Automatic Discovery of Global and Local Equivalence Relationships in Labeled Geo-Spatial Data

Bart Thomee
Yahoo Labs
San Francisco, CA, USA
bthomee@yahoo-inc.com

Gianmarco De Francisci Morales
Yahoo Labs
Barcelona, Spain
gdfm@yahoo-inc.com

ABSTRACT

We propose a novel algorithmic framework to automatically detect which labels refer to the same concept in labeled spatial data. People often use different words and synonyms when referring to the same concept or location. Furthermore these words and their usage vary across culture, language, and place. Our method analyzes the patterns in the spatial distribution of labels to discover equivalence relationships. We evaluate our proposed technique on a large collection of geo-referenced Flickr photos using a semi-automatically constructed ground truth from an existing ontology. Our approach is able to classify equivalent tags with a high accuracy (AUC of 0.85), as well as providing the geographic extent where the relationship holds.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Spatial databases and GIS*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

Keywords

Geotagged data, geo-spatial analysis, folksonomy, relationship discovery, Flickr

1. INTRODUCTION

In recent years, social media and Web applications, such as del.icio.us, have amassed large quantities of user-generated content. Researchers often leverage the "wisdom of the crowd" in order to organize and search this content. In particular, user-supplied *tags*, textual labels assigned to such content, form a powerful and useful feature that has been successfully exploited for this purpose in many different domains.

Tags are free-form, short keywords associated by a user to some content such as a photo, a link, or a blog entry. The categorization system that arises from applying tags is usually called a folksonomy [43]. Unlike ontologies and taxonomies, folksonomies result in unstructured knowledge, i.e. they are flat and have no predefined semantics. However, their unstructured nature is also their strength. For example, tagging has lower cognitive cost than picking categories from an ontology, it allows for greater flexibility and self-expression, and the folksonomy may naturally evolve to reflect emergent properties of the data [40].

The availability of large quantities of data has prompted researchers to try to automatically extract knowledge from it. Formerly ontologies used to be built manually by experts, however with large amounts of data this process has proved to be expensive and unsustainable. Currently, most state-of-the-art methods to generate ontologies use large large text databases [25] as their imput. Often, they leverage crowdsourced data and user-generated content. Many approaches mine Wikipedia or the Web to find interesting relationships, for instance DBPedia[1] and Google's Knowledge Graph[2]. The expectation is that the individual interactions of a large number of agents would lead to global effects that can be observed as semantics. Ontologies would thus become an emergent property of the system as opposed to a fixed formalization of knowledge.

While geo-referenced tags are readily available in large quantities, the problem of referencing relationships geographically has received little attention so far. The presence of the spatial dimension allows us to ask new questions about labeled data, such as *"where does this relationship hold?"*, *"is this relationship valid locally or globally?"*, and *"which tag among several equivalents is most prominent here?"*.

Broadly speaking, we are interested in identifying patterns in the distribution of labels over some domain. In this work we describe a general framework for the spatial domain and focus on geographic patterns for evaluation. Specifically, we look at tags from Flickr[3], a popular photo-sharing website that supports tags and geo-referenced photos. Rather than looking at co-occurrences of tags in photos [39], we analyze the spatial patterns that emerge from the usage of tags, and extract structured information from these patterns in the form of equivalence relationships. The equivalence relationships extracted include synonyms, language variants, misspellings, abbreviations, and even nicknames. For example, our algorithm can determine that the Eiffel Tower may interchangeably be referred to by different tags, such as `la tour eiffel`, `toureiffel`, `eiffelturm`, and `la`

HT'14, September 1–4, 2014, Santiago, Chile.

http://dx.doi.org/10.1145/2631775.2631794.

[1] `http://dbpedia.org`
[2] `http://www.google.com/insidesearch/features/search/knowledge.html`
[3] `http://www.flickr.com`

`dame de fer`, even if these tags would never occur together in the same photo. We further determine whether the equivalence relationship holds globally or holds only locally. For example, in the city of Paris there is a park officially known as 'Le Jardin du Luxembourg', which may be referred to by people when tagging their photos by using the tag `luxembourg` or the tag `park`. In and around this particular park, these two labels can be considered equivalent, and one may be substituted for the other; yet anywhere else in Paris these two labels should not be considered the same, in particular because the label `park` will likely also occur around all other parks in Paris. Conversely, while there are many parks in the city of Luxembourg, the equivalence relationship between these two tags does not hold there either.

There are several applications for a system that automatically uncovers geo-spatial equivalence relationships; next we provide a few examples.

Tag canonicalization and entity linking. Mapping the tags to known entities, and thus implicitly knowing the relationships between entities, can enrich existing knowledge repositories. Such mapping of tags to their canonical forms, results in (*i*) the compression of the tag vocabulary, since several terms can be mapped to a single representative entry, and (*ii*) more emphasis on relevant terms when, for instance, performing term counting for building language models. In fact, any related term that would be considered individually (and thus would be competing for attention) will be considered as part of a larger group as a result of canonicalization. For recommendation and personalization, the vocabulary compression yields a reduced dimensionality of the tag space that needs to be considered, e.g., for matrix factorization. Furthermore, the spatial aspect can be beneficial for contextualization in hyper-local search. For example, a user in a Spanish-speaking country may issue the query "bamba" when searching for a popular Mexican folk song, while a user in Israel issuing this query is more likely to refer to a particular peanut butter-flavored snack.

Tag suggestion, correction, de-duplication, disambiguation and search. The canonicalization process yields relationships between tags; these relationships can hold globally true or be region or area specific. This knowledge can benefit the tagging process by flagging misspelled tags, suggesting tags that have a strong presence in the area where the photo was taken, de-duplicating tags when they refer to the same canonical tag, and disambiguating tags by using geo-spatial and contextual information to find the intended canonical terms (e.g., a user tagging a photo of an orange with "orange" when it is taken inside Orange County in California). In addition, search results can be improved when the user searches for a tag by expanding the search query to also include all other tags that map to the same canonical term, as well as diversifying the results by highlighting all possible meanings of the query if it is ambiguous.

