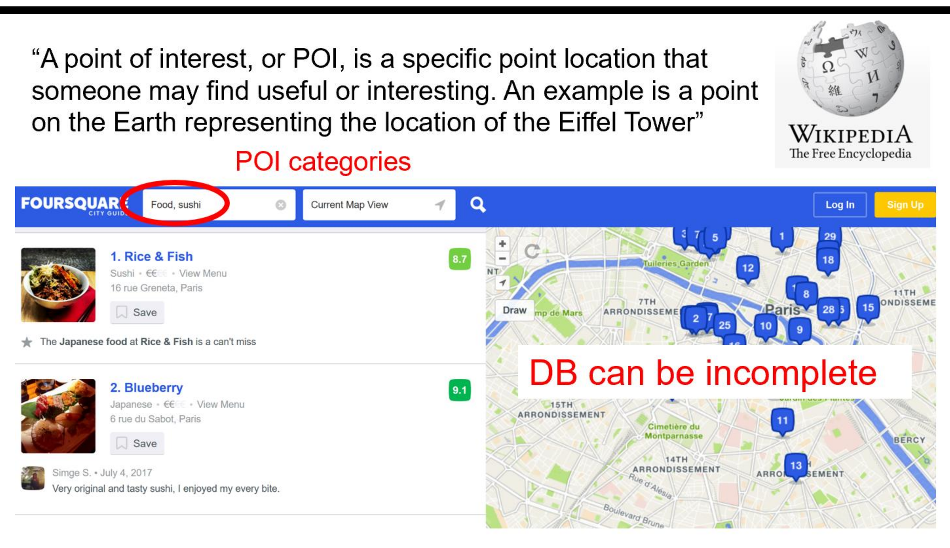
Culture-aware Point-of-Interest Completion

in a Global Location-Based Social Network Database without Access to User Information

NAVER LABS EUROPE

Scope



Motivation

User data has been used to define context
- Users' check-in information i.e. frequency, duration, location and time [1, 2]
- Users' gender and age [3]
- User-defined tags [4,5]

Drawbacks
- Access to users' information can be difficult (see GDPR)
- Only location proximity has been considered as context
  - Cultural background is extremely important [6,7]

Our Contributions
- First study of culture-aware POI categorization in a global, multi-lingual database,
- without access to user data at training time

Example

"La Table du Ramen" located in Paris, France

French Culture (Observed category): Japanese Restaurant

Japanese Culture (Target categories): Ramen Restaurant, Noodle House

Key idea

➢ Insight : Majority of POIs are categorized in a manner that reflects local culture in Location-Based Social Networks

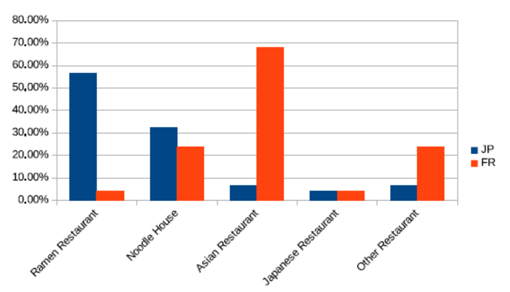


Figure 1: Category distribution of POIs having the token "noodle" in their name in Japan and France. It is obvious that "Ramen Restaurant" is the most popular category in Japan and "Asian Restaurant" in France.

> Use culture related attributes to learn a latent representation of POI's categories at training time. Then, at inference time, replace the corresponding inputs according to the target cultural profile.

