Assessing the influence of point-of-interest features on the classification of place categories

Milias, V.; Psyllidis, A.

DOI
10.1016/j.compenvurbsys.2021.101597

Publication date
2021

Document Version
Final published version

Published in
Computers, Environment and Urban Systems

Citation (APA)
Milias, V., & Psyllidis, A. (2021). Assessing the influence of point-of-interest features on the classification of place categories. *Computers, Environment and Urban Systems*, *86*, Article 101597. https://doi.org/10.1016/j.compenvurbsys.2021.101597

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Computers, Environment and Urban Systems 86 (2021) 101597

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Assessing the influence of point-of-interest features on the classification of place categories

Vasileios Milias, Achilleas Psyllidis *

Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Keywords:
Point of interest
Classification
Feature importance
Feature extraction
POI categories

ABSTRACT

Points of interest (POIs) digitally represent real-world amenities as point locations. POI categories (e.g. restaurant, hotel, museum etc.) play a prominent role in several location-based applications such as social media, navigation, recommender systems, geographic information retrieval tools, and travel-related services. The majority of user queries in these applications center around POI categories. For instance, people often search for the closest *pub* or the best value-for-money *hotel* in an area. To provide valid answers to such queries, accurate and consistent information on POI categories is an essential requirement. Nevertheless, category-based annotations of POIs are often missing. The task of annotating unlabeled POIs in terms of their categories — known as POI classification — is commonly achieved by means of machine learning (ML) models, often referred to as classifiers. Central to this task is the extraction of known features from pre-labeled POIs in order to train the classifiers and, then, use the trained models to categorize unlabeled POIs. However, the set of features used in this process can heavily influence the classification results. Research on defining the influence of different features on the categorization of POIs is currently lacking. This paper contributes a study of feature importance for the classification of unlabeled POIs into categories. We define five feature sets that address operation based, review-based, topic-based, neighborhood-based, and visual attributes of POIs. Contrary to existing studies that predominantly use multi-class classification approaches, and in order to assess and rank the influence of POI features on the categorization task, we propose both a multi-class and a binary classification approach. These, respectively, predict the place category among a specified set of POI categories, or indicate whether a POI belongs to a certain category. Using POI data from Amsterdam and Athens to implement and evaluate our study approach, we show that operation based features, such as opening or visiting hours throughout the day, are the most important place category predictors. Moreover, we demonstrate that the use of feature combinations, as opposed to the use of individual features, improves the classification performance by an average of 15%, in terms of F1-score.

1. Introduction

From a computational perspective, points of interest (POIs) are digital proxies of real-world places, represented as geometric point entities. Nowadays, there is a wealth of online sources, of which POIs are integral components. Examples include geo-enabled social media (e.g. Twitter, Instagram), mapping applications (e.g. OpenStreetMap, Google Maps), travel and tourism-related platforms (e.g. Airbnb, TripAdvisor), among others. Each POI may be characterized by a set of features (often also referred to as attributes or properties). These features vary significantly across different data sources. Location (i.e. geo-coordinates), name, address, and category (i.e. the functional purpose of the establishment that each POI represents, such as *restaurant*, *hotel*, or *museum*) are the most common POI features. Other attributes may include business hours, accessibility information, reviews, ratings, interior and/or exterior pictures, among others.

Out of these features, *POI categories* play a prominent role in the service provision of the above-mentioned online applications. Users of these applications often perform category-based search queries, such as "what is the best *Italian restaurant* in this area?", "what is the closest *hotel* to the *train station*?", or "how can I go from my home to the *library*?". Having consistent and accurate information on POI categories is of critical importance to the functioning of such online applications, as well as to other domains such as site selection, real estate, and urban planning. Even though official registries of business establishments (e.g. business listings of national chambers of commerce) contain detailed

* Corresponding author.
E-mail addresses: V.Milias@tudelft.nl (V. Milias), A.Psyllidis@tudelft.nl (A. Psyllidis).

https://doi.org/10.1016/j.compenvurbsys.2021.101597
Received 8 June 2020; Received in revised form 5 January 2021; Accepted 8 January 2021
Available online 18 January 2021

category-related and other information of each entity, some of these annotations are often either missing in online POI registries or have coding issues (e.g. categorizing a bar as an arts organization, as discussed in Sect 3).

A common process for annotating unlabeled POIs, in terms of their categories, is *POI classification*, usually by means of machine learning (ML) classifiers (Choi, Song, Park, & Lee, 2020; Giannopoulos, Alexis, Kostagiolas, & Skoutas, 2020). Typically, this process involves the selection of a set of categories from already annotated POI data, collected from one or several sources. These categories are, then, used as classes (or labels) for the set of unlabeled POIs (i.e. multi-class classification) (Giannopoulos & Meimaris, 2020). The features that accompany each annotated POI, and which may provide insights into the characteristics of each category, are first extracted and, then, represented as feature vectors (Ding, Zhang, Pan, Wu, & Pu, 2020; Yan, Janowicz, Mai, & Gao, 2020; Zhang, Xiong, Kong, & Zhu, 2020). The features are subsequently fed into a classifier in order to train it and, then, the trained classifier is used for the classification of unlabeled POIs into categories. Existing works on POI classification use either several (Lu et al., 2020; Vandecasteele & Devillers, 2015; Zhou et al., 2020) or limited features (Choi et al., 2020; Funke & Storandt, 2020; Giannopoulos et al., 2020) to categorize new POIs.

One of the main issues with POI classification is that different features could *influence* differently the categorization results. Research on defining the influence of various features on the POI classification task is currently lacking. In other words, it is, to date, unspecified whether all features (e.g. location coordinates, business hours, ratings etc.) play an equal role in classifying a new POI as, for example, a restaurant, hotel or bar. And if this is not the case, then what is the individual contribution of each feature to assigning a category to a POI? Is the ranking of those individual contributions the same in relation to different POI categories? And would this vary regionally, depending on the context within which a POI is located (e.g. differences by city, region, country etc.)?

In this paper, we contribute a study of feature importance for the classification of unlabeled POIs into categories. We extract features from several data sources; namely, Google Places, Foursquare, Twitter, and publicly available street-level imagery, with the aim to create semantically-rich descriptions of the POIs that are used in training our classifiers. We, further, organize the extracted features into five feature sets, based on the POI aspects they represent. Specifically, we define (1) *operation-based* features that, among others, describe the opening and visiting hours, (2) *topic-based* features that refer to topics extracted from textual user-generated content (i.e. tweets) within a radius surrounding each POI at hand, (3) *review-based* features that refer to reviews, ratings, and sentiments, (4) *neighborhood-based* category features that correspond to the categories of neighboring POIs, and (5) *visual* features that describe the external appearance (i.e. based on facade elements) of POIs. Some of the aforementioned features (e.g. categories, business hours) are extracted directly from the sources through their corresponding Application Programming Interfaces (APIs), while others (e.g. reviews, topics from tweets, sentiments, visual features) are extracted by means of natural language processing and topic modeling techniques, as well as through image recognition algorithms using deep learning models.

We incorporate the extracted features and feature sets in several ML-based classifiers. In addition to common practices, which solely follow *multiclass* classification approach (i.e. a classification which predicts the POI category from a given set of labeled POIs), we further apply a *binary* classification, which defines whether a POI belongs to a certain category or not. This allows for an assessment and ranking of the features that influence POI categorization the most. We implement and evaluate our method using POI data for the cities of Amsterdam (the Netherlands) and Athens (Greece) – a North and a South European city, respectively. Both constitute primate cities in the corresponding national urban system hierarchies, though with different population sizes, especially at the metropolitan level (i.e. Athens metropolitan region is about 3 times larger than the Amsterdam metropolitan region, in terms of population size). We, therefore, decided to focus on the city-center areas of both Amsterdam and Athens, which have similar population size and concentration of amenities and business establishments (see Section 3). Our empirical analyses show that *operation-based* features (e.g. opening and closing hours) yield high POI classification performance, in terms of F1-score (i.e. a weighted average of the precision and recall), followed by *review-based* features (e.g. reviews, ratings). Contrariwise, topic-based and visual features contribute the least to the prediction of a POI category. In all cases, the *combination* of the features yields the highest classification results (by an average of about 15% in both classification approaches), meaning that richer POI profiles lead to more accurate POI categorizations.

The remainder of the paper is structured as follows: Section 2 discusses related work on POI classification and the use of various features for POI categorization and enrichment. In Section 3 we outline the data collected from various sources, focusing on urban areas of Amsterdam and Athens. Section 4 introduces the proposed pipeline for POI data matching and feature extraction, as well as the classification approaches. Section 5 presents (a set of) the results obtained through the various experiments. A discussion of the findings is presented in Section 6. Finally, Section 7 summarizes the conclusions and discusses future lines of research.

2. Related work

This section presents related work on state-of-the-art approaches to POI classification, in terms of annotating unlabeled POIs or enriching pre-labeled ones. Common approaches to POI classification focus on ML classifiers and feature extraction methods, using POI profiles with either limited or rich features to describe and categorize new POIs. In the case of ML-based POI classifications with limited metadata, POI features such as the name and location (i.e. geo-coordinates) (Giannopoulos et al., 2020), as well as textual features surrounding the POI name (Choi et al., 2020) played an important role in categorizing POIs or in describing them in further detail (e.g. distinguishing restaurant categories) (Funke & Storandt, 2020). In the case of classification using rich POI features, (Lu et al., 2020) combined mobility patterns and fare amounts from taxi data together with location features and POI categories to infer the lifetime status (booming, decaying, stable) of POIs.