Several challenges need to be faced to successfully mine relationships from geo-spatial labeled user-generated content. First, data is present at several different scales (street, city, region, country, etc.). Second, people do not always behave and tag rationally. This fact translates in a high volume of noise in the data. Finally, the large amount of the data imposes the use of efficient techniques.

The main contributions of this work are the following:

- We investigate the problem of automatically uncovering equivalence relationships from geo-spatial data;
- We propose an algorithm to find equivalence relationships and evaluate it on a collection of Flickr tags;
- We validate our approach on a ground truth generated semi-automatically from an existing ontology.

The remainder of this paper is organized as follows. We first discuss related work in Section 2. Section 3 then introduces our framework while Section 4 describes its implementation. Section 5 presents its evaluation and Section 6 concludes with final remarks and future outlooks.

2. RELATED WORK

At its core, this work is an application of geo-spatial data mining to geo-referenced folksonomies.

Folksonomy is a portmanteau term from *folk* and *taxonomy* [43]. It refers to a classification system that creates meaning from unstructured and collaborative annotations that are used in a practical real-world system. While ontologies are traditionally crafted by experts, and therefore relatively expensive to create, folksonomies arise from the steady-state tagging behavior of users, or so-called "wisdom of the crowd", and are relatively inexpensive to create. Additionally, while ontologies usually have a hierarchical structure, folksonomies are composed by a collection of tags which are all equal, i.e., the structure of a folksonomy is flat.

Passant [29] proposes a method to improve the quality of folksonomies by using ontologies for tag canonicalization, disambiguation and structuring. Knerr [21] proposes a formalization of folksonomies in terms of linked data to enable interoperability across different folksonomies. In the same spirit, MOAT [30] is a framework that enables machine-readable semantics for tags. Limpens et al. [22] provides a survey of proposals that attempt to bridge ontologies and folksonomies. Gruber [12] tries to clarify the relationship between ontologies and folksonomies, and argues that contrasting them is a "false dichotomy".

Differently from these other works, the method proposed in this paper takes advantage of the data available from folksonomies in order to enrich ontologies by uncovering new relationships. Similarly to this work, Mika [27] and Lux and Dosinger [23] generate semantic relationship starting from folksonomies. However, they focus on the social part of the folksonomies and do not take into account the geo-spatial aspect of the problem, which is our main contribution.

Closely related to our work, Schmitz [36] try to induce an ontology from Flickr tags. The ontologies are represented as trees of tags that constitute a taxonomy, i.e., the edges are subsumptions ("is-a" relationships). The method used by the authors is a simple rule based approach based on co-occurences, which resembles the mining of generalized (hierarchical) association rules. However, the authors do not take into account the spatial nature of the data. Furthermore, our framework uses a supervised learning approach.

Geo-spatial data mining refers to knowledge extraction from data with spatial coordinates. Examples include data from geographic information systems (GIS), GPS data and mobile trajectories.

Discovering geographic regions within spatial data has been an active research topic for many years. A variety of

methods can be applied to uncover the hidden spatial structures, most notably techniques such as clustering [9], density estimation [16, 18] and neural networks [15].

In recent years, georeferenced multimedia collections have considerably grown in size, in particular due to the availability of digital cameras with built-in GPS receivers that can automatically attach the geographic location to every photo that is taken. We can therefore find a considerable number of methods in the research literature that try to represent or summarize geographical regions in terms of either the tags associated with the photos in that region [1, 7] or of the photos themselves [35, 2, 19, 17, 44, 5]. For example, language modeling approaches [38, 42, 14] characterize cells that partition the world into discrete units for automatic photo and video geo-localization.

In our work, we make use of geo-spatial areas that describe tags in a folksonomy order to qualify relationships among them. GeoFolk is a bayesian latent topic model that combines tags and geographic coordinates, modeled jointly via a generative process [41]. The authors show its usefulness in the usual contexts of tag recommendation, content classification and clustering. More refined models that can describe non-gaussian distributions have also been proposed more recently [20]. Indeed, these models are one of the possible building blocks that our framework can use to model areas. However, our final goal of finding equivalence relationships is different from what these works address.

Rattenbury et al. [32] extract place and event semantics from Flickr tags. They use patterns in the spatiotemporal distribution of tags to classify them in these two categories. In particular, places exploit spatial patterns, while events use temporal patterns at a given scale. Conversely, we are interested in semantic relationships between tags. Similarly, Zhang et al. [45] try to discover spatiotemporal relationships in collections of tagged photos. However, while their notion of relationship is vague and never explicitly defined, in this work we aim at discovering well-defined semantic relationships. Furthermore, the authors use a fixed coarse grained quantized vector representation of the tag distribution, while we propose a direct estimation of tag distribution from the data, coupled with a dynamic quantization of the distribution based on its spatial extent.

3. ALGORITHMIC FRAMEWORK

In our approach we focus on labeled instances, e.g. tags in the case of online photos, each of which is associated with a location. Our geo-spatial relationship discovery framework employs both unsupervised and supervised learning, and consists of five main steps, namely:

1. **Data representation** – The data distribution of each label is analyzed in order to produce a geo-spatial representation;

2. **Overlap detection** – The data representations of labels are compared to find spatial overlaps that may indicate that a relationship exists between the labels;

3. **Feature generation** – Descriptive features are generated from the geo-spatial representations and their overlap;

4. **Relationship classification** – The features are analyzed in order to determine which, if any, relationship holds between overlapping labels and the geo-spatial extent of this relationship;

5. **Relationship aggregation** – The relationships between labels are optionally analyzed to perform relationship aggregation, e.g., detecting transitivity between three or more relationships.

There are several possible implementations for each of these steps, and our framework does not enforce any particular choice, as long as the various pieces form a consistent pipeline (e.g., the feature generation should work on the chosen data representation). We describe the specific instantiation of the framework we experiment with in Section 4.