Overview of the method

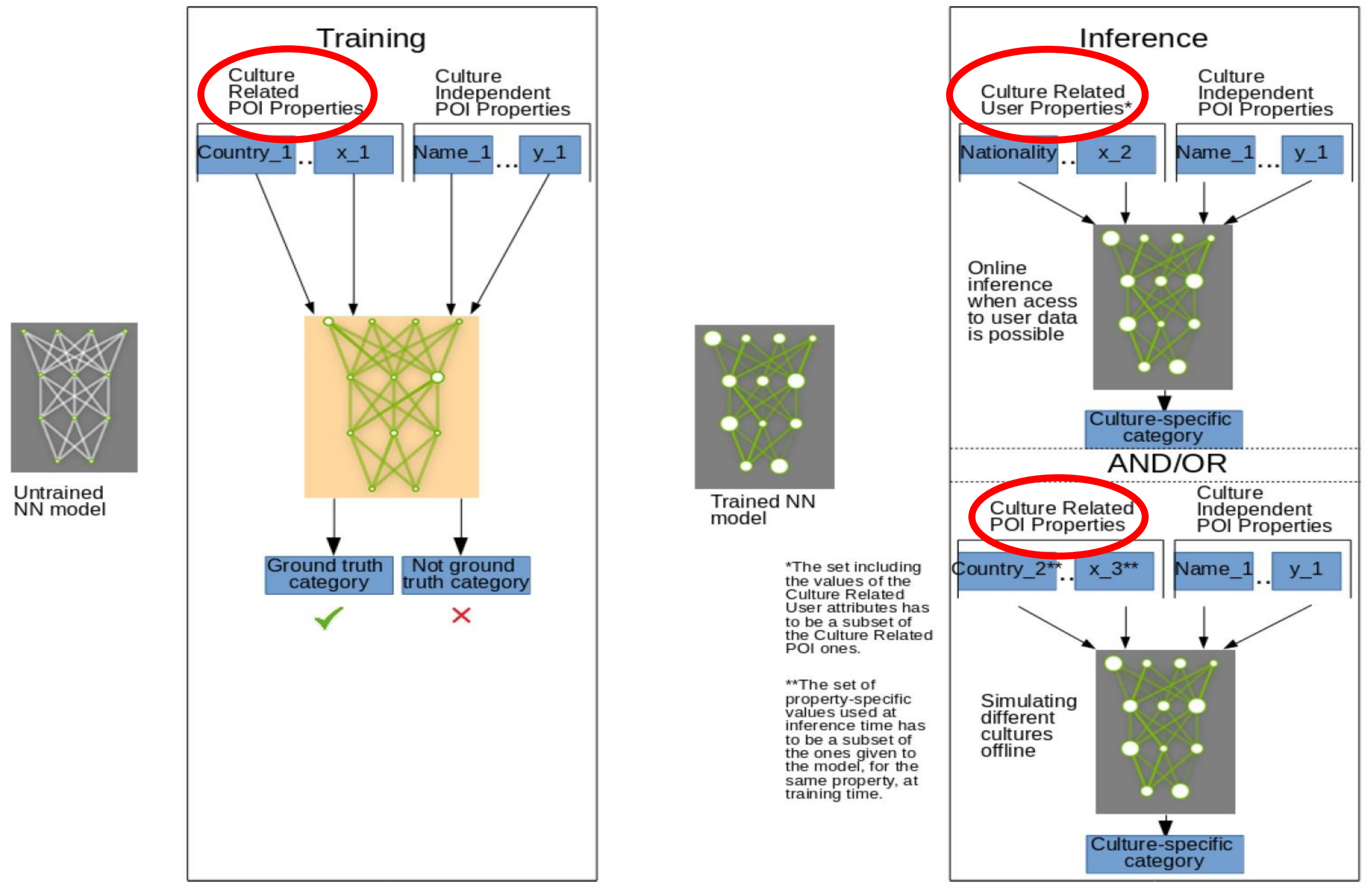


Figure 2: Overview of proposed method. Images of NN models are adapted from [3]

Problem Definition

Observed POI: $p = \{\boldsymbol{x}, \boldsymbol{y_o}\} = \{\boldsymbol{x_C}, \boldsymbol{x_N}, \boldsymbol{y_o}\}$

Target POI: $p = \{\boldsymbol{x}, \boldsymbol{y_c}\}$

Target classifier: $\boldsymbol{y_c} = b_c(\boldsymbol{x})$

Implementation Steps

1. Attribute selection: POI name, location (lat, long)
2. Vectorisation:
- Categorical variables: One hot encoded embeddings
- Sequential variables: 3-gram char LSTM embeddings
- Spatial attributes: lat, long -> discretized to -> Countries (then as categorical) but other granularity is possible (e.g. regions)
- Other cultural variables (optional): e.g. opening times, prices ranges – follow the same approach as for spatial attributes

3. Training:
- Concatenation of all vectorized attributes $\tilde{x} = [\phi_1(a_1), \phi_2(a_2), \ldots, \phi_n(a_n)]$
- Dense layer $\boldsymbol{h} = relu[W^h \tilde{x} + \boldsymbol{b}^h]$
- Output layer $p(\boldsymbol{y}|\boldsymbol{h}, \theta) = sigmoid[W\boldsymbol{h} + \boldsymbol{b}]$

4. Inference: Culture specific-inputs. Apply a threshold of 0.5 for label selection.

Experiments

2.4 M POIs by Foursquare (Train: 40%, Dev: 10%, Test: 50%)

808 categories

Attributes

POI name

Latitude, Longitude

Category	POIs (%)
Café	9.83
Restaurant	5.88
Pizza Place	4.49
Coffee Shop	4.48
Bakery	3.8
Fast Food Restaurant	3.23
...	...
Chinese Restaurant	2.98
Japanese Restaurant	2.5
Asian Restaurant	2.41
...	...
Noodle House	1.02
...	...
Ramen Restaurant	0.65
...	...

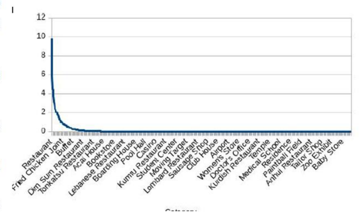


Table 3: Culture-specific prediction results for different POI categories. Green coloured values are significantly higher than in the the rest of the cultures, and red significantly lower, indicating a notable culture-specific influence.

