

Embedding Geographic Locations for Modelling the Natural Environment using Flickr Tags and Structured Data

Shelan S. Jeawak, Christopher B. Jones, and Steven Schockaert

Cardiff University, School of Computer Science and Informatics, Cardiff, UK
{JeawakSS, JonesCB2, SchockaertS1}@cardiff.ac.uk

Abstract. Meta-data from photo-sharing websites such as Flickr can be used to obtain rich bag-of-words descriptions of geographic locations, which have proven valuable, among others, for modelling and predicting ecological features. One important insight from previous work is that the descriptions obtained from Flickr tend to be complementary to the structured information that is available from traditional scientific resources. To better integrate these two diverse sources of information, in this paper we consider a method for learning vector space embeddings of geographic locations. We show experimentally that this method improves on existing approaches, especially in cases where structured information is available.

Keywords: Social media · Text mining · Vector space embeddings · Volunteered Geographic Information · Biodiversity · Ecology.

1 Introduction

Users of photo-sharing websites such as Flickr¹ often provide short textual descriptions in the form of tags to help others find the images. With the availability of GPS systems in current electronic devices such as smartphones, latitude and longitude coordinates are nowadays commonly made available as well. The tags associated with such georeferenced photos often describe the location where these photos were taken, and Flickr can thus be regarded as a source of environmental information. The use of Flickr for modelling urban environments has already received considerable attention. For instance, various approaches have been proposed for modelling urban regions [5], and for identifying points-of-interest [45] and itineraries [7,36]. However, the usefulness of Flickr for characterizing the natural environment, which is the focus of this paper, is less well-understood.

Many recent studies have highlighted that Flickr tags capture valuable ecological information, which can be used as a complementary source to more traditional sources. To date, however, ecologists have mostly used social media to conduct manual evaluations of image content with little automated exploitation of the associated tags [39,6,10]. One recent exception is [21], where bag-of-words

¹ <http://www.flickr.com>

representations derived from Flickr tags were found to give promising result for predicting a range of different environmental phenomena.

Our main hypothesis in this paper is that by using vector space embeddings instead of bag-of-words representations, the ecological information which is implicitly captured by Flickr tags can be utilized in a more effective way. Vector space embeddings are representations in which the objects from a given domain are encoded using relatively low-dimensional vectors. They have proven useful in natural language processing, especially for encoding word meaning [30,34], and in machine learning more generally. In this paper, we are interested in the use of such representations for modelling geographic locations. Our main motivation for using vector space embeddings is that they allow us to integrate the textual information we get from Flickr with available structured information in a very natural way. To this end, we rely on an adaptation of the GloVe word embedding model [34], but rather than learning word vectors, we learn vectors representing locations. Similar to how the representation of a word in GloVe is determined by the context words surrounding it, the representation of a location in our model is determined by the tags of the photos that have been taken near that location. To incorporate numerical features from structured environmental datasets (e.g. average temperature), we associate with each such feature a linear mapping that can be used to predict that feature from a given location vector. This is inspired by the fact that salient properties of a given domain can often be modelled as directions in vector space embeddings [17,8,40]. Finally, evidence from categorical datasets (e.g. land cover types) is taken into account by requiring that locations belonging to the same category are represented using similar vectors, similar to how semantic types are sometimes modelled in the context of knowledge graph embedding [16].

While our point-of-departure is a standard word embedding model, we found that the off-the-shelf GloVe model performed surprisingly poorly, meaning that a number of modifications are needed to achieve good results. Our main findings are as follows. First, given that the number of tags associated with a given location can be quite small, it is important to apply some kind of spatial smoothing, i.e. the importance of a given tag for a given location should not only depend on the occurrences of the tag at that location, but also on its occurrences at nearby locations. To this end, we use a formulation which is based on spatially smoothed version of pointwise mutual information. Second, given the wide diversity in the kind of information that is covered by Flickr tags, we find that term selection is in some cases critical to obtain vector spaces that capture the relevant aspects of geographic locations. For instance, many tags on Flickr refer to photography related terms, which we would normally not want to affect the vector representation of a given location². Finally, even with these modifications, vector space embeddings learned from Flickr tags alone are sometimes outperformed by bag-of-words representations. However, our vector space embeddings

² One exception is perhaps when we want to predict the scenicness of a given location, where e.g. terms that are related to professional landscape photography might be a strong indicator of scenicness.

lead to substantially better predictions in cases where structured (scientific) information is also taken into account. In this sense, the main value of using vector space embeddings in this context is not so much about abstracting away from specific tag usages, but rather about the fact that such representations allow us to integrate numerical and categorical features in a much more natural way than is possible with bag-of-words representations.

The remainder of this paper is organized as follows. In the next section, we provide a discussion of existing work. Section 3 then presents our model for embedding geographic locations from Flickr tags and structured data. Next, in Section 4 we provide a detailed discussion about the experimental results. Finally, Section 5 summarizes our conclusions.