The final output of our framework is a set of tuples of the form $(\lambda_a, \lambda_b, E_{ab})$ that represent equivalence relationships between labels λ_a and λ_b. Each relationship is associated with a geo-spatial extent E_{ab} that represents the spatial area where the relationship holds. In the remainder of this section we detail each of the aforementioned steps in the framework.

3.1 Data representation

Given a set of labels Λ, we can define the data collection[4] as $\mathbb{D}_\Lambda = \biguplus \mathbb{D}_\lambda \mid \lambda \in \Lambda$, where the $\biguplus$ operator performs a union of the data instances associated with each label $\mathbb{D}_\lambda$. In case a data item has multiple labels, it will be included in $\mathbb{D}_\Lambda$ as many times its labels. We further define $\mathbb{D}_\lambda \triangleq \{d\}$, where d has a label λ and is represented by a tuple (x, y), which contains a location that is expressed by a coordinate x and a coordinate y. In the specific case of geo-referenced data items, a location refers to a geographic coordinate, where x represents the longitude and y the latitude of the coordinate.

Modeling. For each label $\lambda \in \Lambda$, we individually analyze the geo-spatial distribution of its instances by using unsupervised learning. The algorithm detects where the label exhibits a strong presence, and yields one or more clusters that can be represented as closed surfaces. For example, in two dimensions the surface of each obtained cluster would be formed by the convex hull of all points in the cluster. The choice of which clustering algorithm to apply depends on various factors, such as (*i*) the complexity of the data, (*ii*) the scale of the data, (*iii*) the volume of noise in the data, and (*iv*) the desired degree of accuracy for the modeling.

For example, mean-shift [3], k-means [24] or DBSCAN [8] can be used to separate the data into disjoint clusters, whereas kernel density estimation [34] or Gaussian mixture decomposition [33] can be used to generate probabilistic models of the data. Some clustering methods require pre-setting the number of clusters to detect, while others implicitly perform clustering without requiring to specify the number of clusters [31]; nonetheless, for the former type of clustering methods there exist practical methods (e.g., Bayesian information criterion) to perform model selection, i.e., to choose the best number of clusters.

The union of the detected clusters for a label forms its geo-spatial representation. Similarly to the definition of the data collection, we define the geo-spatial representation of a label as a cluster collection $\mathbb{C}_\Lambda = \biguplus \mathbb{C}_\lambda \mid \lambda \in \Lambda$, where $\mathbb{C}_\lambda \triangleq \{c\}$ and where each cluster c has a label λ and is represented by its geo-spatial surface S.

[4]While we describe our data model using two spatial dimensions for clarity, the extension of our model to higher-dimensional spaces is straightforward.

3.2 Overlap detection

After the geo-spatial representations have been determined for each label, we perform pair-wise comparisons between the representations to find overlaps that may indicate the existence of a relationship between the labels. For each pair of labels λ_a and λ_b, we identify all geo-spatial overlaps between their associated clusters $\mathbb{C}_{\lambda_a}$ and $\mathbb{C}_{\lambda_b}$. Each overlap is represented by a tuple $(\lambda_a, C_a, \lambda_b, C_b, E_{ab})$, where $C_a \subseteq \mathbb{C}_{\lambda_a}$ and $C_b \subseteq \mathbb{C}_{\lambda_b}$, and where E_{ab} is the extent of the geo-spatial intersection between the clusters C_a and C_b. By definition, when there are multiple overlaps between a pair of labels, then no cluster of either label is able to participate in more than one overlap. A further filtering step can be applied to the obtained overlaps in order to discard those whose extents are too small to allow for reliably detecting meaningful relationships.

3.3 Feature generation

Once the overlaps have been identified, we generate features from the union of involved clusters $M_a = \bigcup c \mid c \in C_a$ and $M_b = \bigcup c \mid c \in C_b$, as well as from their geo-spatial intersection $M_{ab} = M_a \bigcap M_b$, as exemplified in Figure 1. The features aim at describing the inherent properties of the geo-spatial extents and densities of the overlapping clusters, thus capturing the important aspects which allow a classifier to determine whether a relationship holds between the overlapping clusters of the labels. We identify three kind of features: *overlap*, *extent* and *density*.

Overlap-based features indicate which kind of relationship likely holds between the overlapping clusters. We can differentiate between various types of overlap, e.g., by using qualitative spatial reasoning to express the overlap between the clusters through different region connection calculi [4]. For instance, a non-tangential proper part relationship between M_a and M_b, i.e., all clusters in M_a are completely covered by the clusters in M_b, would suggest a subsumption relationship is likely.

Extent-based features indicate the strength of the relationship between the labels – the larger the extent of the overlap, the more likely an actual relationship to exist – as well as facilitating the decision process when several relationships have been identified as candidates for describing the overlap. For instance, the relative extent of the overlap may enable an equivalence relationship to be distinguished from a subsumption one.

Density-based features capture how similar the distribution of the data within the extents of the overlapping clusters are, in particular the extent the clusters have in common with respect to their overall extents. These features allow to detect relationships that would be hard to find based on just their overlap or size. For example, when a cluster with a large extent overlaps one or more clusters with small extent an equivalence relationship may still be uncovered between their labels when their normalized densities are similar.

After the feature generation step, we end up with features $F_{a,b,ab} = \{F_a, F_b, F_{ab}\}$ describing each overlap.

3.4 Relationship classification

Different kinds of relationships between labels may be present within the data, such as equivalence and subsumption, where the relationship may vary according to its location. In general, each type of relationship can be treated as a different class in a multi-class classifier. Such a classifier takes as input the features $F_{a,b,ab}$ of an overlap between a pair of labels $\{\lambda_a, \lambda_b\}$, and produces as output a decision r that indicates which, if any, relationship holds between the two labels. A classifier need not produce a binary decision, but may rather assign a confidence score to each type of relationship, which would provide the relationship aggregation step with more refined information to better resolve situations when several relationships are possible.