Category	Original data	Culture					
		KR	FR	US	BR	TR	GR
Acai House	1488	43	29	0	1534	0	0
Note:Except for Brazil in the rest of the cultures the same POIs are categorised as Snack Place, Juice Bar, Dessert Shop. According to Foursquare's documentation Acai house is a category only supported in Brazil.							
Bistro	879	470	3310	351	1162	95	10
Note: Bistros predicted using the French culture are tagged in the original data as: Café, Bar, Gastropub, Diner. Corresponding predictions in other cultures are: Café, Wine Bar, Bar, Gastropub.							
Brasserie	18	0	91	0	3	0	0
Note: In other cultures Café is the main predicted category for the same POIs (or there is no prediction at all). In the silver standard the POIs are also categorised as Bistro or Café.							
Café	126665	108908	87381	63223	108837	148050	236436
Note: In the US culture a lot of Cafes seem to be categorised as Coffee Shops instead. In the Greek culture Coffee Shop, Breakfast Place, Dessert Shop, Bar, Tea Room POIs are categorised as Café (which is actually representative of the culture).							
Coffee Shop	54291	49069	45769	68102	50513	52742	23629
Note: As explained in the previous row.							
Churrascaria	557	0	0	0	674	0	0
Note: Churrascaria is a Portuguese/Brazilian BBQ. In other cultures the majority of the same POIs are classified as BBQ Joint and a small percentage as Steakhouse (especially in the US).							
Creperie	978	775	2968	932	1138	917	1740
Note: Creperies are obviously common place in the French culture. In the US and KR ones the same POIs are rather categorised as Dessert Shop or Breakfast Shop. In the BR one in addition to Dessert Shop we find also Pastelaria[9].							
Dessert Shop	15164	19422	4747	9651	15460	21941	13736
Note: In the French culture Dessert Shop POIs are rather classified as Café, Bakery, Creperie, Pastry Shop, Chocolate Shop. In the US one we have to note the large number of POIs categorised as Ice Cream Shop, Frozen Yogurt Shop, Candy Store.							
Diner	2590	2289	2388	3980	1539	1593	100
Note: Some of the POIs categorised in the US culture as Diner, are mainly categorised in other cultures as Café or Breakfast Spot or there is no prediction (the difference is really big with Greece where almost all of them are categorised as Café).							
Friterie	656	7	4864	2	419	4	62
Note: The model has learned that Friterie is a culture-specific category (supported in FR, BE, NL in the Foursquare database). It is interesting to note that in the US culture the same POIs are categorised as Burrito Place, Taco Place, Food Truck, in the TR one as Kofte Place and in the GR one as Snack Place. They all seem to share a fast food aspect.							
Meyhane	854	0	0	0	0	2157	0
Note: The model has learned that Meyhane is TR culture-specific category. In other cultures Meyane POIs usually do not get any prediction.							
Pastelaria	340	1	0	0	898	57	0
Note: The model has learned that Pastelaria is BR (and Portuguese) culture-specific category. In FR culture categorised as Creperie, Snack Place, Bakery, US:Bakery,Snack Place, KR:Bakery, Dessert Shop, Snack Place.							
Pastry Shop	60	5	307	0	0	88	0
Note: The model has learned that Pastry Shop is more frequent in FR (i.e. Patisserie). In other cultures we mainly find Dessert Shop or Bakery.							
Souvlaki Shop	94	0	0	0	0	0	2434
Note: The model learned that Souvlaki Shop is specific to GR culture. In other cultures and in the silver standard, the same POIs are categorised mainly as Fried Chicken Joint and BBQ joint. There is a strong correlation to Steakhouse as well, or more precisely to Kebab shops that are also classified as Steakhouse in the silver standard.							
Sports Bar	67	9	2	100	4	5	0
Note: The model learned that Sports Bar is more frequent in the US culture. In the silver standard, the same POIs are categorised also just as Bar and/or Wing Joint. Furthermore, the model learned a strong correlation between the category Wings Joint and sports Bar - the two categories are predicted together 77% of the time.							
Takoyaki Place	186	130	45	34	22	44	30
Note: Takoyaki Place POIs do not get any prediction most of the times in other cultures except for KR. 86% of the predicted POIs are correct according to the silver standard.							

Conclusions

Model learned
- Categories that are only allowed in specific countries by design e.g. Churrascaria in Brazil, Portugal
- To make predictions that reflect local culture reasonably well (as far as we can judge) e.g. Bistro is pop ular in France
- Semantically similar categories for different cultures e.g. Churrascaria (Brazil) -> BBQ Joint (in other c ultures), Pastelaria -> Bakery, Snack Place

But latent similarities are not equivalence!
E.g. Souvlaki Shop (Greek) -> Kebab Shop: both are "fast-food", cooked on a spit BUT one is (traditionall y) made using pork

Further user studies are required