(Vandecasteele & Devillers, 2015) used the semantic similarity of various POI labels to automatically provide OpenStreetMap users with recommendations of richer POI features. (Zhou et al., 2020) further combined user profile and map query data to refine POI labels.

More recent approaches to POI classification use POI or context embeddings (i.e. low-dimensional, continuous vector representations of variables relating to the spatial context of POIs, the co-occurence frequency of categories, and the semantic similarity of names, among others) in the training process, in place of extracted features. These are, further, combined with deep learning (DL) methods to solve the classification task. (Yan et al., 2020) use POI embeddings, based on co-occurrences of neighboring POI categories, in order to predict the similarity between POI categories, in relation to a given category-based hierarchy derived from Yelp. Context embeddings have, further, been used in related classification tasks that, nevertheless, go beyond POI categorization. (Cocos & Callison-Burch, 2020) use data from Google Places and OpenStreetMap to derive geographic contexts for the enrichment of geo-referenced Twitter microposts, by means of word embeddings. (Yao et al., 2017; Zhang et al., 2020) use embeddings related to co-occurrence and neighborhood relationships to, respectively, study the distribution of land uses across space and the spatio-temporal dynamics of human activities. Deep learning methods in combination with context embeddings have been used in tasks of POI enrichment (Xu et al., 2020; Ziegler et al., 2020) and recommendation (Zhao, Zhao, King, & Lyu, 2020).

With regard to enrichment processes, various POI features have been used in classification tasks. (McKenzie, Janowicz, Gao, & Gong, 2015)

use temporal and operation-based POI features to discover the spatio-temporal behaviour of people towards different places and its regional variation. Features extracted from user-generated text of geo-referenced microposts, reviews, and Wikipedia articles has been used in the extraction of neighborhoods' activity signatures (Fu, McKenzie, Frias-Martinez, & Stewart, 2018), the identification of a collective sense of place (Jenkins, Croitoru, Crooks, & Stefanidis, 2016), the quantification of tourist attractions' similarity (McKenzie & Adams, 2018), and in the detection of city regions, based on the social interactions of people, and the prediction of future POI locations (Psyllidis, Yang, & Bozzon, 2018). Visual POI features extracted from street-level imagery have been used to map urban objects and to infer properties of urban areas. Examples include the mapping of urban greenery (Li et al., 2015; Liu, Silva, Wu, & Wang, 2017; Qiu, Psyllidis, Bozzon, & Houben, 2020), the inference of subjective properties, such as liveliness, perception of safety, and attractiveness (Fu, Chen, & Lu, 2020), the qualities of the urban environment (Liu et al., 2017), and the perception of urban regions by people (Zhang et al., 2018). Despite the various approaches to POI classification and the use of numerous features in several applications, none of current literature examines how different features impact the classification task. To the best of our knowledge, our work is the first to assess the influence of different features on the classification of POIs into categories. Our study addresses this gap by extracting a large variety of POI features from different online data sources, and by analyzing the contribution of each feature — and combinations of features — to the categorization of POIs, towards accurate and consistent POI labeling.

3. Data

In generating a rich POI dataset that covers the five feature sets listed in the Introduction, we collected and combined data from several online sources. The selection of the data sources is built upon three basic requirements: (1) the *existence* of common POI attributes across the data sources, which would allow for matching different features to a single POI, (2) the *diversity* of their nature, to mitigate the potential bias introduced by sources that are similar in nature (e.g. combination of both user-generated and non-user-generated content) (Fu et al., 2018), and (3) the *coverage* of all the feature sets listed in Sect. 1, when the data sources are combined.

Based on the above requirements, we collected data from Google Places, Foursquare, Twitter, and publicly available street-level imagery, using the corresponding APIs, across selected urban areas of Amsterdam (the Netherlands) and Athens (Greece). The selection of the cities under study is based on the availability of data and on our familiarity with them, which facilitates a qualitative interpretation of the results. Even though they both constitute primate cities, the population size at the metropolitan level differs substantially. For this reason, we focus on the city-center areas in both cities, as delineated in (Fig. 1), which are characterized by similar population size and concentration of amenities and business establishments. At the same time, the differences in terms of the spatial configuration of the urban fabric and the aforementioned amenities allow us to evaluate the potential impact of the urban spatial structure on the obtained results, and the generalizability of our methodology.

POI data from Google Places and Foursquare contain *operation-based* (e.g. opening hours, visiting hours etc.), and *review-based* (e.g. reviews, ratings etc.) features. Based on the latitude and longitude of the collected POIs, we further retrieved the categories of nearby places to address *neighborhood-based* features. A total of 109,539 POIs (64,906 in Amsterdam, and 44,633 in Athens) were retrieved from the Google Places API, whereas 26,918 POIs (15,624 in Amsterdam, and 11,294 in Athens) were retrieved from the Foursquare Places API.

We implement a *data matching* step (described in Sect. 4) after the "data collection phase A" part (Fig. 4). This step is required in order to mitigate the existence of different taxonomies and the categorical misalignment between different data sources. In particular, Foursquare organizes POI categories into a two-level hierarchy. Higher-level categories describe a POI more generally, such as restaurant, hotel, museum or stadium. Lower-level categories specify even further what the POI is about (e.g. Italian restaurant, airport hotel, archaeological museum, football stadium etc.). There exist 10 high-level categories, which are further specified into 2881 lower-level categories. On the other hand, Google Places organizes POIs into 96 categories without including any subcategories, at least in the publicly available API[1].

In addition to these organizational schemes, category annotations of the same POI could sometimes vary across sources. For example, Ruigoord Kerk in Amsterdam is categorized as a bar in Foursquare[2], whereas Google Places[3] uses the term arts organization to describe the same place.

Considering the number of instances per POI category in the newly created matched dataset our analysis was conducted on the basis of the ten most frequent POI categories. The total number of Google and Foursquare POIs under consideration was, therefore, reduced to 5755 (3304 in Amsterdam, and 2451 in Athens) (Figs. 2, 3).

The geographic coordinates of those matched POIs were used for collecting tweets, generated within a 50 m radius from each POI location (see "data collection phase B", Fig. 4). Twitter data are used for the extraction of *topic-based* features (e.g. topics discussed in the vicinity of a POI at hand), with the aim of capturing the context surrounding a POI, which could in turn contribute to defining its category. The reasoning behind this is that, besides the knowledge of categories of neighboring labeled POIs (i.e. defined in the neighborhood-based feature set), which could sometimes be outdated, usergenerated content created in the vicinity of a POI could give an indication of its category (e.g. retail stores tend to cluster) (Janowicz, McKenzie, Hu, Zhu, & Gao, 2019). We collected tweets in the period between January 1, 2017 and October 20, 2018. For each tweet, we collected the text, language, creation date, number of retweets, number of likes, and geo-location. For the purposes of this work, we selected tweets written either in English or in the official language of the country where each of the case-study cities belongs to (i.e. Dutch for Amsterdam, and Greek for Athens). The number of tweets, generated within a 50 m radius from each of the 5755 POI locations, reached a total of 214108 (120250 in Amsterdam, and 93858 in Athens).

Lastly, the geo-coordinates of the matched POIs are also used for collecting 360° panoramic images at ground level, from which *visual* features are extracted. Each location is represented by four perpendicular images, altogether forming a 360° panorama, so as to extract facade elements of POIs and their immediate surroundings. This, further, mitigates the bias introduced when representing a POI location with a single image (e.g. its street frontage). In some occasions, street-level images depict the same location on different dates. In this case, we only keep the most recent ones. We collected a total of 22944 street-level images (13104 in Amsterdam, and 9840 in Athens).

4. Method

This section presents the proposed approach to assessing the influence of different features on the classification of POIs into categories. The architecture of the developed pipeline (Fig. 4) consists of three main components: (1) the *Data Collection and Matching* module gathers data from various sources and, subsequently, combines them by identifying common POI attributes through a matching algorithm, (2) the *POI*

[1] Besides Foursquare and Google Places, the multilevel POI categorization scheme is further used by other popular location-based services such as Yelp, which organizes POIs into a hierarchy of 22 higher-level categories with more than 1200 subcategories.

[2] https://foursquare.com/v/ruigoord-kerk/4f1e2df66aa3156a2ef46a90.

[3] https://www.google.com/maps/place/Ruigoord/@52.410303,4.746685,17z/data=!3m1!4b1!4m5!3m4!1s0x47c5e4940400f37f:0xc9af93d6d0acfbfe!8m2!3d52410,303!4d4.7488737.