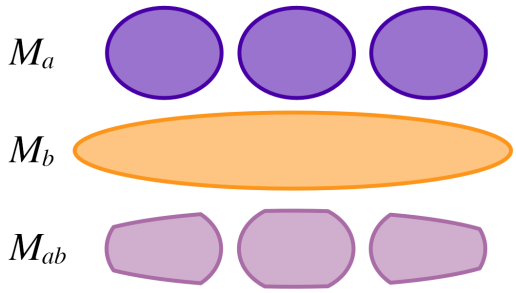

Figure 1: The intersection M_{ab} of the clusters M_a and M_b.

3.5 Relationship aggregation

The final step of our framework optionally performs an aggregation of the relationships detected between each pairs of labels, in order to more accurately deduce which relationships between labels actually hold. For example, if a relationship r is transitive and the output of the classification is as follows:

$$(\lambda_a, \lambda_b, r, E_{ab}), (\lambda_b, \lambda_c, r, E_{bc})$$

we can infer the additional relationship $(\lambda_a, \lambda_c, r, E_{ac})$, where $E_{ac} = E_{ab} \cap E_{bc}$, provided that $E_{ac} \neq \emptyset$. Furthermore, if a relationship between two labels holds across n extents, e.g,

$$(\lambda_a, \lambda_b, r, E^1_{ab}), (\lambda_a, \lambda_b, r, E^2_{ab}), \ldots, (\lambda_a, \lambda_b, r, E^n_{ab})$$

then the relationship also holds on their union $(\lambda_a, \lambda_b, r, E^1_{ab} \cup E^2_{ab} \cup \ldots \cup E^n_{ab})$.

Aggregating the individual relationships for overlaps between pairs of labels provides an opportunity to correct those that have been misclassified. Furthermore adjacent geo-spatial extents may be connected to also cover the extent between them. For instance, the relationship between the labels `holidays` and `vacaciones` may be initially classified as an equivalence relationship that holds in many places in Spanish-speaking countries; these equivalence relationships could be expanded to cover the entire world, even though these labels may have co-occurred in just a few instances.

While our framework is generic in the sense that various types of relationships could be detected with suitable features, classifiers and aggregation, in this paper we solely focus on detecting equivalence relationships between tags, and leave the detection of other relationships to future work.

4. IMPLEMENTATION

In this section we describe one particular instantiation of our algorithmic framework. We discuss the implementation

choices of each of the steps of the framework presented in the previous section.

4.1 Data representation

As explained in Section 3.1, there are several ways to generate an intermediate representation of the raw data that is amenable for detecting overlaps and for generating features. In this work, we represent the data via a probabilistic generative model, which allows us to model the data as a probability density function. We adopt the Gaussian mixture model (GMM) [33] to describe the data, in particular because it provides a sound statistical framework for the approximation of unknown distributions that are not necessarily Gaussian themselves [26].

To obtain the spatial representation of a label we first fit a Gaussian mixture model – this can be done for multiple labels in parallel – in order to obtain one or more mixtures. Each mixture component represents a different subpopulation within the distribution of the data for a label, and is described by a mean μ, a covariance matrix Σ and a prior α that can be considered the "height" of its underlying Gaussian. To fit the model to the data, we employ the Expectation-Maximization (EM) [6] algorithm. There are various options to choose the hyper-parameter k of the model (the number of mixture components). One approach is to use the Bayesian information criterion (BIC) [37]. For simplicity, we choose a practical iterative approach based on cross-validation as implemented in WEKA [13]. In each iteration, we increase the number of components by one and we compute the log-likelihood of the model on a 10-fold cross-validation. If the log-likelihood increases, we repeat the procedure, otherwise we stop.

In line with our earlier terminology we will refer to mixtures as "clusters" henceforth. The spatial representation of a label $\mathbb{C}_\lambda$ is then the set formed by all of its clusters, i.e. $\mathbb{C}_\lambda \triangleq \{c\}$, where each cluster c has label λ and is represented by the tuple (μ, Σ, α), see Figure 2 for an example.

4.2 Overlap detection

Since the values of the Gaussian function never become zero, any cluster represented as a mixture of Gaussians always overlaps with every other cluster. However, in practice, the value of the Gaussian function rapidly decreases to zero and becomes negligible a few standard deviations away from the mean. We therefore apply a threshold ϵ to limit the extent of each cluster $c \in \mathbb{C}_\lambda$ to the area bounded by the level curve $L_c = \{(x, y) \mid \mathcal{N}(\mu_c, \Sigma_c) = \epsilon/\alpha_c\}$. This thresholding makes pairwise comparison of distributions feasible by pruning unrelated clusters.

The cluster prior α_c affects the level curve of a cluster, since the level curve will encompass a larger area given a higher prior than the level curve given a lower prior, due to its more influential Gaussian. The level curve of a cluster may be the empty set when all of its densities are below the threshold; such clusters are unlikely to significantly contribute to any kind of relationship and can be filtered out. For each pair of labels λ_a and λ_b we can now determine all geo-spatial overlaps between their associated clusters $\mathbb{C}_{\lambda_a}$ and $\mathbb{C}_{\lambda_b}$ by checking for intersections between the level curves of their clusters, yielding zero or more overlaps. As mentioned earlier, each overlap is represented by a tuple $(\lambda_a, C_a, \lambda_b, C_b, E_{ab})$, where $C_a \subseteq \mathbb{C}_{\lambda_a}$ and $C_b \subseteq \mathbb{C}_{\lambda_b}$, and where E_{ab} is the extent of the intersection between the clusters C_a and C_b, see Figure 3 for an example.

