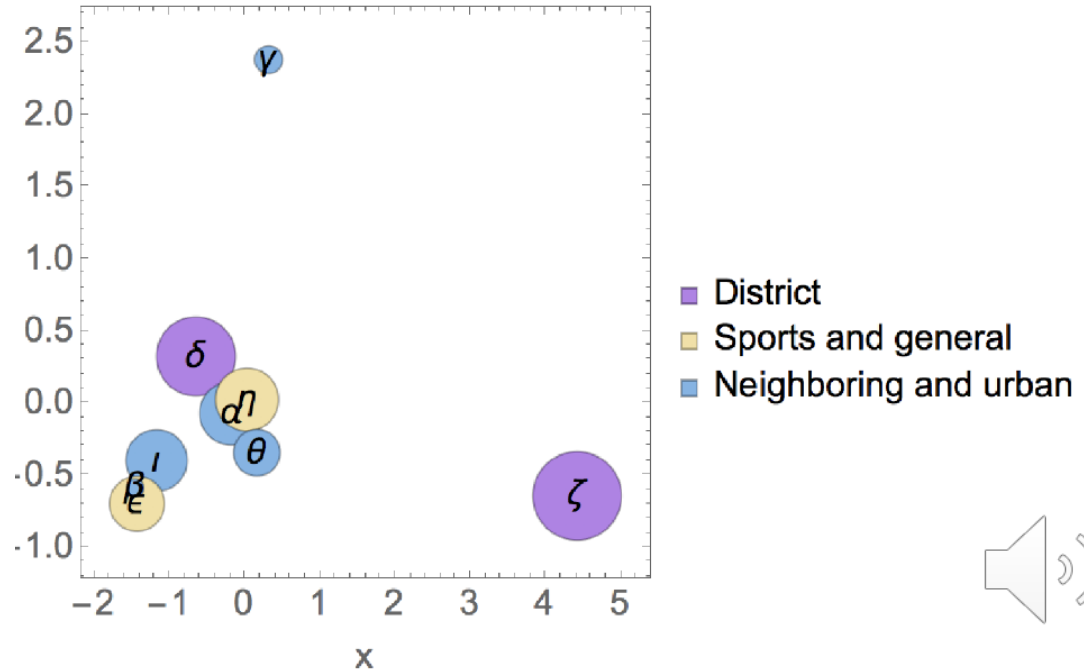


# Two-dimensional PCA visualizations for embeddings of nine parks

Proposed method



Skip gram



# AGENDA

---

01

## Background

About POI atmosphere categorization

02

## Proposed Method

LSTM with DSSM training

03

## Experiment

POI categorization,  
POI atmosphere categorization

04

## Conclusion

Take home messages



# Conclusions



- We proposed a new POI categorization method that can represent the atmosphere of a POI beyond its geospatial features.
- We focus on users' search behavior with LSTM based encoding using DSSM pre-training.
- We demonstrated the effectiveness of our method, especially for POI atmosphere in comparison with the widely used skip-gram models.
- We believe that our method complements the existing methods based on geographic features, and will encourage their use in the real world.



# Conclusions



- We proposed a new POI categorization method that can represent the atmosphere of a POI beyond its geospatial features.
- We focus on users' search behavior with LSTM based encoding using DSSM pre-training.
- We demonstrated the effectiveness of our method, especially for POI atmosphere in comparison with the widely used skip-gram models.
- We believe that our method complements the existing methods based on geographic features, and will encourage their use in the real world.



# THANKS

Do you have any questions?

[ktsubouc@yahoo-corp.jp](mailto:ktsubouc@yahoo-corp.jp)



Kota Tsubouchi,  
Yahoo Japan Corporation



Hayato Kobayashi,  
Yahoo Japan Corporation



Toru Shimizu  
Yahoo Japan Corporation