2 Related Work

2.1 Vector space embeddings

The use of low-dimensional vector space embeddings for representing objects has already proven effective in a large number of applications, including natural language processing (NLP), image processing, and pattern recognition. In the context of NLP, the most prominent example is that of word embeddings, which represent word meaning using vectors of typically around 300 dimensions. A large number of different methods for learning such word embeddings have already been proposed, including Skip-gram and the Continuous Bag-of-Words (CBOW) model [30], GloVe [34], and fastText [13]. They have been applied effectively in many downstream NLP tasks such as sentiment analysis [43], part of speech tagging [35,28], and text classification [26,12]. The model we consider in this paper builds on GloVe, which was designed to capture linear regularities of word-word co-occurrence. In GloVe, there are two word vectors w_i and \tilde{w}_j for each word in the vocabulary, which are learned by minimizing the following objective:

$$J = \sum_{i,j=1}^V f(x_{ij})(w_i \cdot \tilde{w}_j + b_i + \tilde{b}_j - \log x_{ij})^2$$

where x_{ij} is the number of times that word i appears in the context of word j , V is the vocabulary size, b_i is the target word bias, \tilde{b}_j is the context word bias. The weighting function f is used to limit the impact of rare terms. It is defined as 1 if $x > x_{max}$ and as $(\frac{x}{x_{max}})^\alpha$ otherwise, where x_{max} is usually fixed to 100 and α to 0.75. Intuitively, the target word vectors w_i correspond to the actual word representations which we would like to find, while the context word vectors \tilde{w}_j model how occurrences of j in the context of a given word i affect the representation of this latter word. In this paper we will use a similar model, which will however be aimed at learning location vectors instead of the target word vectors.

Beyond word embeddings, various methods have been proposed for learning vector space representations from structured data such as knowledge graphs

[2,52,44], social networks [15,48] and taxonomies [47,31]. The idea of combining a word embedding model with structured information has also been explored by several authors, for example to improve the word embeddings based on information coming from knowledge graphs [50,42]. Along similar lines, various lexicons have been used to obtain word embeddings that are better suited at modelling sentiment [43] and antonymy [33], among others. The method proposed by [27] imposes the condition that words that belong to the same semantic category are closer together than words from different categories, which is somewhat similar in spirit to how we will model categorical datasets in our model.

2.2 Embeddings for geographic information

The problem of representing geographic locations using embeddings has also attracted some attention. An early example is [41], which used principal component analysis and stacked autoencoders to learn low-dimensional vector representations of city neighbourhoods based on census data. They use these representations to predict attributes such as crime, which is not included in the given census data, and find that in most of the considered evaluation tasks, the low-dimensional vector representations lead to more faithful predictions than the original high-dimensional census data.

Some existing works combine word embedding models with geographic coordinates. For example, in [4] an approach is proposed to learn word embeddings based on the assumption that words which tend to be used in the same geographic locations are likely to be similar. Note that their aim is dual to our aim in this paper: while they use geographic location to learn word vectors, we use textual descriptions to learn vectors representing geographic locations.

Several methods also use word embedding models to learn representations of Points-of-Interest (POIs) that can be used for predicting user visits [11,29,55]. These works use the machinery of existing word embedding models to learn POI representations, intuitively by letting sequences of POI visits by a user play the role of sequences of words in a sentence. In other words, despite the use of word embedding models, many of these approaches do not actually consider any textual information. For example, in [29] the Skip-gram model is utilized to create a global pattern of users' POIs. Each location was treated as a word and the other locations visited before or after were treated as context words. They then use a pair-wise ranking loss [49] which takes into account the user's location visit frequency to personalize the location recommendations. The methods of [29] were extended in [55] to use a temporal embedding and to take more account of geographic context, in particular the distances between preferred and non-preferred neighboring POIs, to create a "geographically hierarchical pairwise preference ranking model". Similarly, in [53] the CBOW model was trained with POI data. They ordered POIs spatially within the traffic-based zones of urban areas. The ordering was used to generate characteristic vectors of POI types. Zone vectors represented by averaging the vectors of the POIs contained in them, were then used as features to predict land use types. In the CrossMap method [54] they learned embeddings for spatio-temporal hotspots obtained from social media

data of locations, times and text. In one form of embedding, intended to enable reconstruction of records, neighbourhood relations in space and time were encoded by averaging hotspots in a target location’s spatial and temporal neighborhoods. They also proposed a graph-based embedding method with nodes of location, time and text. The concatenation of the location, time and text vectors were then used as features to predict peoples’ activities in urban environments. Finally, in [51], a method is proposed that uses the Skip-gram model to represent POI types, based on the intuition that the vector representing a given POI type should be predictive of the POI types that found near places of that type.

Our work is different from these studies, as our focus is on representing locations based on a given text description of that location (in the form of Flickr tags), along with numerical and categorical features from scientific datasets.