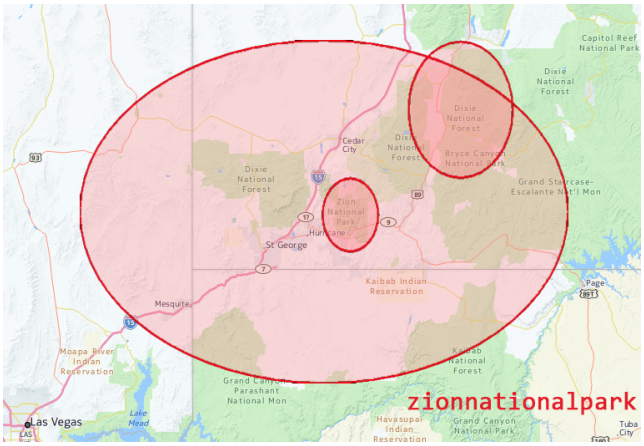

Figure 2: The clusters detected for the tag `zionnationalpark` ($\epsilon = 0.0001$). The densities within the clusters are not shown. The smallest cluster is centered on Zion National Park and has a large prior of 0.92. The largest cluster is also centered on the park, but has a smaller prior of 0.07. The third cluster is centered on Dixie National Forest, albeit only with a prior of 0.01.

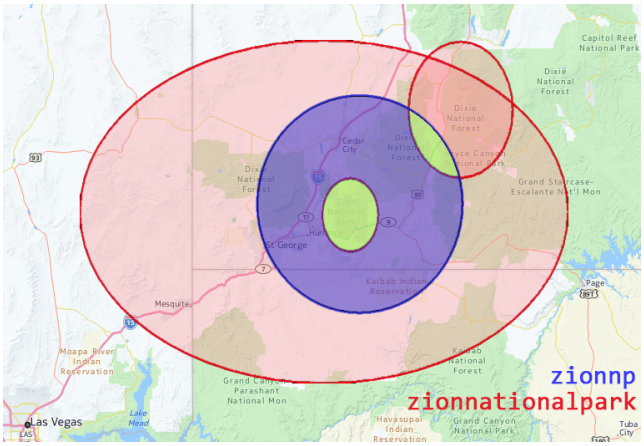

Figure 3: The clusters detected for the tags `zionnationalpark` and `zionnp` ($\epsilon = 0.0001$) and their intersection marked in yellow. The densities within the clusters are not shown.

4.3 Feature generation

For each overlap we extract a number of overlap-based, extent-based and density-based features from the union of involved clusters $M_a = \bigcup c \mid c \in C_a$ and $M_b = \bigcup c \mid c \in C_b$, and from their intersection $M_{ab} = M_a \bigcap M_b$.

Overlap-based features. Following the work in the area of qualitative spatial reasoning, we express the overlaps using region connection calculus. The RCC8 model [4] distinguishes between eight different types of region connections that express all possible ways regions can interact with each other, such as partial overlapping or touching as is illustrated in Figure 4; the model can also be applied when multiple regions are involved. Per overlap we first compute the combined level curves L_a from M_a and L_b from M_b beyond

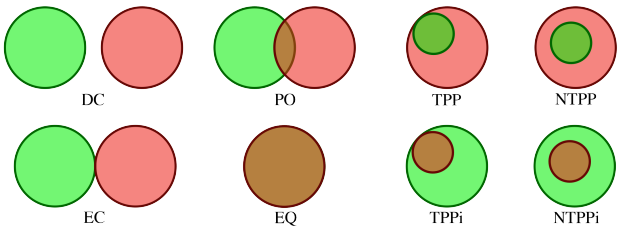

Figure 4: The RCC8 connections. DC = disconnected, EC = externally connected, PO = partially overlapping, EQ = equal, TPP(i) = tangential proper part (inverse), and NTTP(i) = non-tangential proper part (inverse).

which the distributions have value zero, after which we can then compare the areas delineated by the level curves to assign one of the RCC8 region connection types to the overlap.

Extent-based features. We compute several features from the area covered by M_a, M_b and M_{ab}. In order to do so, we first represent the data for each M_i as a two-dimensional binary histogram $f_i(x, y)$, where a histogram bin is activated when its corresponding location falls inside the area enclosed within the combined level curve. All three histograms $f_a(x, y)$, $f_b(x, y)$ and $f_{ab}(x, y)$ span the same geographic area. For computational reasons we resample the geographic area to fit a histogram with a maximum dimension of 2000 bins, either widthwise or lengthwise depending on the horizontal or vertical orientation of the geographic area. From each of the three histograms we then first compute the second and third order complex translation, rotation and scaling invariant moments [10]. Translation, rotation and scaling invariance allows for a robust description of the shapes of M_a and M_b and their intersection M_{ab} by ignoring their horizontal and vertical displacements, orientational differences and their disparity in size. The second order moments capture the distribution of the mass of each shape with respect to the mean, while the third order moments capture their skewness. In addition to these moments we further calculate the sizes of the areas and compute the Jaccard index, i.e. the area of the intersection M_{ab} divided by the area of the union $M_a \bigcup M_b$.

Density-based features. We compute the same translation, rotation and scaling invariant moments as above, but this time we use the actual densities of the clusters rather than their shapes in the histograms $f_i(x, y)$. We furthermore include the location and value of the maximum density and the average density per histogram cell. Finally, we compute the Jensen-Shannon divergence between M_a and M_b, to capture how similar their density distributions are.

In total, we generate 8 overlap-based features, 28 extent-based features and 40 density-based features, yielding a total of 76 features per overlap, see also Table 1.

4.4 Relationship classification

To find an equivalence relationship, we set up the classification task as a binary classification problem, where the positive class represents the existence of the equivalence relationship and the negative class its absence. While in principle any binary classifier can be used for the task, we choose to use tree-based classifiers to gain insight on the relative importance of the features. In particular, we experiment standard decision trees, with random forests and with gradient boosted decision trees. We use WEKA [13] for the decision trees and the random forests, and an in-house implementation for the gradient boosted decision trees. We present the results of the classification task in Section 5.

