Point-of-Interest Type Prediction using Text and Images

Danae Sánchez Villegas, Nikolaos Aletras
Computer Science Department, University of Sheffield, UK

POIs in Social Media

Text and **images** of social media posts shared by users from specific **points-of-interest (POI)** contribute to shaping a place's identity.
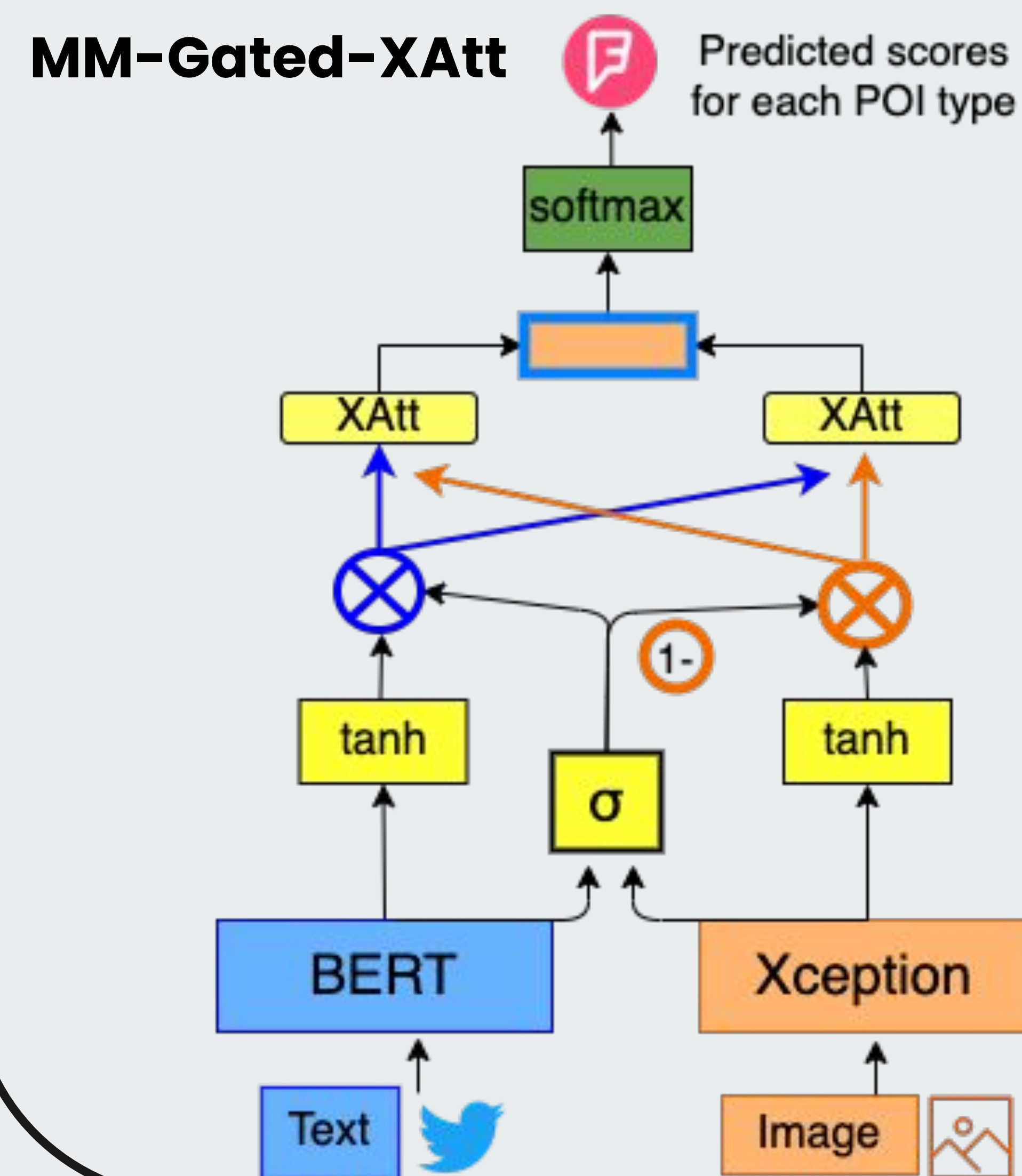
Next stop NYC ✈️ ← Text

← Image

POI {
Great Outdoors
Shop & Service
Travel & Transport
Food
}

from Tampa International Airport (TPA)

POI Type Prediction

MM-Gated-XAtt

F — Predicted scores for each POI type

softmax

XAtt — XAtt

⊗ — ⊗

tanh — σ — 1- — tanh

BERT — Xception

Text 🐦 — Image 🖼️

Models and Results

Macro F1

Model	Macro F1
BERT	43.67
ResNet	21.11
EfficientNet	24.72
Xception	23.64
Concat-BERT+ResNet	43.28
Concat-BERT+EfficientNet	41.56
Concat-BERT+Xception	44
Attention-BERT+Xception	42.89
Guided Attention-	41.53
LXMERT	40.17
Ensemble-BERT+LXMERT	43.82
MM-Gate	44.64
MM-XAtt	27.31
MM-Gated-XAtt	47.21

Ours
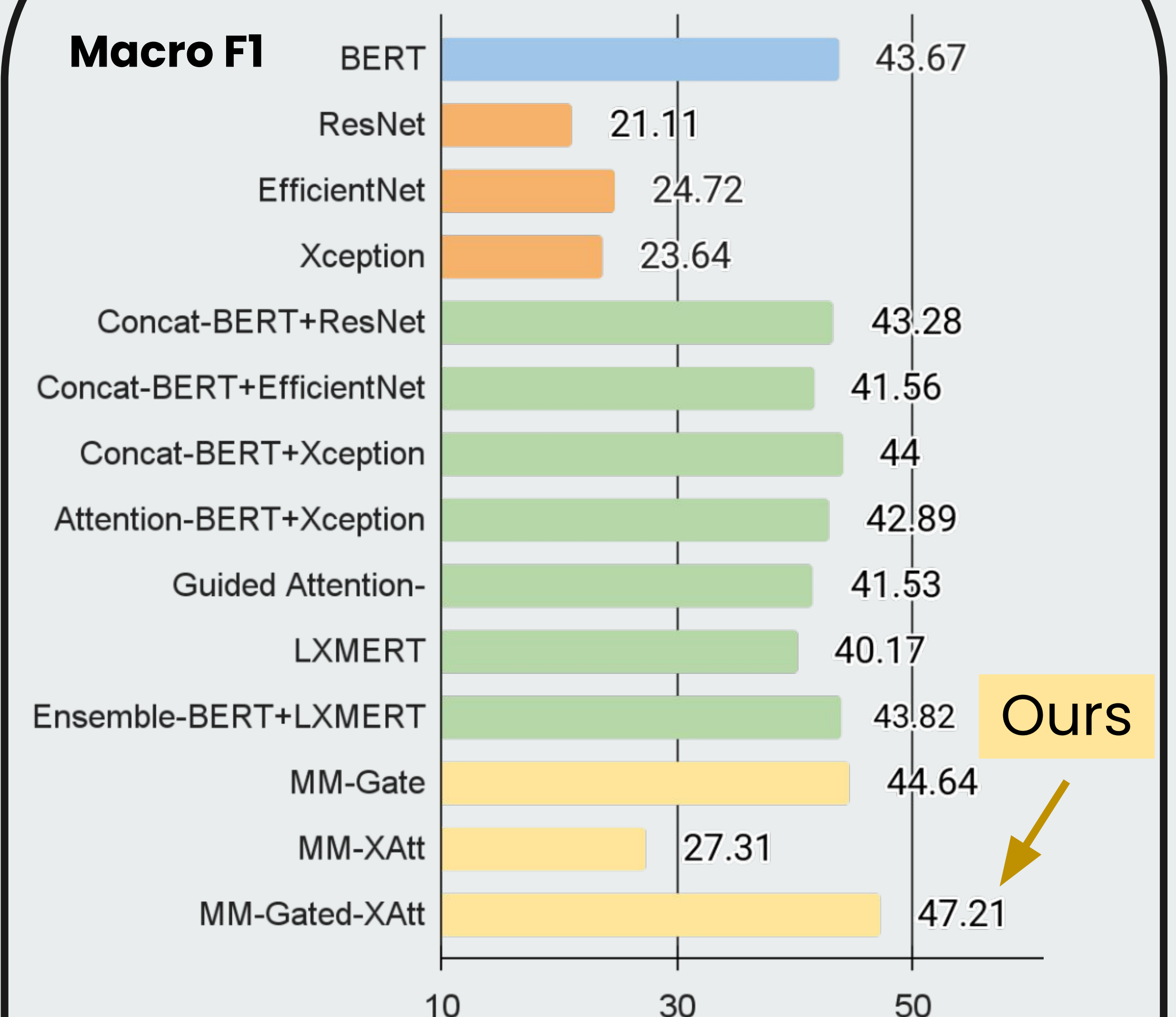
MM-Gated-XAtt combines text and image content in two levels:
1. gated multimodal fusion to manage the flow of information from each modality
2. cross-attention to capture interactions

MM-Gated-XAtt outperforms unimodal and multimodal approaches

Cross-Attention Analysis

it's getting cold up here <mention> <url>

Staff only, please!

BERT (text-only): Arts & Entertainment

✓ **MM-Gated-XAtt (Ours): Shop & Service**
- Text: 60% - Image: 40%
- Image information can help to address the ambiguity in short texts

Error Analysis

miso creamed kale with mushrooms <mention>

True: Nightlife Spot

MM-Gated-XAtt (Ours): Food

Most errors occur identifying for POI categories where people might perform similar activities in each of them such as:

- Food and Shop & Service
- Food and Nightlife Spot

Acknowledgements