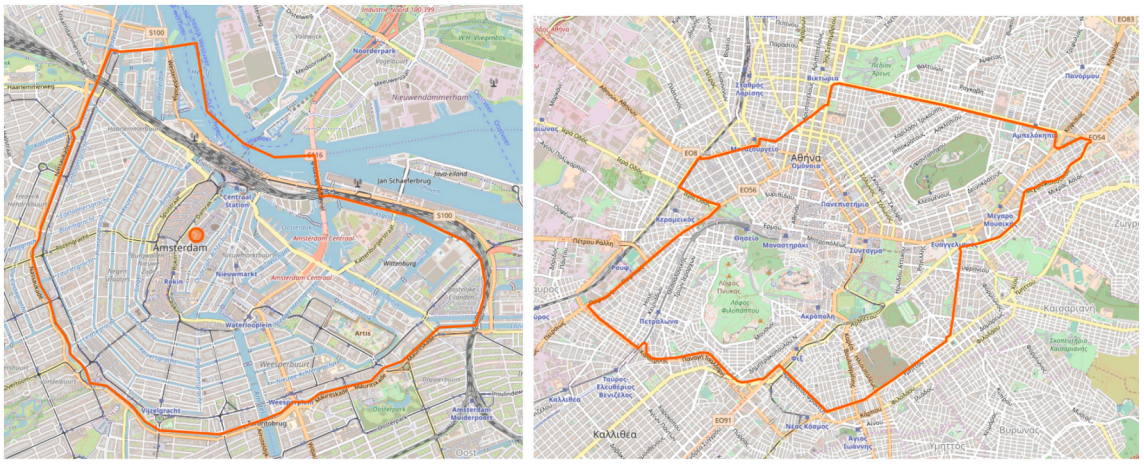

Fig. 1. The selected areas under study include the center of Amsterdam (left) and the inner ring of Athens (right).

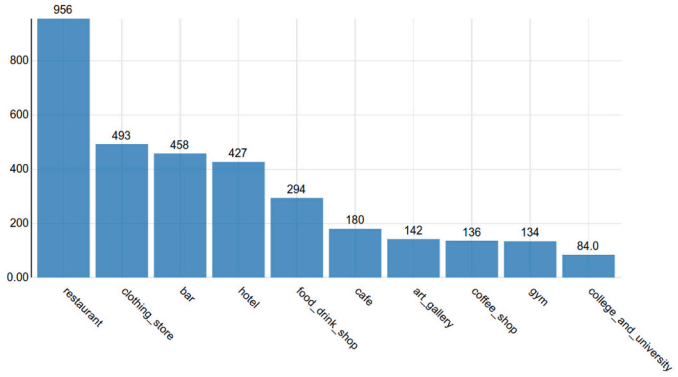

Fig. 2. Number of selected POI instances in Amsterdam per category.

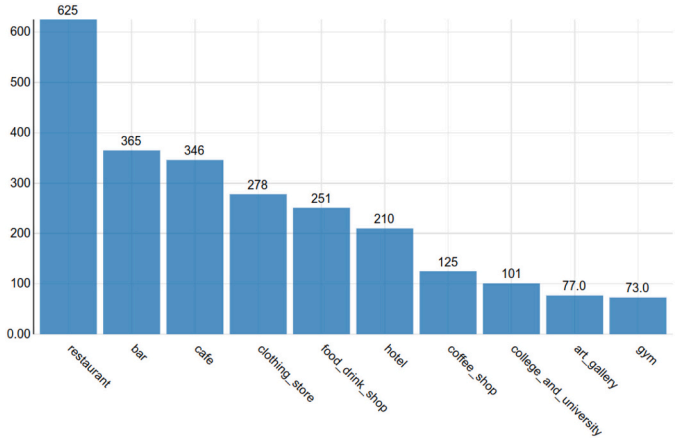

Fig. 3. Number of selected POI instances in Athens per category.

Feature Extraction module mines the various POI features from the combined dataset, and (3) the *Classification and Ranking* module predicts the category of a POI, based on the previously extracted POI features, and generates a ranking of the features that contribute the most to the classification of the place categories. The following paragraphs present each of the aforementioned components in further detail.

4.1. Data collection and matching

The data collection and matching module comprises two sub-modules that are, respectively, designed for gathering POI data from a variety of sources, and for matching them on the basis of common features across the data sources. The data collection part of this component has been discussed in detail in Sect. 3. Therefore, this sub-section focuses on the matching process.

Matching refers to the process of identifying whether POI entities belonging to different data sources correspond to the same physical place (McKenzie, Janowicz, & Adams, 2020). Given the nature of the data sources used in this work, a first matching is carried out with POIs collected from Google Places and Foursquare. As mentioned in Sect. 3, we collect tweets within a 50 m radius of each matched POI, and street-level images for each matched POI. After matching the Google Places and Foursquare POIs, all the data sources used in this work are eventually matched together. The resulting combined dataset comprises enriched POI entities of physical places with attributes that span the feature sets considered here.

The matching algorithm developed in this work is inspired by the tree structure approach introduced in (Jiang, Alves, Rodrigues, Ferreira Jr, & Pereira, 2015). The common POI features used for performing the matching are the *geo-coordinates*, *name*, *address*, and *phone*. In the few occasions where some of the aforementioned features are missing, matching is carried out on the basis of POI geo-coordinates and name, which are always present. A set of similarity metrics are used for comparing each of the feature values between the sources. Regarding geo-coordinates, a buffer of 300 m radius is used as a threshold geographical distance between the POIs. That is, for each Foursquare POI, we retrieve all the Google Place POIs within a radius of 300 m. Then, every Foursquare POI is compared with each Google POI, based on the remaining attributes mentioned above.

Given that the name, address, and phone attributes are essentially strings, from a data type perspective, we apply a set of string similarity metrics to facilitate the comparison (part of which, inspired by (McKenzie et al., 2020)). Specifically, we use the (a) Levenshtein distance, (b) Damerau-Levenshtein distance, (c) Phonetic similarity, (d) Ratcliff and Obershelp's algorithm, and (e) Longest subsequence metric. The overall string similarity score of each string-type feature is calculated as the mean of all the above-mentioned string similarity scores.

After the matching algorithm groups the POIs that belong to each source and are less than 300 m apart, it generates three scores based on: (1) name similarity and distance, (2) name similarity and street name similarity, and (3) name similarity, phone number and distance. If these scores are higher than given predefined thresholds, the POIs are matched. If more than one pair of POIs could be matched, the one with the higher score is selected. In defining the corresponding rules and thresholds, we follow a heuristic approach. The evaluation of the matching algorithm is based on a sample of 200 POIs for each city and is

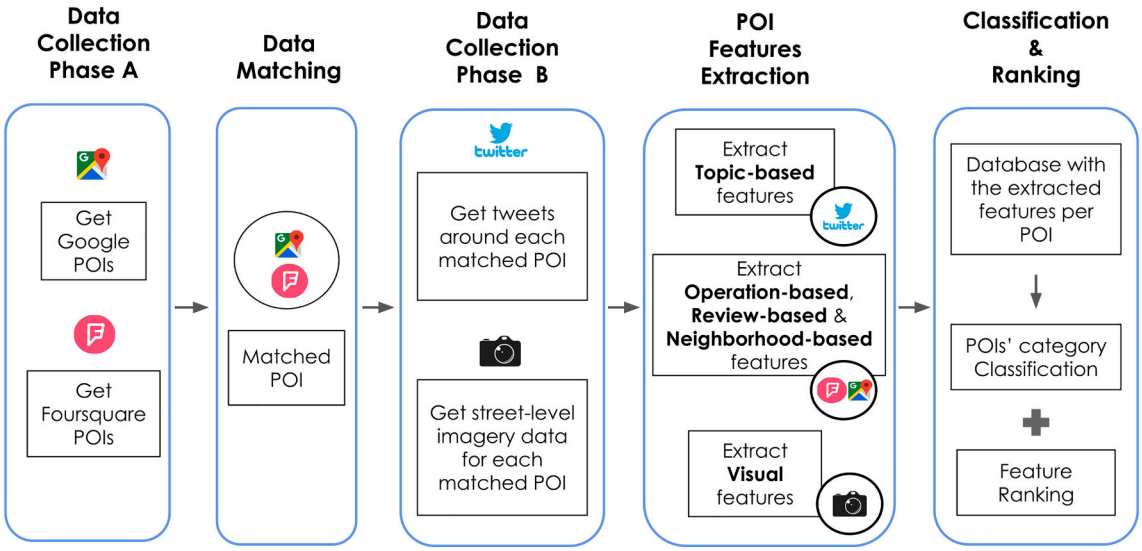

Fig. 4. Pipeline of the proposed approach: (1) Data Collection and Matching, (2) POI Features Extraction and (3) Classification and Ranking.

performed manually. The algorithm achieves 97% accuracy of correct matches in the case of Amsterdam, and 98% in the case of Athens.

In the resulting dataset of matched POIs, there exist several POI categories that only occur once or twice. To gain better insight of the contribution of features to the classification of POI categories, the following ten most frequent categories are used in the analysis: Hotel, Bar, Coffee Shop, Restaurant, Cafe, Clothing Store, Art Gallery, Food and Drink Shop, Gym, and College and University.

4.2. POI feature extraction

After matching POI data from various sources, a set of features is extracted from the resulting matched dataset, in order to ultimately assess the contribution of each feature to the classification of POI categories. As discussed in Sect. 1, the features are organized into five broader sets. This sub-section describes the feature extraction process for each of the aforementioned set.