4.5 Relationship aggregation

Equivalence is a transitive relationship, thus given the binary relationships uncovered in the previous step, we can compute their transitive closure and therefore discover more relationships, as was also earlier described in Section 3.5. Recall that the aggregated relationship will hold on the union of the intersections of the original extents. Since we evaluate our algorithm on a selection of tags that were in part randomly selected, the number of transitive relationships discovered in this manner is therefore small in our dataset. It is nevertheless valuable to analyze the transitive aggregation from a qualitative point of view; we present examples from our dataset in Section 5.3.

5. EXPERIMENTS

We experiment with our relationship discovery framework and validate it on the specific task of finding *synonyms*. Synonyms represent a very well defined subset of equivalence relationships, which is easy to assess. However, manually building a ground truth for such a task is a tedious job. For this reason, we devise a procedure to generate it semi-automatically, so that we only need to check the synonyms and non-synonyms identified by the evaluated algorithms for correctness. Before presenting the experimental results, in the next section we describe the dataset used in our experiments and the procedure used to generate the ground truth.

5.1 Dataset

Flickr. Our Flickr dataset contains a sample of over 56 million geo-referenced images uploaded to Flickr before the end of 2010. Each photo is represented by a geographic location, indicated by longitude and latitude, and one or more tags assigned to the photo by the user. We consider each instance of a tag associated with a photo as a label and annotate each label with the location information of its source photo. A single photo may thus generate multiple instances of annotated labels.

We performed sanitization on the data by removing all non-latin characters, reducing all remaining latin characters to their lowercase representation and removing all diacritics, so that tags like `España` and `Gaudí` become `espana` and `gaudi` respectively. We removed tags that referred to years, such as `2006` and `2007`, as well as camera manufacturer names, such as `canon` and `nikon`, because these at times get automatically added by capture devices or photo applications and thus are not representative of the tagging behavior of users.

We additionally removed infrequently used tags by ensuring we have at least 50 instances per tag occurring around the world. While infrequently used tags could yield a cluster that characterizes an area, Sigurbjornsson and van Zwol [39] show that the tag frequency distribution in Flickr follows a power law where the long tail contains words categorized as occurring incidentally, making it unlikely for such tags to generate a good cluster.

Ground truth. To generate the ground truth, we first extracted all (inter-language and intra-language) synonyms

Table 1: The overlap-based (top), extent-based (middle) and density-based (bottom) features used by our algorithm.

DC	M_a and M_b do not overlap.
EC	M_a and M_b touch each other.
PO	M_a and M_b partially overlap.
EQ	M_a and M_b exactly overlap.
TPP	M_a is a tangential proper part of M_b.
$TPPi$	M_b is a tangential proper part of M_a.
$NTPP$	M_a is a non-tangential proper part of M_b.
$NTPPi$	M_b is a non-tangential proper part of M_a.
$Jaccard$	Jaccard index of M_a and M_b.
$Area_{\{a,b,ab\}}$	Area sizes of the extents of M_a, M_b and of their intersection M_{ab}, respectively.
$\Phi_e(1,1)_{\{r,i\}\{a,b,ab\}}$	Second order invariant (real and complex) moments [10] of the extents of M_a, M_b and M_{ab}.
$\Phi_e(2,0)_{\{r,i\}\{a,b,ab\}}$	Second order invariant (real and complex) moments [10] of the extents of M_a, M_b and M_{ab}.
$\Phi_e(2,1)_{\{r,i\}\{a,b,ab\}}$	Third order invariant (real and complex) moments [10] of the extents of M_a, M_b and M_{ab}.
$\Phi_e(3,0)_{\{r,i\}\{a,b,ab\}}$	Third order invariant (real and complex) moments [10] of the extents of M_a, M_b and M_{ab}.
$JensenShannon$	Jensen-Shannon divergence of M_a and M_b.
$Sum_{\{a,b,ab\}}$	Cumulative densities of M_a, M_b and M_{ab}.
$Avg_{\{a,b,ab\}}$	Average densities of M_a, M_b and M_{ab}.
$MaxX(0,0)_{\{a,b,ab\}}$	Normalized longitude bin $(0-1)$ of maximum density in M_a, M_b and M_{ab}.
$MaxY(0,0)_{\{a,b,ab\}}$	Normalized latitude bin $(0-1)$ of maximum density in M_a, M_b and M_{ab}.
$MaxV(0,0)_{\{a,b,ab\}}$	Maximum density values of M_a, M_b and M_{ab}.
$\Phi_d(1,1)_{\{r,i\}\{a,b,ab\}}$	Second order invariant (real and complex) moments [10] of the densities of M_a, M_b and M_{ab}.
$\Phi_d(2,0)_{\{r,i\}\{a,b,ab\}}$	Second order invariant (real and complex) moments [10] of the densities of M_a, M_b and M_{ab}.
$\Phi_d(2,1)_{\{r,i\}\{a,b,ab\}}$	Third order invariant (real and complex) moments [10] of the densities of M_a, M_b and M_{ab}.
$\Phi_d(3,0)_{\{r,i\}\{a,b,ab\}}$	Third order invariant (real and complex) moments [10] of the densities of M_a, M_b and M_{ab}.

from the BabelNet multilingual ontology [28] and applied the same sanitization described above to each synonym. To avoid any ambiguity in interpretation we retained only those terms that were assigned to a single concept. We computed the intersection between the selected terms and the tag vocabulary in Flickr, ensuring that for each concept at least two terms had been used as photo tags. We then manually inspected all terms per concept to validate whether they truly were synonyms and discarded all terms whose meaning was (possibly) ambiguous. For example, `kivu` and `lakekivu` are considered synonyms in BabelNet for the concept "Lake Kivu", even though in reality the former tag describes a much larger region that encompasses the lake; we thus did not consider these two terms as synonyms. Ultimately, we generated 1240 examples. We formed an equal-sized set of non-synonym tags by including a portion of the terms we discarded earlier and by randomly sampling the remainder from the tag vocabulary that were not in the synonym set.

As described in Section 4, we proceeded to fit a Gaussian mixture model to the data instances of each of the synonyms and non-synonyms to obtain one or more clusters per tag, and then detected the overlapping clusters. Whenever the clusters of a pair of tags generated multiple overlaps we only retained the one where the cumulative priors were highest of all clusters that were part of the overlap. Finally, we computed the features for all clusters involved in the remaining overlaps, and we assigned the class `positive` if both tags in the pair were present in BabelNet as a synonym and `negative` otherwise. We determined the overlaps for varying values of the threshold ϵ. In case the pair of tags did not generate overlapping clusters, the overlap-based feature was set to DC and the extent features and density features of the intersection were set to zero.

5.2 Empirical Results

We evaluate three different classifiers on the ground truth for several values of the threshold ϵ: decision tree (DT), random forests (RF) and gradient boosted decision trees (GBDT). For each classifier and value of ϵ we run a 10-fold cross validation and average results across the folds. For DT we set the minimum number of instances in each leaf to 10 and the confidence factor for pruning to 0.10. For RF we use 50 trees, and for GBDT 20 trees. These settings were chosen empirically in a preliminary test in order to reduce over-fitting. We leave all other settings at their default value.

Table 2 summarizes the results. We report the positive rate (TPR), the false positive rate (FPR) and the area under the ROC curve (AUC). The TPR and FPR are obtained at the optimal point of operation on the ROC curve by giving equal weights to misclassifications. For all three classifiers the AUC is high, where the most powerful classifiers in our evaluation setting is RF, which is able to reach an AUC of 0.85 for $\epsilon = 0.001$. While the GBDT obtains the best TPR, at the same time the FPR is highest, yielding a lower AUC than the RF and one that is comparable to the DT.

When inspecting the generated clusters and their overlaps for different values of ϵ, we observe that higher values lead to fewer and more spatially constrained clusters that less frequently overlap, whereas the reverse is the case for smaller values. Intuitively, the classification performance should increase for higher thresholds of ϵ, since any remaining clusters that do overlap have strong spatial support and similar spatial footprints, and thus are more likely to refer to synonyms than to non-synonyms. Yet, we notice that the performance of the classifiers is relatively stable across all evaluated thresholds ϵ, indicating that the clusters formed for synonymous tags do not always have strong spatial support

nor may they necessarily overlap. To this end we investigate which features are more important than others for a correct classification.

Feature importance. By analyzing our model from the GBDT classifier, we can infer which features contribute the most to the classification, as described by Friedman [11]. Across all models, all of the features in the top-10 are density based. Six of these features refer to the relative location and height of the peaks of the clusters and their intersection, one feature refers to the Jensen-Shannon divergence and the remaining three features are second and third order image density moments. Interestingly, when we look beyond the top-10, the overlap-based features from the RCC8 model seem to contribute little to the classification, which may be due to the mismatch between the GMM and the model for which the region connection calculus was originally devised, and deserves further study. Overall our results indicate that the Gaussian mixtures model is very suitable for modeling geo-spatial labeled data and that the features extracted from the clusters and their intersection, primarily the density-based features, are sufficiently informative for the classification task.

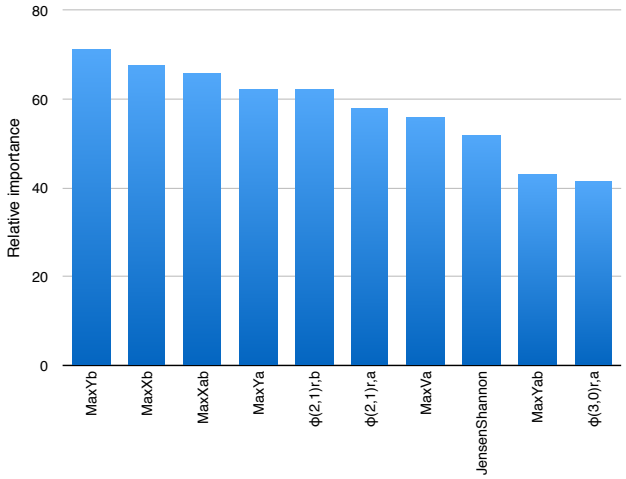

Figure 5: Estimated relative importance of the top-10 features for GBDT.

Comparison. We implement a baseline based on spatial tag correspondences to place the results of our method into context. Here, we discretize the world by considering it as a large histogram, where each cell has a size of ρ° longitude by ρ° latitude. For example, $\rho = 0.0001$ yields cells of about 10x10m each, while $\rho = 0.01$ yields cells of about 1x1km. We quantize the location of each tag instance to a cell in the histogram, after which we count the number of instances per cell. We then represent the global distribution of non-empty cells for a tag by a sparse feature vector, after which we compare the vectors of pairs of tags using cosine similarity, producing a single feature representing their similarity. We finally evaluate the performance of the baseline, again using the Decision Tree, Random Forest and Gradient Boosted Decision Tree, and report the same metrics as before.

Table 3 summarizes the results. We can see that the baseline achieves worse results in comparison with our method; our method outperforms the baseline for all three classifiers.

Table 2: Evaluation results for our algorithm for the Decision Tree (DT), Random Forest (RF) and Gradient Boosted Decision Tree (GBDT) classifiers. We report the true positive rate (TPR), false positive rate (FPR) and the area under the curve (AUC) for several parameter values ϵ.

ϵ	DT			RF			GBDT		
	TPR	FPR	AUC	TPR	FPR	AUC	TPR	FPR	AUC
0.000001	0.762	0.274	0.759	0.777	0.223	0.832	0.773	0.257	0.737
0.000005	0.746	0.254	0.767	0.763	0.237	0.822	0.808	0.287	0.748
0.00001	0.729	0.271	0.759	0.776	0.224	0.819	0.818	0.318	0.752
0.00005	0.737	0.263	0.773	0.778	0.222	0.823	0.729	0.241	0.721
0.0001	0.733	0.267	0.753	0.770	0.230	0.819	0.829	0.409	0.679
0.0005	0.733	0.267	0.761	0.771	0.229	0.844	0.869	0.313	0.737
0.001	0.762	0.238	0.792	0.787	0.213	0.848	0.848	0.240	0.791

Table 3: Evaluation results for the spatial tag correspondence baseline for the Decision Tree (DT), Random Forest (RF) and Gradient Boosted Decision Tree (GBDT) classifiers. We report the true positive rate (TPR), false positive rate (FPR) and the area under the curve (AUC) for several parameter values ρ.

ρ	DT			RF			GBDT		
	TPR	FPR	AUC	TPR	FPR	AUC	TPR	FPR	AUC
0.0001	0.545	0.455	0.537	0.531	0.469	0.554	0.101	0.014	0.180
0.0005	0.592	0.408	0.602	0.587	0.413	0.611	0.205	0.012	0.290
0.001	0.612	0.388	0.615	0.615	0.385	0.645	0.252	0.025	0.354
0.005	0.663	0.337	0.670	0.658	0.342	0.701	0.350	0.035	0.485
0.01	0.697	0.303	0.675	0.688	0.312	0.747	0.402	0.028	0.559

We notice that the baseline performs particularly badly for the GBDT when ρ is small, although even when ρ is large it still underperforms with respect to the DT and RF; while the GBDT classifier is effective in achieving a low FPR, it is not able to produce a high enough TPR to yield a competitive performance.

The baseline is an approach that only considers the occurrences of tags at a global level, and as such it is incapable of detecting equivalence relationships that only hold locally and nowhere else. In contrast, our method only classifies tags as being equivalent in the area spanned by the union of their overlapping clusters, unless the aggregation step expands this area. Furthermore, the size of the cells can considerably influence the accuracy of the classifier, where for very small cells the number of instances per individual cell may be very small and produce many cells per tag, which may result in dissimilar feature vectors for equivalent tags. Conversely, for very large cells the number of instances per individual cell may be very large and produce few cells per tag, which may result in similar feature vectors for non-equivalent tags. In contrast, our method automatically detects the optimal scale at which to represent the tag distribution. The advantage of the baseline over our method, however, is that it is not as complex and is able to detect equivalence relationships that are globally valid.

5.3 Anecdotal Results

In this section we provide several examples of clusters and the obtained classified overlaps to provide a visual illustration of the output of our algorithm and the classifiers. Correctly classified positive examples include the tag pairs `baleari` – `balearics` (an archipelago of Spain) and `occupylondon` – `occupylsx` (the Occupy London movement), while incorrectly classified positive examples include `arcdetriom-`

phe – grandearche (two famous landmarks in Paris, France) and nijocastle – nijojo (a castle located in Kyoto, Japan). Incorrectly classified negative examples also occur, such as the tag pair portknockie – bowfiddlerock (a coastal village town in Scotland and its characteristic landmark). We show a few examples in Figure 6. We observe that positive tag pair examples may be misclassified as a result of so-called "bulk tagging" of photos, where people apply the same set of tags to all photos they took that day, while negative tag pair examples are typically misclassified due to subsumption relationships with both having similar tag distributions (e.g. the tags barcelona and spain).

In terms of detecting transitive relationships, the tags louvre – museedulouvre (referring to the famous museum the Louvre in Paris) are positively identified as an equivalence relationship, as were louvremuseum – museedulouvre, both spanning roughly the same geographic area. Thus, we are able to connect the tags louvre – louvremuseum together, whereby their relationship holds in the intersection of their extents.

Interestingly enough our classifiers pointed out mistakes in our ground truth labeling. For example, what we thought to be a negative example, barajas – lemd, is classified as positive; it turns out that the tags refer to the shortened name of Madrid-Barajas Airport and its four-character ICAO airport code.

6. CONCLUSIONS

In this paper we introduced a novel algorithmic framework to automatically discover equivalence relationships in georeferenced folksonomies. Differently from previous work in the literature, our method is based on the geo-spatial patterns of the tags in the folksonomy, rather than on their co-occurrences in the content. Furthermore, by using a combination of unsupervised and supervised learning, we discover well-defined equivalence relationship rather than generic clusters of related tags as done previously in the literature. Our method is generic and can be applied to any labeled data with geo-spatial coordinates. We validated our framework on the task of discovering synonyms from their geo-spatial patterns. We built a ground truth semi-automatically by leveraging an existing ontology. Our method is able to correctly classify the synonyms with very high accuracy, and the best classifier achieves an AUC of 0.85.

This work can be extended in several directions. On the modeling part, we plan to explore different data representations, such as non-parametric DP-GMM. Furthermore, in this paper we used elliptical Gaussians as the basic mixture. However, by using a diagonal or full covariance matrix, which allows the mixtures to be more freely placed, more complex shapes and more powerful models can be obtained. This characteristic might be useful in discovering relationships different from equivalence.

In this work we mainly focused on validating the approach on a known ground truth, however the capabilities of the framework still need to be evaluated further. We plan to perform a large scale evaluation of the output of the classification. Given the size of the task, crowdsourcing the evaluation of the discovered relationships seems a viable option.

Finally, while in this paper we focused on equivalence relationships, we will expand the relationships that can be discovered to include more different kinds. For example, subsumption relationships are a prime candidate that would require minimal modifications to the framework. Furthermore, by including temporal features in our analysis, we will explore the possibility of uncovering periodic relationships in the data.

7. REFERENCES

[1] S. Ahern, M. Naaman, R. Nair, and J. Yang. World explorer: visualizing aggregate data from unstructured text in georeferenced collections. *JDCL*, pp. 1–10, 2007.

[2] W.-C. Chen, A. Battestini, N. Gelfand, and V. Setlur. Visual summaries of popular landmarks from community photo collections. *ICMR*, pp. 782–789, 2009.

[3] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[4] A. G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts. Qualitative spatial representation and reasoning with the Region Connection Calculus. *GeoInformatica*, 1(3):275–316, 1997.

[5] M. Cristani, A. Perina, U. Castellani, and V. Murino. