Culture-aware Point-of-Interest Completion

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

Nikolaos Lagos, Ioan Calapodescu

30 March 2020

DARLI-AP workshop

NAVER LABS

Multilingual & Multicultural POI DBs

What's the relation between *data imputation techniques*, the city of Paris, and a bowl of noodles?

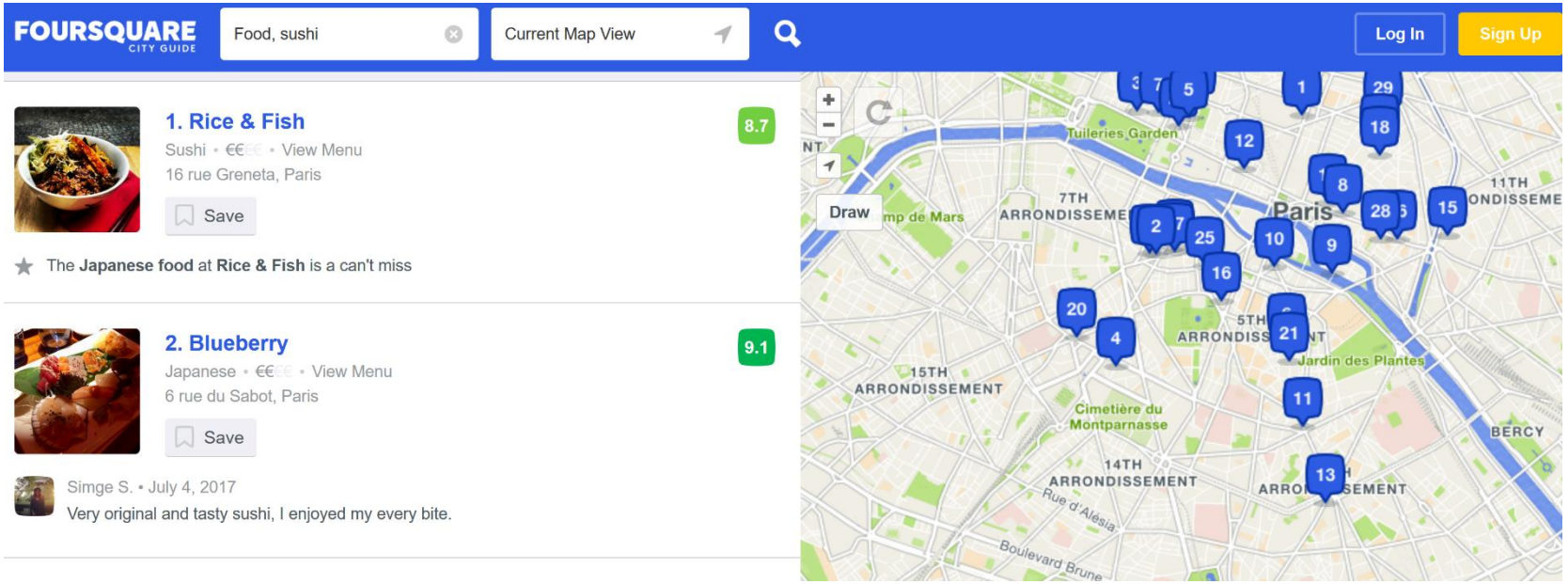
Point of Interest

“A point of interest, or POI, is a specific point location that someone may find useful or interesting. An example is a point on the Earth representing the location of the Eiffel Tower.”

Point of Interest

“A point of interest, or POI, is a specific point location that someone may find useful or interesting. An example is a point on the Earth representing the location of the Eiffel Tower”

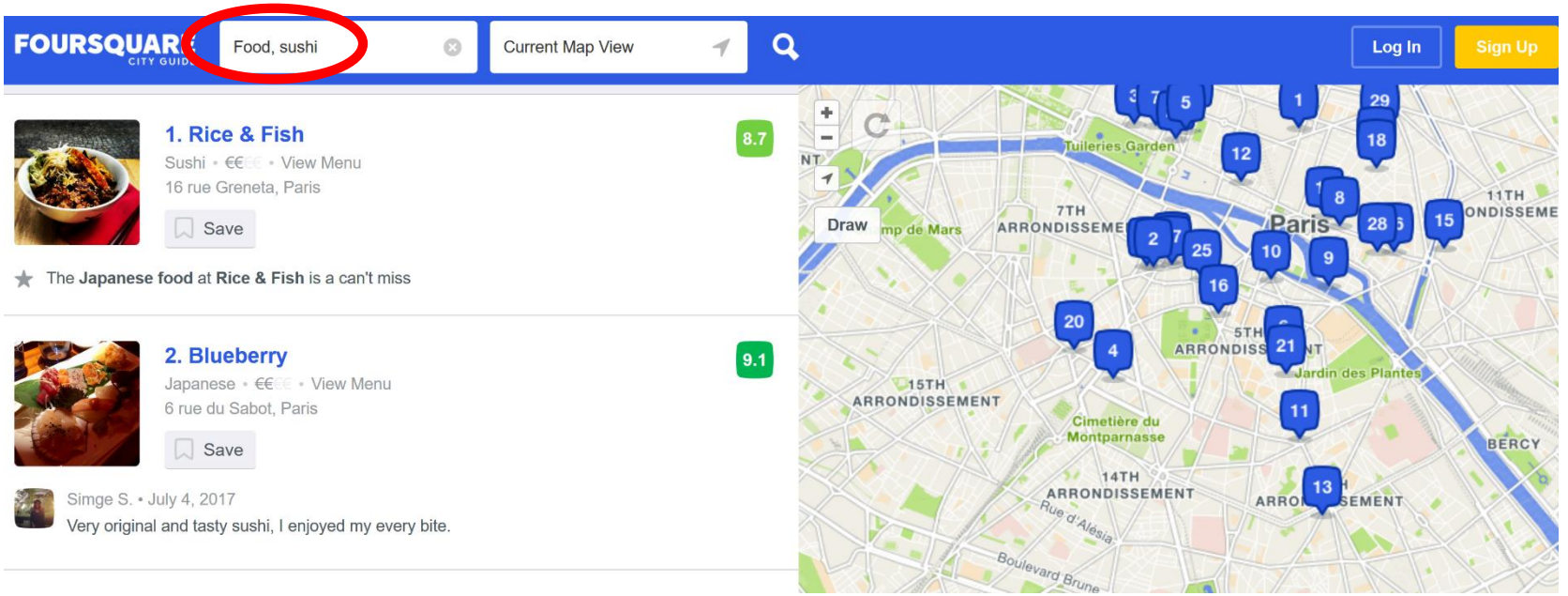
WIKIPEDIA
The Free Encyclopedia



Point of Interest Categories

“A point of interest, or POI, is a specific point location that someone may find useful or interesting. An example is a point on the Earth representing the location of the Eiffel Tower”

WIKIPEDIA
The Free Encyclopedia

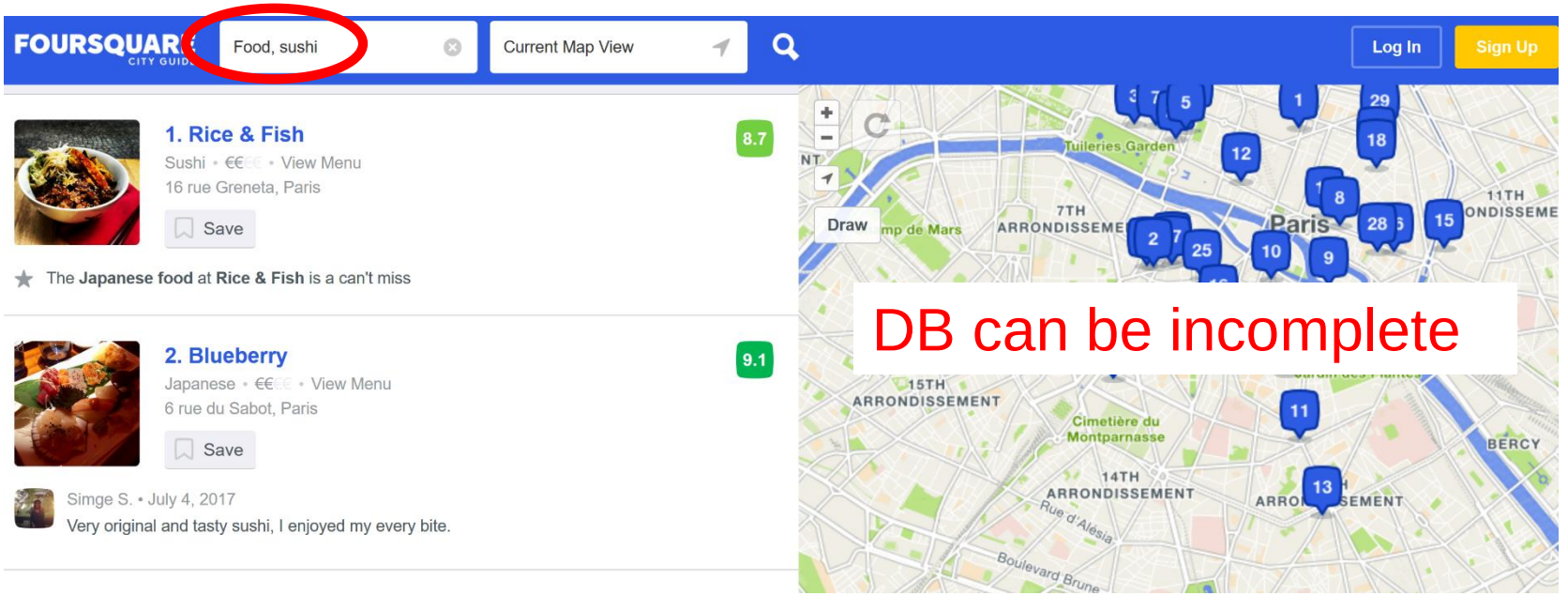


Point of Interest Categories

“A point of interest, or POI, is a specific point location that someone may find useful or interesting. An example is a point on the Earth representing the location of the Eiffel Tower”

WIKIPEDIA
The Free Encyclopedia

POI categories



Automated POI Categorization

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

Problem Definition

Goal

“La Table du Ramen” located in Paris, France

French Culture (Observed category): Japanese Restaurant

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

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

Category Prediction Method - Insight

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

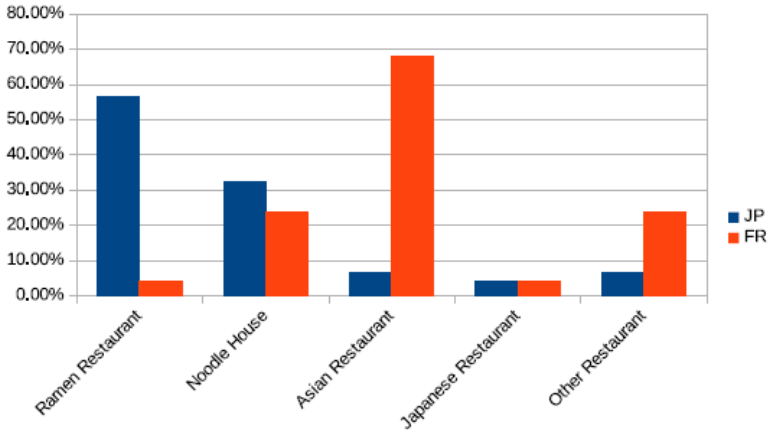


Figure 1: Category distribution of POIs having the token "noodle" in their name in Japan and France. It is obvious that "Ramen Restaurant" is the most popular category in Japan and "Asian Restaurant" in France.

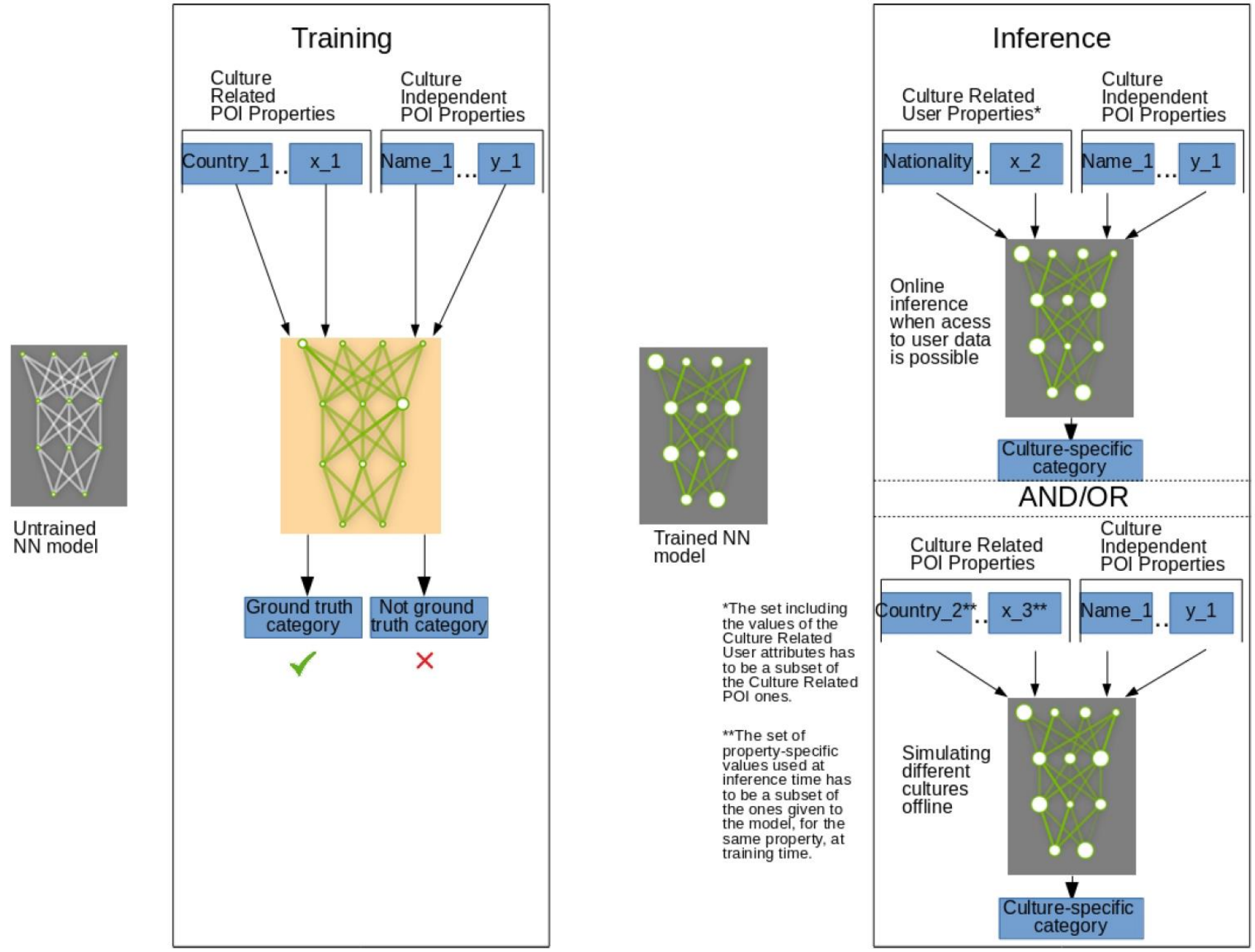


Figure 2: Overview of proposed method. Images of NN models are adapted from [3]

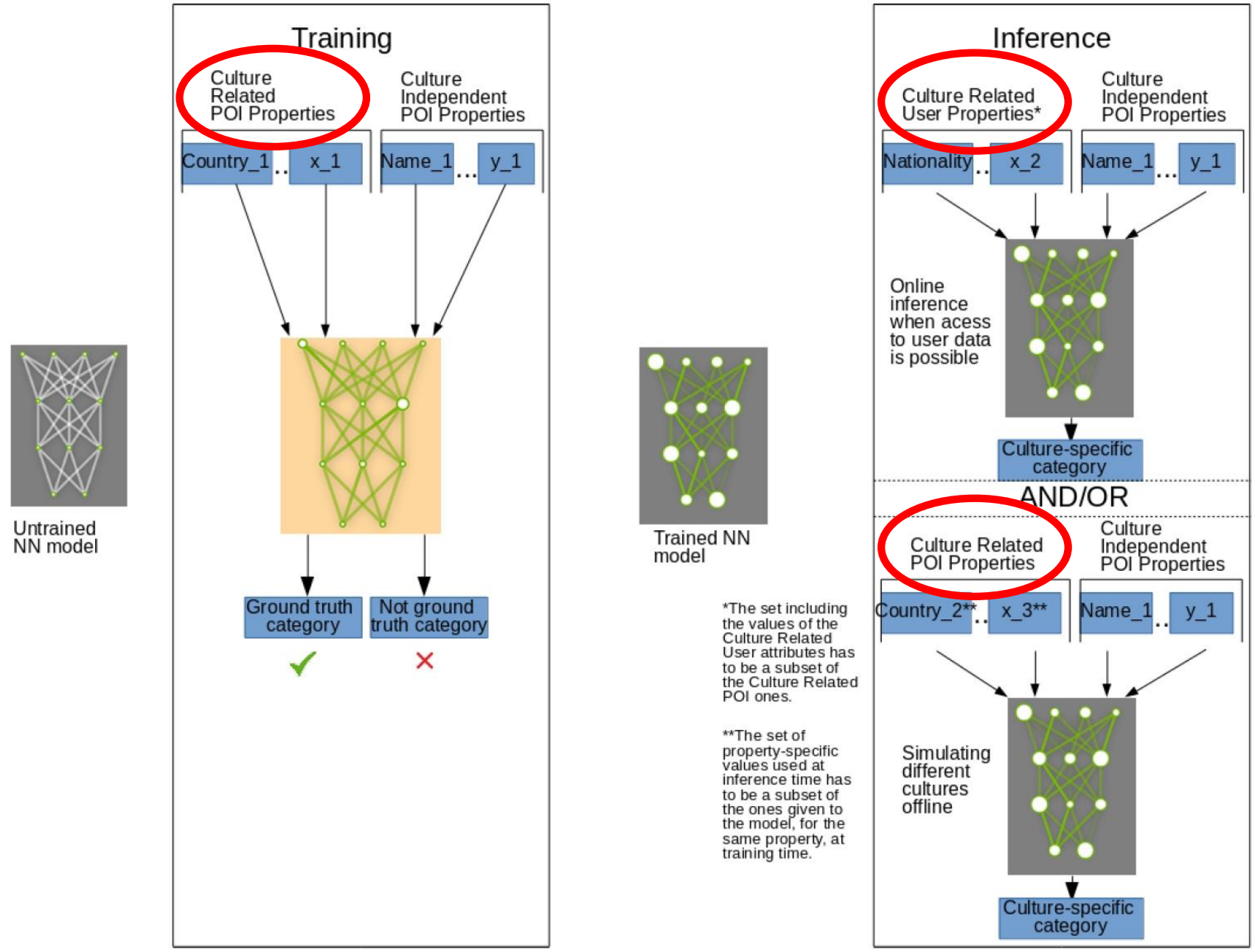


Figure 2: Overview of proposed method. Images of NN models are adapted from [3]

Category prediction method - Steps

Attribute selection: POI name, location (lat, long)

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

Training:

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

Inference:

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

Experiments - Data

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

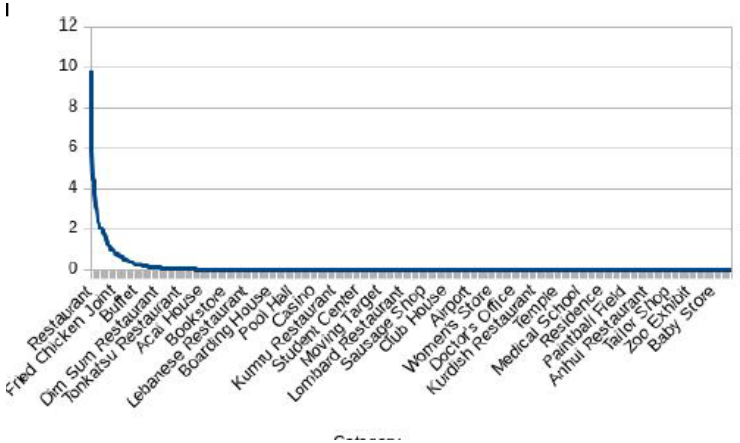
808 categories

<u>Attributes</u>

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
...	...



Experiments - Results

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 popular in France
- Semantically similar categories for different cultures e.g. Churrascaria (Brazil) -> BBQ Joint (in other cultures), 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 (traditionally) made using pork

Further user studies are required

Multilingual & Multicultural POI DBs

What's the relation between *data imputation techniques*, the city of Paris, and a bowl of noodles?

Culture-aware data imputation ☺

Q&A