4.2.1. Extraction of operation-based features

The features belonging to this category refer to a place's *opening hours*, *website*, *phone number*, *price range*, and *visiting hours*. These features are directly extracted from the Google Places and Foursquare APIs. Given that each source adopts different levels of detail, regarding the description of a POI category, in this work we make use of the more specialized categorizations (e.g. restaurant, hotel etc.), as opposed to higher-level classifications (e.g. Travel and Transport). In cases where there is misalignment of the POI category between Foursquare and Google Places, we use the Foursquare POI category, given that it is more accurate than the one of Google Places (Martí, Serrano-Estrada, & Nolasco-Cirugeda, 2019). For the analysis of the temporal features, we aggregate *visiting hours* into four time windows: morning (05:00–12:00), afternoon (13,00–17,00), evening (18,00–21,00), and night (22:00–04:00).

4.2.2. Extraction of topic-based features

The extraction of the topic-based features aims at capturing the context surrounding a POI. All features belonging to this category are extracted from the collected Twitter data and, namely, refer to the: *number of tweets* around each place and per language, *average number of words*, *sentiment*, *average time difference* between consecutive tweets around each place, and *topics*. For the extraction of sentiments, the TextBlob[4] (for tweets in English) and Polyglot[5] (for tweets in Dutch and Greek) natural language processing Python libraries are used.

To derive topics discussed in the tweets, we train several Latent Dirichlet Allocation (LDA) models. LDA is an unsupervised probabilistic model that enables the discovery of latent topics in a given textual corpus, consisting of documents (Blei, Ng, & Jordan, 2003). In this work, the text corpus comprises all the tweets that have been collected in the two use-case cities, whereas each document refers to the set of tweets collected around each POI. After pre-processing the collected tweets (i.e. language-based stopword and punctuation removal, lemmatization, and conversion of emojis into text), a probabilistic distribution of topics is assigned to each POI. Thereby, POIs are expressed as unique topic probability patterns. In total, four LDA models are built; one for each city and language (i.e. Dutch and English for Amsterdam, and Greek and English for Athens). Given that in order to train an LDA model, the number of topics (k) has to be predefined, we calculate optimal k -topics using the *coherence* and *perplexity* metrics of the trained models for several k -values. Drawing on this, we train the LDA models with $k = 10$ topics. For instance, some examples of the probabilistic distribution of topics are:

- 0.047 * noordholland (north holland) + 0.031 * goed (good) + 0.019 * man +0.018 * cafe +0.012 * north (Amsterdam).
- 0.091 * day +0.081 * love +0.023 * hotel +0.020 * coffee islandco + 0.019 * thissio (Athens).

4.2.3. Extraction of review-based features

We extract a number of features that approximate the experience of people in different places. These features could include the *ratings*, number of *likes*, number of *photos* taken at a POI, and visitors' *reviews*. Ratings are used here as a proxy for the quality of people's experience in a given place. Likes and photos attached to a POI could indicate its popularity. Lastly, reviews about a place could give insight into the variety of experiences across different social groups (e.g. age range, locals or tourists etc.), and people's sentiments. In extracting these latent features from the written reviews, we employ a set of LDA models and sentiment analysis tools, as described in the previous paragraph. For instance, some examples of the probabilistic distribution of topics

[4] https://textblob.readthedocs.io/en/dev/.
[5] https://pypi.org/project/polyglot/.

include:

- 0.048 * room +0.035 * hotel +0.022 * staff +0.021 * location +0.018 * clean (Amsterdam).
- 0.030 * room +0.026 * hotel +0.018 * view +0.017 * staff +0.015 * breakfast (Athens).

4.2.4. Extraction of visual features

The extraction of visual features has a twofold aim: first, to gain insight into the external appearance of POIs (i.e. how POIs look like from the outside); and second, to detect and map urban objects (e.g. trees, signs, visual labels etc.) on the facade or the surroundings of a POI, which could indicate the category of a place. In accordance with the previous aims, this component of the pipeline is further split into two parts: a *scene recognition* and an *object detection* part, respectively addressing the two goals.

Street-level imagery is the source from which visual features are extracted. In this work, we take a deep learning approach to achieve this. We make use of pre-trained and openly available state-of-the-art deep learning models. Specifically, for scene recognition, we employ the Places Database (Zhou, Lapedriza, Khosla, Oliva, & Torralba, 2017), which includes 1,803,460 images labeled with 365 scene categories, together with a pre-trained deep Residual Network (ResNet) model with 50 layers. For object detection, we again use deep ResNet models, pre-trained on the Google Open Images[6] and COCO[7] datasets. The former contains around 9 M images with 600 box classes, whereas the latter includes 2.5 M labeled instances of 328 K images. Given that each POI is represented by four street-level images (together forming a 360° panorama), the set of visual features of a POI results from the aggregation of the recognized scenes and detected objects on each individual image of the panorama.

4.2.5. Extraction of neighborhood-based features

This set of features refers to the number of neighboring POIs that belong to the same category as the POI at hand. We consider this specific set of features, given the well-known co-location patterns of specific POI categories (e.g. retail stores and food businesses) (Koster, Pasidis, & van Ommeren, 2019; Sevtsuk, 2014). Thereby, knowledge of nearby places could help indicate the category of an unlabeled POI. To achieve this, for each POI we retrieve the number of places of a given category that exist within a radius of 100 m, 1 km, and 5 km, and create an index of co-located places with the POI at hand.

4.3. Classification and ranking

Following the extraction of POI features, the last component of the proposed pipeline includes the classification and ranking modules. These modules assess the influence of the various features on the classification of POIs and, correspondingly, generate a feature ranking. We follow two POI classification approaches: (1) prediction of POI categories among a specified set of categories (*multi-class* classification) and (2) exploration of the possibility that a POI belongs to a certain category (*binary* classification).

The selection of classifiers is perplexed by the large number of features used and the different nature of each feature. After exploring our data we arrived at the following requirements: the selected classifier should be able to handle high-dimensional data (i.e. data with several features), missing data, and provide a clear and concrete method to calculate the contribution of each feature. Note that our overall goal is not to obtain the highest classification results, but to gain a better understanding of how each feature contributes to the identification of a POI category. In this light, we use tree-based classifiers which fulfill all the above requirements.

To evaluate our reasoning, we trained different types of classification models that have frequently and successfully been applied to relevant multi-class problems. Specifically, we trained the following models: Linear Discriminant Analysis (lda), Support Vector Machines (SVM), k-Nearest Neighbor (kNN), Random Forest (RF), eXtreme Gradient Boosting (XGB), and an Ensemble classifier, using majority voting on the results of the four best-performing classifiers. The selected list of classifiers does not include deep learning models for two reasons: first, the amount of the data is not that large to support a deep learning solution and second, the machine learning approaches include a clearer understanding of the features' importance which is crucial for the feature ranking part.

As expected, the XGB, a decision-tree-based classifier, and the ensemble method led to similar results and outperformed all the other classifiers, in terms of F1-score. As an example, when dealing with the problem of predicting the category of a POI among a specified set of categories, using a classifier trained on the extracted POI features, the XGB and ensemble classifier lead to an F1-score of around 60%, followed by the RF and LDA classifiers which scored around 52%, for both cities (Amsterdam results depicted in Fig. 5). In handling the imbalanced dataset, two approaches are followed: (a) a datacentric approach, based on the Synthetic Minority Oversampling Technique (SMOTE) (Chawla, Bowyer, Hall, & Kegelmeyer, 2002), and (b) a model-centric approach, based on adjusting the weights of the XGB classifier. After performing extensive experiments, the XGB classifier, with adjusted weights for handling imbalances in the dataset was selected.

5. Results

In this section, we quantify and rank the influence of the features on POI classification, using the two approaches mentioned in Section 4, i.e. multiclass and binary. In both approaches, the focus is on: (1) the *feature set contribution*, where the extracted features are studied and ranked per category (i.e. operation-based, topic-based, review-based, neighborhood-based, and visual), and (2) the *individual feature contribution*, where the features are analyzed individually. To evaluate the performance of the selected classifier (i.e. XGB), four metrics are used: (a) accuracy, (b) precision, (c) recall, and (d) (macro) F1-score. The *F1-score* is emphasized and is considered to be the most valuable metric as it takes into consideration both precision and recall, and it is not affected by imbalances in the dataset. A qualitative interpretation of the obtained results is discussed in Section 6.

5.1. Multi-class classification

The aim of the multi-class classification is to assess how different POI features contribute to predicting the POI category, among the ten selected categories (i.e. Hotel, Bar, Coffee Shop, Restaurant, Cafe, Clothing Store, Art Gallery, Food and Drink Shop, Gym, and College and University).

5.1.1. Feature set contribution

Table 1 presents the performance of the XGB classifier when trained on each feature set separately, and when all feature sets are combined. As expected, the combination of all feature sets improves the performance of the classifier (by an average of 27%) for both cities, achieving an F1-score of 0.606 for Amsterdam, and 0.609 for Athens. In both cities under study, the most important feature set is the *operation-based*, closely followed by the review-based (achieving an F1-score of 0.464 for Amsterdam, and 0.365 for Athens) while the least important ones are the topic-based (F1-score of 0.194 for Amsterdam, and 0.143 for Athens) and the visual (0.145 for Amsterdam, and 0.154 for Athens).

Furthermore, a direct comparison of the two cities is realized, so as to evaluate the effect of regional characteristics on the ranking of the feature set contribution (Table 2). The classifier performs better for

[6] https://storage.googleapis.com/openimages/web/index.html.
[7] http://cocodataset.org.