2.3 Analyzing Flickr tags

Many studies have focused on analyzing Flickr tags to extract useful information in domains such as linguistics [9], geography [5,14], and ecology [1,21,22]. Most closely related to our work, [21] found that the tags of georeferenced Flickr photos can effectively supplement traditional scientific environmental data in tasks such as predicting climate features, land cover, species occurrence, and human assessments of scenicness. To encode locations, they simply combine a bag-of-words representation of geographically nearby tags with a feature vector that encodes associated structured scientific data. They found that the predictive value of Flickr tags is roughly on a par with that of the scientific datasets, and that combining both types of information leads to significantly better results than using either of them alone. As we show in this paper, however, their straightforward way of combining both information sources, by concatenating the two types of feature vectors, is far from optimal.

Despite the proven importance of Flickr tags, the problem of embedding Flickr tags has so far received very limited attention. To the best of our knowledge, [18] is the only work that generated embeddings for Flickr tags. However, their focus was on learning embeddings that capture word meaning (being evaluated on word similarity tasks), whereas we use such embeddings as part of our method for representing locations.

3 Model Description

In this section, we introduce our embedding model, which combines Flickr tags and structured scientific information to represent a set of locations L . The proposed model has the following form:

$$J = \alpha J_{tags} + (1 - \alpha) J_{nf} + \beta J_{cat} \quad (1)$$

where $\alpha \in [0, 1]$ and $\beta \in [0, +\infty]$ are parameters to control the importance of each component in the model. Component J_{tags} will be used to constrain the

representation of the locations based on their textual description (i.e. Flickr tags), J_{nf} will be used to constrain the representation of the locations based on their numerical features, and J_{cat} will impose the constraint that locations belonging to the same category should be close together in the space. We will discuss each of these components in more detail in the following sections.

3.1 Tag Based Location Embedding

Many of the tags associated with Flickr photos describe characteristics of the places where these photos were taken [19,37,38]. For example, tags may correspond to place names (e.g. Brussels, England, Scandinavia), landmarks (e.g. Eiffel Tower, Empire State Building) or land cover types (e.g. mountain, forest, beach). To allow us to build location models using such tags, we collected the tags and meta-data of 70 million Flickr photos with coordinates in Europe (which is the region our experiments will focus on), all of which were uploaded to Flickr before the end of September 2015. In this section we first explain how tags can be weighted to obtain bag-of-words representations of locations from Flickr. Subsequently we describe a tag selection method, which will allow us to specialize the embedding depending on which aspects of the considered locations are of interest, after which we discuss the actual embedding model.

Tag weighting. Let $L = \{l_1, \dots, l_m\}$ be a set of geographic locations, each characterized by latitude and longitude coordinates. To generate a bag-of-words representation of a given location, we have to weight the relevance of each tag to that location. To this end, we have followed the weighting scheme from [21], which combines a Gaussian kernel (to model spatial proximity) with Positive Pointwise Mutual Information (PPMI) [3,32].

Let us write $U_{t,l}$ for the set of users who have assigned tag t to a photo with coordinates near l . To assess how relevant t is to the location l , the number of times t occurs in photos near l is clearly an important criterion. However, rather than simply counting the number of occurrences within some fixed radius, we use a Gaussian kernel to weight the tag occurrences according to their distance from that location:

$$w(t, l) = \sum_{d(l,r) \leq D} |U_{t,l}| \cdot \exp\left(-\frac{d^2(l,r)}{2\sigma^2}\right)$$

where the threshold $D > 0$ is assumed to be fixed, r is the location of a Flickr photo, d is the Haversine distance, and we will assume that the bandwidth parameter σ is set to $D/3$. A tag occurrence is counted only once for all photos by the same user at the same location, which is important to reduce the impact of bulk uploading. The value $w(t, l)$ reflects how frequent tag t is near location l , but it does not yet take into account the total number of tag occurrences near l , nor how popular the tag t is overall. To measure how strongly tag t is associated with location l , we use PPMI, which is a commonly used measure of association in natural language processing. However, rather than estimating PPMI scores

from term frequencies, we will use the $w(t, l)$ values instead:

$$PPMI(t, l) = \max \left(0, \log \left(\frac{p_{t,l}}{p_t p_l} \right) \right)$$

where:

$$p_{t,l} = \frac{w(t, l)}{N} \quad p_t = \frac{\sum_{l' \in L} w(t, l')}{N} \quad N = \sum_{t' \in T} \sum_{l' \in L} w(t', l') \quad p_l = \frac{\sum_{t' \in T} w(t', l)}{N}$$

with T the set of all tags, and L the set of locations.

Tag selection. Inspired by [25], we use a term selection method in order to focus on the tags that are most important for the tasks that we want to consider and reduce the impact of tags that might relate only to a given individual or a group of users. In particular, we obtained good results with a method based on Kullback-Leibler (KL) divergence, which is based on [46]. Let C_1, \dots, C_n be a set of (mutually exclusive) properties of locations in which we are interested (e.g. land cover categories). For the ease of presentation, we will identify C_i with the set of locations that have the corresponding property. Then, we select tags from T that maximize the following score:

$$KL(t) = \sum_{i=1}^n P(C_i|t) \log \frac{P(C_i|t)}{Q(C_i)}$$

where $P(C_i|t)$ is the probability that a photo with tag t has a location near C_i and $Q(C_i)$ is the probability that an arbitrary tag occurrence is assigned to a photo near a location in C_i . Since $P(C_i|t)$ often has to be estimated from a small number of tag occurrences, it is estimated using Bayesian smoothing:

$$P(C_i|t) = \frac{(\sum_{l \in C_i} w(t, l)) + \gamma \cdot Q(C_i)}{N + \gamma}$$

where γ is a parameter controlling the amount of smoothing, which will be tuned in the experiments. On the other hand, for $Q(C_i)$ we can simply use a maximum likelihood estimation:

$$Q(C_i) = \frac{\sum_{l \in C_i} \sum_{t \in T} w(t, l)}{\sum_{j=1}^n \sum_{l \in C_j} \sum_{t \in T} w(t, l)}$$

Location embedding. We now want to find a vector $v_{l_i} \in V$ for each location l_i such that similar locations are represented using similar vectors. To achieve this, we use a close variant of the GloVe model, where tag occurrences are treated as context words of geographic locations. In particular, with each location l we associate a vector v_l and with each tag t we associate a vector \tilde{w}_t and a bias term \tilde{b}_{t_j} , and consider the following objective (which in our full model (1) will be combined with components that are derived from the structured information):

$$J_{tags} = \sum_{l_i \in L} \sum_{t_j \in T} (v_{l_i} \tilde{w}_{t_j} + \tilde{b}_{t_j} - PPMI(t_j, l_i))^2$$

Note how tags play the role of the context words in the GloVe model, while instead of learning target word vectors we now learn location vectors. In contrast to GloVe, our objective does not directly refer to co-occurrence statistics, but instead uses the $PPMI(t_j, l_i)$ scores. One important consequence of this is that we can also consider pairs (l_i, t_j) for which t_j does not occur in l_i at all; such pairs are usually called *negative examples*. While they cannot be used in the standard GloVe model, some authors have already reported that introducing negative examples in variants of GloVe can lead to an improvement [20]. In practice, evaluating the full objective above would not be computationally feasible, as we may need to consider millions of locations and millions of tags. Therefore, rather than considering all tags in T for the inner summation, we only consider those tags that appear at least once near location l_i together with a sample of negative examples.

3.2 Structured Environmental Data

There is a wide variety of structured data that can be used to describe locations. In this work, we have restricted ourselves to the same datasets as [21]. These include nine (real-valued) numerical features, which are latitude, longitude, elevation³, population⁴, and five climate⁵ related features (avg. temperature, avg. precipitation, avg. solar radiation, avg. wind speed, and avg. water vapor pressure). In addition, 180 categorical features were used, which are CORINE⁶ land cover classes at level 1 (5 classes), level 2 (15 classes) and level 3 (44 classes) and 116 soil types (SoilGrids⁷). Note that each location should belong to exactly 4 categories: one CORINE class at each of the three levels and a soil type.

Numerical features. Numerical features can be treated similarly to the tag occurrences, i.e. we will assume that the value of a given numerical feature can be predicted from the location vectors using a linear mapping. In particular, for each numerical feature f_k we consider a vector \tilde{w}_{f_k} and a bias term \tilde{b}_{f_k} , and the following objective:

$$J_{nf} = \sum_{l_i \in L} \sum_{f_k \in NF} (v_{l_i} \cdot \tilde{w}_{f_k} + \tilde{b}_{f_k} - \text{score}(f_k, l_i))^2$$

where we write NF for set of all numerical features and $\text{score}(f_k, l_i)$ is the value of feature f_k for location l_i , after z-score normalization.

Categorical features. To take into account the categorical features, we impose the constraint that locations belonging to the same category should be close together in the space. To formalize this, we represent each category type cat_i as

³ <http://www.eea.europa.eu/data-and-maps/data/eu-dem>

⁴ <http://data.europa.eu/89h/jrc-luisa-europopmap06>

⁵ <http://worldclim.org>

⁶ <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-2>

⁷ <http://www.soilgrids.org>

a vector w_{cat_l} , and consider the following objective:

$$J_{cat} = \sum_{l_i \in R} \sum_{cat_l \in C} (v_{l_i} - w_{cat_l})^2$$

4 Experimental Results

4.1 Evaluation Tasks

We will use the method from [21] as our main baseline. This will allow us to directly evaluate the effectiveness of embeddings for the considered problem, since we have used the same structured datasets and same tag weighting scheme. For this reason, we will also follow their evaluation methodology. In particular, we will consider three evaluation tasks:

1. Predicting the distribution of 100 species across Europe, using the European network of nature protected sites Natura 2000⁸ dataset as ground truth. For each of these species, a binary classification problem is considered. The set of locations L is defined as the 26,425 distinct sites occurring in the dataset.
2. Predicting soil type, again each time treating the task as a binary classification problem, using the same set of locations L as in the species distribution experiments. For these experiments, none of the soil type features are used for generating the embeddings.
3. Predicting CORINE land cover classes at levels 1, 2 and level 3, each time treating the task as a binary classification problem, using the same set of locations L as in the species distribution experiments. For these experiments, none of the CORINE features are used for generating the embeddings.

In addition, we will also consider the following regression tasks:

1. Predicting 5 climate related features: the average precipitation, temperature, solar radiation, water vapor pressure, and wind speed. We again use the same set of locations L as for species distribution in this experiment. None of the climate features is used for constructing the embeddings for this experiment.
2. Predicting people’s subjective opinions of landscape beauty in Britain, using the crowdsourced dataset from the ScenicOrNot website⁹ as ground truth. The set L is chosen as the set of locations of 191 605 rated locations from the ScenicOrNot dataset for which at least one georeferenced Flickr photo exists within a 1 km radius.

4.2 Experimental Setup

In all experiments, we use Support Vector Machines (SVMs) for classification problems and Support Vector Regression (SVR) for regression problems to make

⁸ http://ec.europa.eu/environment/nature/natura2000/index_en.htm

⁹ <http://scenic.mysociety.org/>

	Prec	Rec	F1
BOW-Tags	0.57	0.11	0.18
BOW-KL(Tags)	0.11	0.86	0.19
GloVe	0.10	0.88	0.17
EGEL-Tags	0.10	0.88	0.18
EGEL-Tags+NS	0.12	0.82	0.21
EGEL-KL(Tags+NS)	0.15	0.64	0.25
BOW-All	0.65	0.50	0.56
EGEL-All	0.56	0.60	0.58

Table 1: Results for predicting species distribution.

	Prec	Rec	F1
BOW-Tags	0.17	0.44	0.24
BOW-KL(Tags)	0.30	0.43	0.36
GloVe	0.32	0.39	0.35
EGEL-Tags	0.32	0.40	0.36
EGEL-Tags+NS	0.30	0.44	0.36
EGEL-KL(Tags+NS)	0.32	0.44	0.37
BOW-All	0.39	0.43	0.41
EGEL-All	0.33	0.67	0.44

Table 2: Results for predicting soil type.

	CORINE level 1			CORINE level 2			CORINE level 3		
	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
BOW-Tags	0.49	0.45	0.47	0.20	0.13	0.16	0.14	0.08	0.10
BOW-KL(Tags)	0.40	0.47	0.43	0.39	0.12	0.18	0.24	0.13	0.17
GloVe	0.20	0.90	0.33	0.12	0.53	0.19	0.12	0.25	0.17
EGEL-Tags	0.20	0.89	0.33	0.12	0.56	0.20	0.16	0.21	0.18
EGEL-Tags+NS	0.23	0.73	0.35	0.12	0.52	0.20	0.18	0.22	0.19
EGEL-KL(Tags+NS)	0.26	0.62	0.37	0.14	0.58	0.23	0.19	0.25	0.22
BOW-All	0.52	0.51	0.51	0.27	0.19	0.22	0.18	0.11	0.13
EGEL-All	0.45	0.66	0.54	0.27	0.48	0.35	0.23	0.33	0.27

Table 3: Results for predicting CORINE land cover classes, at levels 1, 2 and 3.

	Temp		Precip		Solar rad		Water vap		Wind speed	
	MAE	ρ	MAE	ρ	MAE	ρ	MAE	ρ	MAE	ρ
BOW-Tags	1.62	0.84	11.66	0.68	926	0.83	0.08	0.71	0.54	0.75
BOW-KL(Tags)	1.69	0.81	12.85	0.65	1057	0.75	0.08	0.71	0.53	0.73
GloVe	1.96	0.44	15.37	0.31	1507	0.36	0.11	0.47	0.74	0.28
EGEL-Tags	1.95	0.47	15.03	0.31	1426	0.41	0.10	0.46	0.73	0.32
EGEL-Tags+NS	1.97	0.44	14.93	0.32	1330	0.44	0.10	0.46	0.72	0.36
EGEL-KL(Tags+NS)	1.48	0.73	13.55	0.52	1008	0.77	0.08	0.66	0.65	0.59
BOW-All	0.72	0.94	10.52	0.75	484	0.93	0.05	0.91	0.43	0.84
EGEL-All	0.71	0.95	10.03	0.79	436	0.95	0.05	0.92	0.43	0.88

Table 4: Results for predicting average climate data.

predictions from our representations of geographic locations. In both cases, we used the SVM^{light} implementation¹⁰ [23]. For each experiment, the set of locations L was split into two-thirds for training, one-sixth for testing, and one-sixth for tuning the parameters. All embedding models are learned with Adagrad using 30 iterations. The number of dimensions is chosen for each experiment from $\{10, 50, 300\}$ based on the tuning data. For the parameters of our model in Equation 1, we considered values of α from $\{0.1, 0.01, 0.001, 0.0001\}$ and values of β from $\{1, 10, 100, 1000, 10\,000, 100\,000\}$.

To compute KL divergence, we need to determine a set of classes C_1, \dots, C_n for each experiment. For classification problems, we can simply consider the given categories, but for the regression problems we need to define such classes by discretizing the numerical values. For the scenicness experiments, we considered scores 3 and 7 as cut-off points, leading to three classes (i.e. less than 3, between 3 and 7, and above 7). Similarly, for each climate related features, we consider two cut-off values for discretization: 5 and 15 for average temperature, 50 and 100 for average precipitation, 10 000 and 17 000 for average solar radiation, 0.7 and 1 for average water vapor pressure, and 3 and 5 for wind speed. The smoothing parameter γ was selected among $\{10, 100, 1000\}$ based on the tuning data. In all experiments where term selection is used, we select the top 100 000 tags. We fixed the radius D at 1km when counting the number of tag occurrences. Finally, we set the number of negative examples as 10 times the number of positive examples for each location, but with a cap at 1000 negative examples in each region for computational reasons. We tune all parameters with respect to the F1 score for the classification tasks, and Spearman ρ for the regression tasks.

4.3 Variants and Baseline Methods

We will refer to our model as EGEL (Embedding GEographic Locations), and will consider the following variants. **EGEL-Tags** only uses the information from the Flickr tags (i.e. component J_{tags}), without using any negative examples and without feature selection. **EGEL-Tags+NS** is similar to **EGEL-Tags** but with the addition of negative examples. **EGEL-KL(Tags+NS)** additionally considers term selection. **EGEL-All** is our full method, i.e. it additionally uses the structured information. We also consider the following baselines. **BOW-Tags** represents locations using a bag-of-words representation, using the same tag weighting as the embedding model. **BOW-KL(Tags)** uses the same representation but after term selection, using the same KL-based method as the embedding model. **BOW-All** combines the bag-of-words representation with the structured information, encoded as proposed in [21]. **GloVe** uses the objective from the original GloVe model for learning location vectors, i.e. this variant differs from **EGEL-Tags** in that instead of $PPMI(t_j, l_i)$ we use the number of co-occurrences of tag t_j near location l_i , measured as $|U_{t_j l_i}|$.

¹⁰ http://www.cs.cornell.edu/people/tj/svm_light/

	MAE	ρ
BOW-Tags	1.01	0.57
BOW-KL(Tags)	1.09	0.51
GloVe	1.27	0.19
EGEL-Tags	1.12	0.37
EGEL-Tags+NS	1.14	0.40
EGEL-KL(Tags+NS)	1.05	0.53
BOW-All	1.00	0.58
EGEL-All	0.94	0.64

Table 5: Results for predicting scenicness.

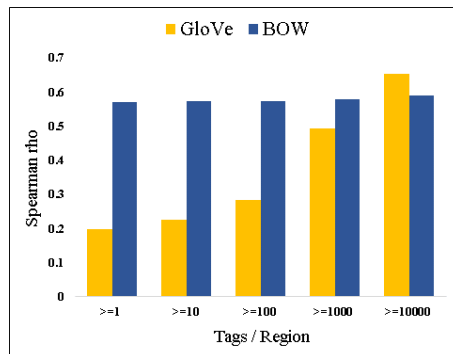


Fig. 1: Comparison between the performance of the GloVe and bag-of-words models for predicting scenicness, as a function of the number of tag occurrences at the considered locations.

4.4 Results and Discussion

We present our results for the binary classification tasks in Tables 1–3 in terms of average precision, average recall and macro average F1 score. The results of the regression tasks are reported in Tables 4 and 5 in terms of the mean absolute error between the predicted and actual scores, as well as the Spearman ρ correlation between the rankings induced by both sets of scores. It can be clearly seen from the results that our proposed method (EGEL-All) can effectively integrate Flickr tags with the available structured information. It outperforms the baselines for all the considered tasks. Furthermore, note that the PPMI-based weighting in EGEL-Tags consistently outperforms GloVe and that both the addition of negative examples and term selection lead to further improvements. The use of term selection leads to particularly substantial improvements for the regression problems.

While our experimental results confirm the usefulness of embeddings for predicting environmental features, this is only consistently the case for the variants that use both the tags and the structured datasets. In particular, comparing BOW-Tags with EGEL-Tags, we sometimes see that the former achieves the best results. While this might seem surprising, it is in accordance with the findings in [24,54], among others, where it was also found that bag-of-words representations can sometimes lead to surprisingly effective baselines. Interestingly, we note that in all cases where EGEL-KL(Tags+NS) performs worse than BOW-Tags, we also find that BOW-KL(Tags) performs worse than BOW-Tags. This suggests that for these tasks there is a very large variation in the kind of tags that can inform the prediction model, possibly including e.g. user-specific tags. Some of the information captured by such highly specific but rare tags is likely to be lost in the embedding.

To further analyze the difference in performance between BoW representations and embeddings, Figure 1 compares the performance of the GloVe model with the bag-of-words model for predicting place scenicness, as a function of the number of tag occurrences at the considered locations. What is clearly noticeable in Figure 1 is that GloVe performs better than the bag-of-words model for large corpora and worse for smaller corpora. This issue has been alleviated in our embedding method by the addition of negative examples.

5 Conclusions

In this paper, we have proposed a model to learn geographic location embeddings using Flickr tags, numerical environmental features, and categorical information. The experimental results show that our model can integrate Flickr tags with structured information in a more effective way than existing methods, leading to substantial improvements over baseline methods on various prediction tasks about the natural environment.

Acknowledgments

Shelan Jeawak has been sponsored by HCED Iraq. Steven Schockaert has been supported by ERC Starting Grant 637277.

References

1. Barve, V.V.: Discovering and developing primary biodiversity data from social networking sites. Ph.D. thesis, University of Kansas (2015)
2. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: *Advances in neural information processing systems*, pp. 2787–2795 (2013)
3. Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. *Computational linguistics* **16**(1), 22–29 (1990)
4. Cocos, A., Callison-Burch, C.: The language of place: Semantic value from geospatial context. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. vol. 2, pp. 99–104 (2017)
5. Cunha, E., Martins, B.: Using one-class classifiers and multiple kernel learning for defining imprecise geographic regions. *International Journal of Geographical Information Science* **28**(11), 2220–2241 (2014)
6. Daume, S.: Mining twitter to monitor invasive alien species an analytical framework and sample information topologies. *Ecological Informatics* **31**, 70–82 (2016)
7. De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., Yu, C.: Constructing travel itineraries from tagged geo-temporal breadcrumbs. In: *Proceedings of the 19th International Conference on World Wide Web*. pp. 1083–1084 (2010)
8. Derrac, J., Schockaert, S.: Inducing semantic relations from conceptual spaces: a data-driven approach to plausible reasoning. *Artificial Intelligence* pp. 74–105 (2015)

9. Eisenstein, J., O'Connor, B., Smith, N.A., Xing, E.P.: A latent variable model for geographic lexical variation. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. pp. 1277–1287 (2010)
10. ElQadi, M.M., Dorin, A., Dyer, A., Burd, M., Bukovac, Z., Shrestha, M.: Mapping species distributions with social media geo-tagged images: Case studies of bees and flowering plants in australia. *Ecological Informatics* **39**, 23–31 (2017)
11. Feng, S., Cong, G., An, B., Chee, Y.M.: Poi2vec: Geographical latent representation for predicting future visitors. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 102–108 (2017)
12. Ge, L., Moh, T.S.: Improving text classification with word embedding. In: IEEE International Conference on Big Data. pp. 1796–1805 (2017)
13. Grave, E., Mikolov, T., Joulin, A., Bojanowski, P.: Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. pp. 427–431 (2017)
14. Grothe, C., Schaab, J.: Automated footprint generation from geotags with kernel density estimation and support vector machines. *Spatial Cognition & Computation* **9**(3), 195–211 (2009)
15. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 855–864 (2016)
16. Guo, S., Wang, Q., Wang, B., Wang, L., Guo, L.: Semantically smooth knowledge graph embedding. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. pp. 84–94 (2015)
17. Gupta, A., Boleda, G., Baroni, M., Padó, S.: Distributional vectors encode referential attributes. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 12–21 (2015)
18. Hasegawa, M., Kobayashi, T., Hayashi, Y.: Social image tags as a source of word embeddings: A task-oriented evaluation. In: LREC. pp. 969–973 (2018)
19. Hollenstein, L., Purves, R.: Exploring place through user-generated content: Using flickr tags to describe city cores. *Journal of Spatial Information Science* (1), 21–48 (2010)
20. Jameel, S., Schockaert, S.: D-glove: A feasible least squares model for estimating word embedding densities. In: Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. pp. 1849–1860 (2016)
21. Jeawak, S., Jones, C., Schockaert, S.: Using flickr for characterizing the environment: an exploratory analysis. In: 13th International Conference on Spatial Information Theory. vol. 86, pp. 21:1–21:13 (2017)
22. Jeawak, S., Jones, C., Schockaert, S.: Mapping wildlife species distribution with social media: Augmenting text classification with species names. In: Proceedings of the 10th International Conference on Geographic Information Science. pp. 34:1–34:6 (2018)
23. Joachims, T.: Making large-scale svm learning practical. Tech. rep., SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund (1998)
24. Joulin, A., Grave, E., Bojanowski, P., Nickel, M., Mikolov, T.: Fast linear model for knowledge graph embeddings. arXiv preprint arXiv:1710.10881 (2017)
25. Kuang, S., Davison, B.D.: Learning word embeddings with chi-square weights for healthcare tweet classification. *Applied Sciences* **7**(8), 846 (2017)
26. Lilleberg, J., Zhu, Y., Zhang, Y.: Support vector machines and word2vec for text classification with semantic features. In: *Cognitive Informatics & Cognitive Com-*

- puting (ICCI* CC), 2015 IEEE 14th International Conference on. pp. 136–140 (2015)
27. Liu, Q., Jiang, H., Wei, S., Ling, Z.H., Hu, Y.: Learning semantic word embeddings based on ordinal knowledge constraints. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. pp. 1501–1511 (2015)
 28. Liu, Q., Ling, Z.H., Jiang, H., Hu, Y.: Part-of-speech relevance weights for learning word embeddings. arXiv preprint arXiv:1603.07695 (2016)
 29. Liu, X., Liu, Y., Li, X.: Exploring the context of locations for personalized location recommendations. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. pp. 1188–1194 (2016)
 30. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119 (2013)
 31. Nickel, M., Kiela, D.: Poincaré embeddings for learning hierarchical representations. In: Advances in Neural Information Processing Systems. pp. 6341–6350 (2017)
 32. Niwa, Y., Nitta, Y.: Co-occurrence vectors from corpora vs. distance vectors from dictionaries. In: Proceedings of the 15th conference on Computational linguistics-Volume 1. pp. 304–309 (1994)
 33. Ono, M., Miwa, M., Sasaki, Y.: Word embedding-based antonym detection using thesauri and distributional information. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 984–989 (2015)
 34. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing. pp. 1532–1543 (2014)
 35. Qiu, L., Cao, Y., Nie, Z., Yu, Y., Rui, Y.: Learning word representation considering proximity and ambiguity. In: AAAI. pp. 1572–1578 (2014)
 36. Quercia, D., Schifanella, R., Aiello, L.M.: The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In: Proceedings of the 25th ACM conference on Hypertext and social media. pp. 116–125 (2014)
 37. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from flickr tags. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 103–110 (2007)
 38. Rattenbury, T., Naaman, M.: Methods for extracting place semantics from flickr tags. *ACM Transactions on the Web* **3**(1), 1 (2009)
 39. Richards, D.R., Friess, D.A.: A rapid indicator of cultural ecosystem service usage at a fine spatial scale: content analysis of social media photographs. *Ecological Indicators* **53**, 187–195 (2015)
 40. Rothe, S., Schütze, H.: Word embedding calculus in meaningful ultradense subspaces. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 512–517 (2016)
 41. Saeidi, M., Riedel, S., Capra, L.: Lower dimensional representations of city neighbourhoods. In: AAAI Workshop: AI for Cities (2015)
 42. Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 4444–4451 (2017)
 43. Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B.: Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd

- Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 1555–1565 (2014)
44. Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: International Conference on Machine Learning. pp. 2071–2080 (2016)
 45. Van Canneyt, S., Schockaert, S., Dhoedt, B.: Discovering and characterizing places of interest using flickr and twitter. *International Journal on Semantic Web and Information Systems (IJSWIS)* **9**(3), 77–104 (2013)
 46. Van Laere, O., Quinn, J.A., Schockaert, S., Dhoedt, B.: Spatially aware term selection for geotagging. *IEEE transactions on Knowledge and Data Engineering* **26**, 221–234 (2014)
 47. Vendrov, I., Kiros, R., Fidler, S., Urtasun, R.: Order-embeddings of images and language. arXiv preprint arXiv:1511.06361 (2015)
 48. Wang, X., Cui, P., Wang, J., Pei, J., Zhu, W., Yang, S.: Community preserving network embedding. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 203–209 (2017)
 49. Weston, J., Bengio, S., Usunier, N.: Large scale image annotation: Learning to rank with joint word-image embeddings. *Machine. Learning* **81**(1), 21–35 (Oct 2010)
 50. Xu, C., Bai, Y., Bian, J., Gao, B., Wang, G., Liu, X., Liu, T.: Rc-net: A general framework for incorporating knowledge into word representations. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. pp. 1219–1228 (2014)
 51. Yan, B., Janowicz, K., Mai, G., Gao, S.: 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. 35:1–35:10 (2017)
 52. Yang, B., Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: Proc. of ICLR-15 (2015)
 53. Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K.: Sensing spatial distribution of urban land use by integrating points-of-interest and google word2vec model. *International Journal of Geographical Information Science* **31**(4), 825–848 (2017)
 54. Zhang, C., Zhang, K., Yuan, Q., Peng, H., Zheng, Y., Hanratty, T., Wang, S., Han, J.: 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 (2017)
 55. Zhao, S., Zhao, T., King, I., Lyu, M.R.: 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 (2017)