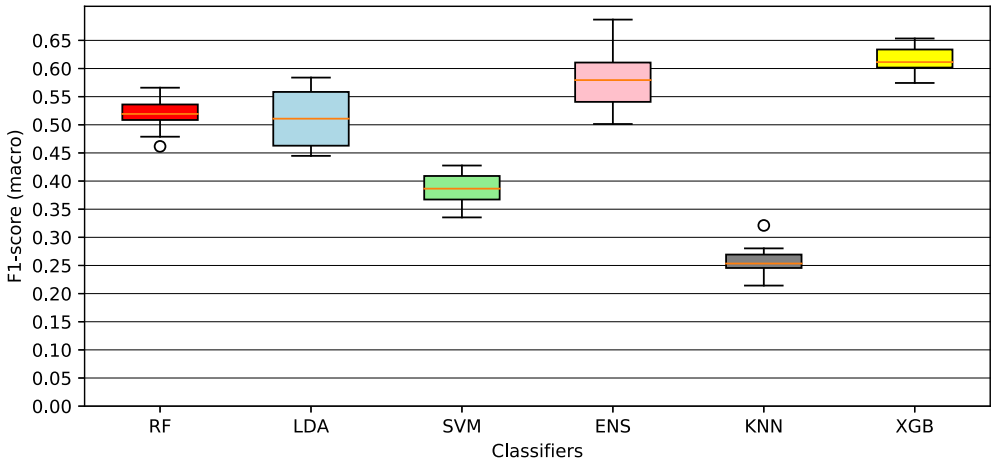

Fig. 5. Amsterdam - F1 (macro) Score per Classifier. The predicted classes (Set A) are: restaurant, clothing store, bar, hotel, food and drink shop, cafe, gym, coffee shop, college and university, art gallery and nightclub. Evaluation method: 10-fold cross validation.

Table 1
Performance prediction based on different feature sets. Predicted classes: restaurant, clothing store, bar, hotel, food and drink shop, cafe, gym, coffee shop, college and university, art gallery, and nightclub.

XGB Classifier					
Amsterdam Data					
Feature Set	Accuracy	Precision	Recall	F1-score	Data Sources Used
Operation-based	0.645	0.590	0.540	0.540	Google & FSQ POIs
Review-based	0.624	0.532	0.456	0.464	Google-FSQ POIs & G. Rev.
Neighborhood-based	0.389	0.383	0.234	0.244	Foursquare POIs
Topic-based	0.364	0.278	0.199	0.194	Tweets
Visual	0.326	0.245	0.156	0.145	Street Level Imagery
All	0.728	0.631	0.599	0.606	All
Athens Data					
Operation-based	0.528	0.556	0.458	0.481	Google & Foursquare POIs
Review-based	0.455	0.415	0.359	0.365	Google-FSQ POIs & G. Rev.
Neighborhood-based	0.436	0.480	0.350	0.364	Foursquare POIs
Topic-based	0.263	0.171	0.150	0.143	Tweets
Visual	0.251	0.186	0.163	0.154	Street Level Imagery
All	0.641	0.651	0.588	0.609	All

Table 2
Prediction performance based on different features for Amsterdam and Athens. Predicted classes: restaurant, clothing store, bar, hotel, food and drink shop, cafe, gym, coffee shop, college and university, art gallery, and nightclub.

XGB classifier - Classes set: A (cities comparison)				
Feature set	F1-score - AMS	F1-score - ATH	Difference	Best Case
Operation-based	0.540	0.481	0.049	Amsterdam
Review-based	0.464	0.365	0.099	Amsterdam
Neighborhood-based	0.244	0.364	0.120	Athens
Topic-based	0.194	0.143	0.051	Amsterdam
Visual	0.141	0.154	0.013	Athens
All	0.606	0.609	0.003	Athens

Athens in three cases: when trained on (1) the neighborhood-based features, (2) the visual features and (3) the combination of all features together. For the rest of the feature sets the F1-score is higher for Amsterdam. The largest difference between the two cities is found for the neighborhood-based feature set (+0.12 for Athens). The second and third largest differences relate to the review-based (+0.099 for Amsterdam) and the operation-based (+0.049 for Amsterdam) feature sets. When all features are used the difference between the two cities is very low (+0.003 for Athens). Thus, the relatively low differences, found in the performance of the classifier between the two cities, are close to zero when the feature sets are combined. Overall, the obtained results suggest that the proposed approach worked equally well for the two cities.

In order to gain further insight on the distinguishability of each POI category, we computed the confusion matrices when using all the feature sets, for both Amsterdam and Athens (Figs. 6 and 7). In both cities, the three most accurately predicted classes are the same, namely clothing store, hotel, and restaurant. More specifically, regarding Amsterdam, clothing store is the best-predicted label with 91% of instances being correctly classified, whereas cafe is the worst predicted label, with only 10% of them being correctly classified. Most cafe-related POIs are, in fact, wrongly classified as either “bars” (31%) or “restaurants” (32%). The second most mis-classified label is college and university with 28% of correct predictions. College and university instances tend to be classified mostly as “art gallery” (16%) or “hotel” (19%). In the case of Athens, the best predicted label is hotel, for which 86% of the POI instances are correctly classified. The worst predicted label is coffee shop for which the 29% are correctly classified. Most of the mis-classified coffee shops are classified as either “cafe” (29%) or “food and drink shop” (19%).

5.1.2. *Individual feature contribution*

The individual feature contribution is calculated according to the “Gain” of the XGBoost algorithm[8]. In principle, each split of the decision tree corresponds to a feature. Adding a new feature leads to the addition of new splits over the tree. “Gain” measures the improvement of the overall performance when those splits are in use or, in other words, when a feature is used compared to when it is not used. This individual “Gain” of each feature/split represents its relative contribution. The 15 most important features, in this regard, for Amsterdam and Athens are depicted in Fig. 8. The features in the form *“Topic (m/n): specific topic”* express the topics extracted from the English Google Reviews using the

[8] https://xgboost.readthedocs.io/en/latest/tutorials/model.html.

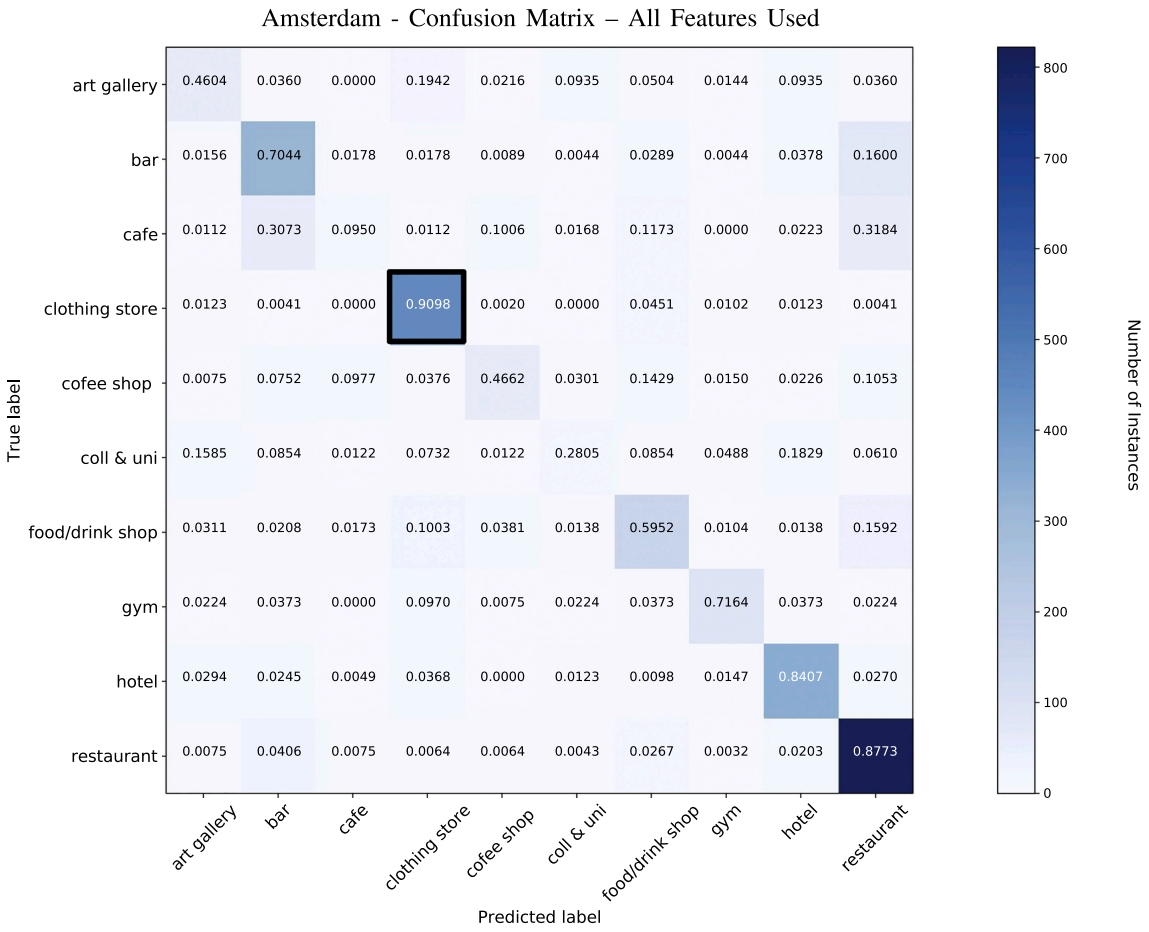

Fig. 6. Confusion Matrix for Amsterdam. The different colors represent the absolute number of predicted instances. Accuracy is denoted by numerical values.

LDA method, where n is the total number of the extracted topics, m is the order of the specified extracted topic, and "*specific topic*" is the labeling of the topic according to the keywords associated with it.

The feature contribution scores support the results shown in Table 1 for both cities. The majority of the 15 most contributing features indeed belong to the *operation-based*, *review-based*, and *neighborhood-based* feature sets. For Athens, two *neighborhood-based* features are included, whereas for Amsterdam only one, and this comes also in agreement with the previous results, as the *neighborhood-based* feature set proved to be more important for Athens than for Amsterdam. In addition, features which are categoryspecific, such as the extracted topics which are interpreted as Hotel, rank relatively high. Thus, it is indicated that if a feature is able to represent accurately just one of the POI categories, its importance will also be high.

5.2. Binary classification

The multi-class classification reveals the contribution of the POI features to predicting POI categories. However, these features might not be the same when dealing with a binary problem, such as in predicting whether a POI category is clothing store or not. In this case, it is possible that other features better capture the characteristics of this category and, thereby, contribute more to the classification. For instance, one could argue that clothing stores cluster spatially more often than, for example, hotels do. Although this particular feature might not be among the most contributing ones in the multi-class approach, it could nevertheless rank high when predicting whether a POI belongs to a specific category (e.g. clothing store) or not. This section explores this issue in further detail.

The most correctly predicted POI categories according to the confusion matrices in Figs. 6 and 7, for both cities, are hotel, clothing store, and restaurant. Given that the extracted features appear to be able to better capture the characteristics of these three POI categories, this section focuses on these specific categories. Thus, in the following paragraphs, three cases are being studied: predicting if the category of a POI is (1) a hotel, (2) a clothing store, or (3) a restaurant.

The first step is to balance the dataset since our focus is on identifying the descriptive characteristics of the selected POI categories. A representative dataset would be highly unbalanced, given that the number of instances of a single POI category is very low, compared to all the rest of the categories combined. Thus, even if a representative dataset could improve the overall performance of the classifier, it would not serve our goal which is to assess the influence of different POI features on their classification into categories. In balancing the dataset, the following process is followed: let p be the number of instances of the category to be predicted and n the total number of types. Then from each POI category, a random sample of instances equal to $\frac{p}{n-1}$ is retrieved and used, so that their sum is equal to the number of the instances of the POI category to be predicted. Thus, the dataset is balanced and consists of two classes: one representing the category to be predicted, and another one representing all the rest. For the latter, the instances are equally distributed among the nine remaining categories and combined they lead to a number of instances equal to the one of the predicted class.

5.2.1. Feature set contribution

Table 3 presents the F1-scores of the XGB classifier, when trained on the different feature sets, for each binary classification problem (hotel, clothing store and restaurant) and for both cities. As previously stated, it is important to notice that for each POI category and for both cities the

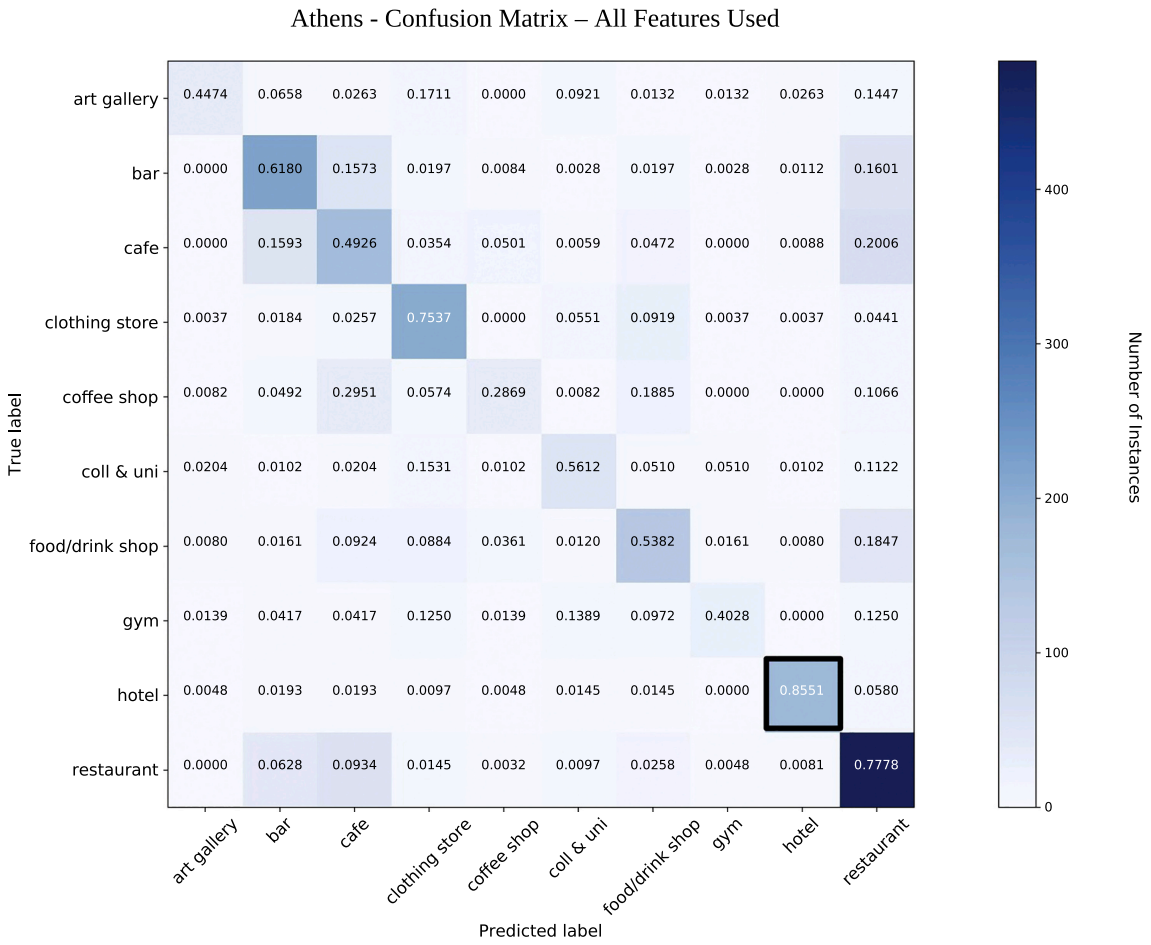

Fig. 7. Confusion Matrix for Athens. The different colors represent the absolute number of predicted instances. Accuracy is denoted by numerical values.

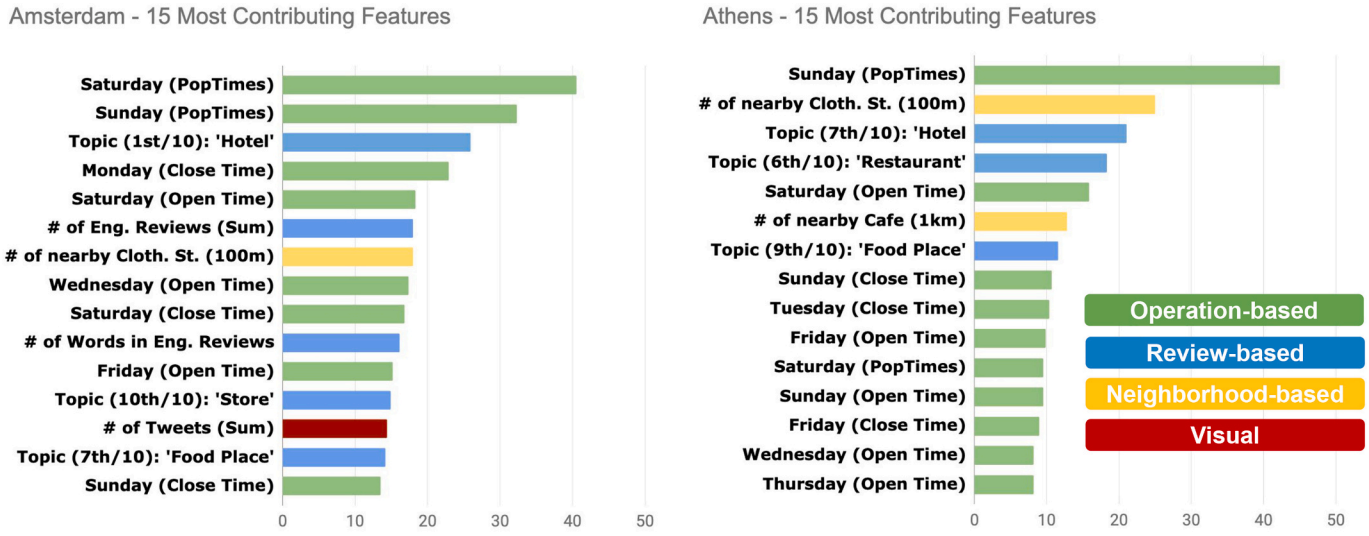

Fig. 8. Top 15 most contributing features -Amsterdam (left) & Athens (right). The "Topic" features have been extracted using the LDA method. Thus, Review Topic (1) expresses the first extracted topic from the English Google reviews.

combination of all feature sets leads to the highest F1-scores. However, the performance of the classifier when trained on different feature sets is not always stable among the three POI categories under study. An example of this is presented for the case of Amsterdam, regarding the topic-based feature set. For clothing store the F1-score is 77% while for restaurant it is 58%. This implies that different POI categories are "special" for different reasons.

In the case of Amsterdam, the F1-score is higher when predicting whether a POI represents a clothing store (92%) than a hotel (89%) or a restaurant (0.88%). Again, the most contributing feature sets seem to be

Table 3
Prediction performance based on different feature sets. The columns Hotel, Clothing store, and Restaurant represent the three binary classification problems that are analyzed.

XGB classifier - Classes set: A				
Amsterdam data				
Feature Set	Hotel F1-score	Cl. Store F1-score	Restaurant F1-score	Data sources used
Operation-based	0.832	0.857	0.823	Google & FSQ POIs
Review-based	0.889	0.784	0.853	Google-FSQ POIs & G. Rev.
Neighborhood-based	0.633	0.783	0.600	Foursquare POIs
Topic-based	0.610	0.774	0.579	Tweets
Visual	0.550	0.650	0.558	Street Level Imagery
All	0.892	0.924	0.877	All
Athens Data				
Operation-based	0.790	0.776	0.775	Google & Foursquare POIs
Review-based	0.881	0.661	0.774	Google-FSQ POIs & G. Rev.
Neighborhood-based	0.818	0.781	0.580	Foursquare POIs
Topic-based	0.65	0.6	0.558	Tweets
Visual	0.550	0.601	0.567	Street Level Imagery
All	0.920	0.820	0.831	All

the operation-based, review-based and neighborhood-based while the topic-based and visual tend to be less important. However, for each POI category the results are quite different and, therefore, general statements are hard to make.

In the case of Athens, the best results are obtained for the hotel (92%) followed by the restaurant (83%) and then by the clothing store (82%). It is worth mentioning that in the cases of hotel and clothing store the *neighborhood-based* feature set leads to better results than the operation-based or review-based feature sets. On the contrary, for restaurants, the neighborhood-based feature set does not work that well.

Overall, the results obtained when the classifier is trained on the *operation-based* and *review-based* feature sets are quite consistent for both cities, and always relatively high. In contrast, the topic-based and visual feature sets both tend to not play an important role in almost every case. Finally, the contribution of the neighborhood-based feature set to predicting the POI category is in some cases quite high and in others not.

5.2.2. Individual feature contribution

Diving into the contribution of the individual features, the 15 most contributing features regarding the hotel and clothing store categories (restaurants had similar results), are presented in Fig. 9. In general terms, the results presented in Table 3 agree with the results of those figures, meaning that the most contributing features indeed belong to the feature sets that lead to the highest performance of the classifier. Particularly, the majority of the 15 most contributing features in both cases are either *operation-based* or *review-based*. This consistency, however, is not observed in all cases.

Regarding hotel POIs, the most contributing features tend to be *operation-based* (e.g. opening/closing hours, visiting hours) and *review-based* (e.g. features extracted from reviews) for both cities. Not surprisingly, the extracted topics from the Google reviews relating to hotels, are present in those figures among the three most contributing features. Other topics which rank high are the Gym (for Amsterdam) and the Food Place for Athens, both representing amenities that could be offered by hotels. In addition, some of the neighborhood-based features are also among the 15 most contributing features. These are "nearby hotels" and "nearby Art Galleries" for Amsterdam and Athens, respectively. In the

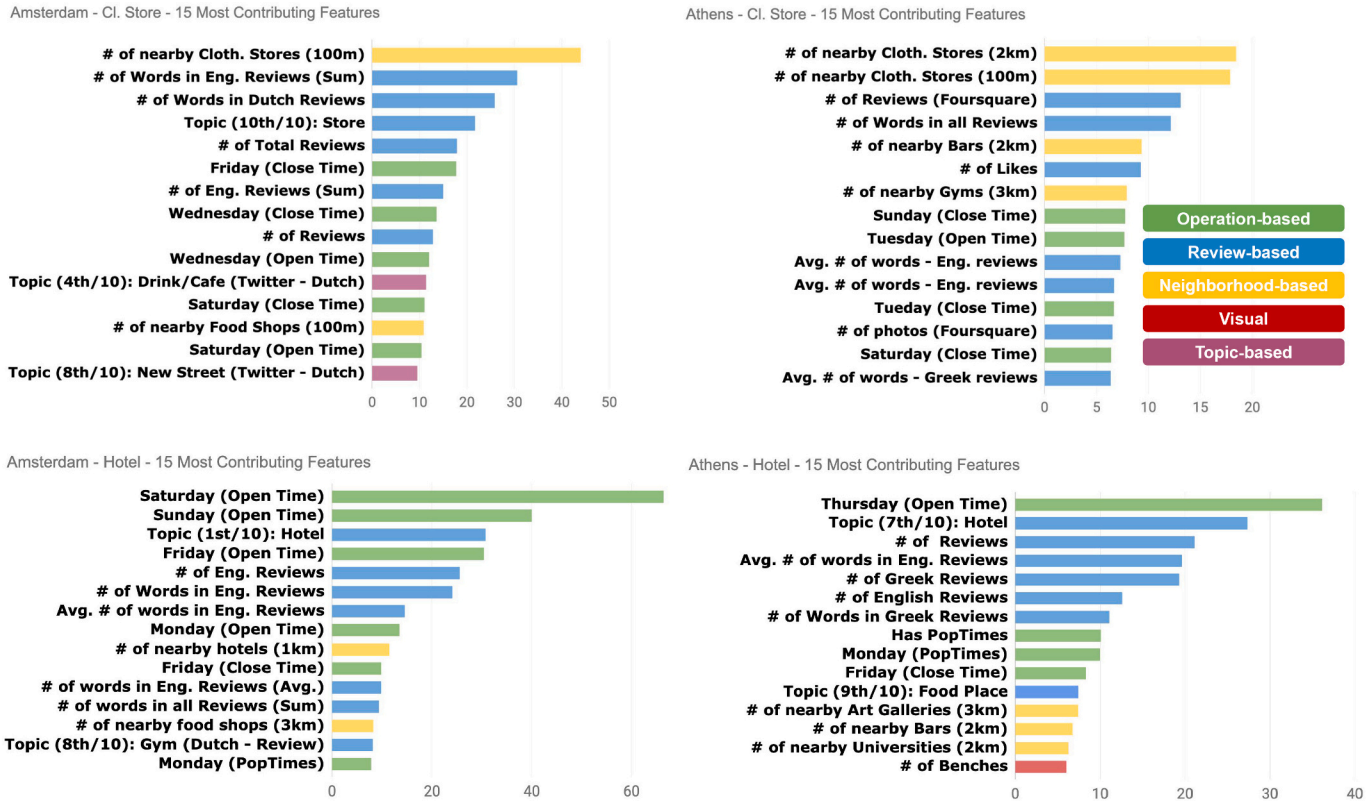

Fig. 9. The 15 most contributing features for predicting if a POI is a Hotel or a Clothing Store in Amsterdam (left) and Athens (right). The "Topic" features refer to the extracted topics from the Google Reviews.

case of Athens, there is also a visual feature among the 15 most contributing features: the number of benches. Thus, even if the use of a feature set does not lead to good results, an individual feature belonging to this set could still be important. Overall, the features which mostly characterize a hotel seem to be the opening/closing hours, the number, size and topics of the written reviews and their location with respect to other places of the same and different category.

Regarding clothing store POIs, the "number of nearby clothing stores" is the most contributing feature for both Amsterdam and Athens. The neighborhood - based features are more contributing for Athens than for Amsterdam, whereas for the topic-based features the opposite occurs. This comes in agreement with the results presented in Table 3, where the topic-based feature set lead to comparable performance with the review-based and operation-based (e.g. the topic "Brand" which is extracted from Twitter is the 11th most contributing feature for Amsterdam) in this case. Overall, the features which seem to better characterize clothing stores are the opening and closing hours, the number, size and topic of the written reviews, and their location especially in respect to other clothing stores.

The above results suggest that, depending on the POI category to be predicted, the features which perform the best as predictors of the category differ. However, even if the most contributing individual features vary, in every case the *operation-based* and *review-based* features are consistently among the most contributing ones. In both use cases, the *combination* of the features improves the results of the classification (by an average of 19%). The results support that while the contribution of each feature and feature set varies, the overall ranking of those features is similar.

6. Discussion

Given that in all our experiments the performance of the classifier fluctuates quite consistently when trained on different feature sets, this Section discusses in further detail the influence of each feature set on both classification problems (i.e. multi-class and binary).

Starting with the *operation-based* features, the results support that they contribute the most when categorizing unlabeled POIs. These features proved to rank high and, correspondingly, the performance of our classifier declined in the experiments where they were not available. Our results particularly suggest that the time-related operation-based characteristics of a POI contribute the most to describing its category.

The *review-based* feature set also contributes substantially to the classification of POI categories. The review-based features consist mostly of the number, length, and topics of Google reviews per POI. Given that the reviews are category-specific, they tend to include valuable information about the POI category itself. For instance, it is reasonable that the clothing store reviews include different characteristics/topics from the restaurant reviews as they focus on different characteristics of the places they refer to (e.g. in the case of clothing stores, reviews focus on products and prices, whereas in restaurants they focus on food). The high ranking of the review-based features suggests that short user-generated texts from reviews, contribute substantially to identifying the category of a POI.

As observed in both use cases, the *topic-based* and *visual* features contribute the least to the categorization of POIs. However, topic-based features could contribute more to the further characterization of POIs that belong to the same category (e.g. to characterize bars as "artistic" or "sport"), or to the identification of neighborhood characteristics, as in (Fu et al., 2018). Nevertheless, in some cases specific topic-based features, such as the number of tweets in the multi-class classification problem (Fig. 8) or the extracted topics in the binary problem (Fig. 9), proved to rank high, in terms of how much they influence the classification. Another notable aspect, is that the used natural language processing libraries tend to work better in the English language, meaning that the influence of the topic-based features in our experiments might have been affected by the percentage of the English written tweets over the total number of tweets.

Regarding the *visual* features, their relatively low influence could be explained by the fact that the selected POI categories, tend to be quite similar in visual terms. For instance, a bar, cafe, coffee shop, or restaurant are often hard to distinguish solely by their storefronts (if labels and logos are excluded). The existence of other objects (e.g. trees etc.) in the exterior space of each place did not seem to correlate with its category. However, it is important to note that, in this work, we did not consider the signage on facades, which is often found in many business-related storefronts (Sharifi Noorian, Qiu, Psyllidis, Bozzon, & Houben, 2020). Moreover, we only considered exterior but not interior pictures, which could be promising predictors of POI categories.

Lastly, the largest difference in our experiments when predicting the various POI categories, is found when using the *neighborhood-based* feature set. This difference is potentially affected by the tendency of certain amenities (e.g. food businesses and retail stores) to spatially concentrate, due to endogenous or exogenous externalities (Koster et al., 2019; Sevtsuk, 2014). However, it is worth mentioning that the spatial clustering of places appears to be predominantly dependent on the category of each POI rather than on the region or the city. For instance, the clustering of clothing stores across space appears to be irrespective of the city at hand. This led to the fluctuation of the neighborhood-based features' influence throughout our experiments, and it is an important factor to consider when categorizing POIs. Further research on regional variations is, however, needed to further support this, by including POIs in cities belonging to different continents.

7. Conclusion

In this paper, we presented a study of the influence that different POI features have on the classification of unlabeled POIs into categories. We defined a range of feature sets, which cover operation-based, review-based, topic-based, neighborhood-based, and visual attributes of places, and further assessed and ranked their influence — both in sets and individually — using multi-class and binary classification approaches. As expected, our results suggest that the extracted features do not contribute equally to the classification of POIs into categories. More specifically, the features which have a larger influence on POI categorization are the ones relating to the *operation-based* (e.g. opening/closing hours) and *review-based* (e.g. topics extracted from the reviews) attributes of a POI. The similarity of the obtained results in all our experiments, and in both cities under study, could indicate that the contribution of these features is not affected significantly by either the local context or the selection of specific POI categories. However, further analysis needs to be conducted — in which cities from different continents, with substantially different urban spatial configurations (e.g. Asian or American cities) will be included — to further ground this indication and provide additional support to this statement. The influence of urban spatial configuration was more evident in other feature sets, such as the neighborhood-based ones. These tend to rank higher, in terms of contribution to POI categorization, in cases where POIs cluster spatially by category (e.g. "bar" areas). To improve the scalability of our approach and to further support our results, we plan to extend the scope of our experiments to other cities, and also include interior pictures of POIs, where available.

Competing interests statement

The authors declare no potential conflicts of interest of any sort.

Acknowledgements

This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 874724.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Choi, S. J., Song, H. J., Park, S. B., & Lee, S. J. (2020). A poi categorization by composition of onomastic and contextual information. In *, 2. 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (pp. 38–45). IEEE.

Cocos, A., & Callison-Burch, C. (2020). The language of place: Semantic value from geospatial context. In *, 2. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2* (pp. 99–104). Short Papers.

Ding, D., Zhang, M., Pan, X., Wu, D., & Pu, P. (2020). Geographical feature extraction for entities in location-based social networks. In *Proceedings of the 2018 World Wide Web Conference* (pp. 833–842).

Fu, C., McKenzie, G., Frias-Martinez, V., & Stewart, K. (2018). *Identifying spatiotemporal urban activities through linguistic signatures*. Computers, Environment and Urban Systems.

Fu, K., Chen, Z., & Lu, C.-T. (2020). Streetnet: preference learning with convolutional neural network on urban crime perception. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 269–278).

Funke, S., & Storandt, S. (2020). Automatic tag enrichment for points-of-interest in open street map. In *International Symposium on Web and Wireless Geographical Information Systems* (pp. 3–18). Springer.

Giannopoulos, G., Alexis, K., Kostagiolas, N., & Skoutas, D. (2020). Classifying points of interest with minimum metadata. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations* (pp. 1–4). Geosocial Networks and Geoadvertising.

Giannopoulos, G., & Meimaris, M. (2020). Learning domain driven and semantically enriched embeddings for poi classification. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases* (pp. 214–217).

Janowicz, K., McKenzie, G., Hu, Y., Zhu, R., & Gao, S. (2019). Using semantic signatures for social sensing in urban environments. In *Mobility Patterns, Big Data and Transport Analytics* (pp. 31–54). Elsevier.

Jenkins, A., Croitoru, A., Crooks, A. T., & Stefanidis, A. (2016). Crowdsourcing a collective sense of place. *PLoS One, 11*.

Jiang, S., Alves, A., Rodrigues, F., Ferreira, J., Jr., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems, 53*, 36–46.

Koster, H. R., Pasidis, I., & van Ommeren, J. (2019). Shopping externalities and retail concentration: Evidence from dutch shopping streets. *Journal of Urban Economics, 114*, 103194.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing streetlevel urban greenery using google street view and a modified green view index. *Urban Forestry & Urban Greening, 14*, 675–685.

Liu, L., Silva, E. A., Wu, C., & Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems, 65*, 113–125.

Lu, X., Yu, Z., Liu, C., Liu, Y., Xiong, H., & Guo, B. (2020). Inferring lifetime status of point-of-interest: A multitask multiclass approach. *ACM Transactions on Knowledge Discovery from Data (TKDD), 14*, 1–27.

Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems, 74*, 161–174.

McKenzie, G., & Adams, B. (2018). A data-driven approach to exploring similarities of tourist attractions through online reviews. *Journal of Location Based Services, 12*, 94–118.

McKenzie, G., Janowicz, K., & Adams, B. (2020). Weighted multi-attribute matching of user-generated points of interest. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 440–443). ACM.

McKenzie, G., Janowicz, K., Gao, S., & Gong, L. (2015). How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. *Computers, Environment and Urban Systems, 54*, 336–346.

Psyllidis, A., Yang, J., & Bozzon, A. (2018). Regionalization of social interactions and points-of-interest location prediction with geosocial data. *IEEE Access, 6*, 34334–34353.

Qiu, S., Psyllidis, A., Bozzon, A., & Houben, G.-J. (2020). Crowd-mapping urban objects from street-level imagery. In *The World Wide Web Conference* (pp. 1521–1531).

Sevtsuk, A. (2014). Location and agglomeration: The distribution of retail and food businesses in dense urban environments. *Journal of Planning Education and Research, 34*, 374–393.

Sharifi Noorian, S., Qiu, S., Psyllidis, A., Bozzon, A., & Houben, G.-J. (2020). Detecting, classifying, and mapping retail storefronts using street-level imagery. In *Proceedings of the 2020 International Conference on Multimedia Retrieval* (pp. 495–501).

Vandecasteele, A., & Devillers, R. (2015). Improving volunteered geographic information quality using a tag recommender system: the case of openstreetmap. In *OpenStreetMap in GIScience* (pp. 59–80). Springer.

Xu, C., Li, J., Luo, X., Pei, J., Li, C., & Ji, D. (2020). Dlocrl: A deep learning pipeline for fine-grained location recognition and linking in tweets. In *The World Wide Web Conference* (pp. 3391–3397).

Yan, B., Janowicz, K., Mai, G., & Gao, S. (2020). From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 1–10).

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2017). Sensing spatial distribution of urban land use by integrating points-of-interest and google word2vec model. *International Journal of Geographical Information Science, 31*, 825–848.

Zhang, C., Zhang, K., Yuan, Q., Peng, H., Zheng, Y., Hanratty, T., … Han, J. (2020). Regions, periods, activities: Uncovering urban dynamics via cross-modal representation learning. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 361–370).

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning, 180*, 148–160.

Zhang, Y., Xiong, Y., Kong, X., & Zhu, Y. (2020). Learning node embeddings in interaction graphs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 397–406).

Zhao, S., Zhao, T., King, I., & Lyu, M. R. (2020). Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 153–162).

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(6), 1452–1464.

Zhou, J., Gou, S., Hu, R., Zhang, D., Xu, J., Jiang, A., … Xiong, H. (2020). A collaborative learning framework to tag refinement for points of interest. In *Proceedings of the 25th ACM SIGKDD international conference on Knowledge Discovery & Data Mining* (pp. 1752–1761).

Ziegler, K., Caelen, O., Garchery, M., Granitzer, M., He-Guelton, L., Jurgovsky, J., … Zwicklbauer, S. (2020). Injecting semantic background knowledge into neural networks using graph embeddings. In *2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)* (pp. 200–205). IEEE.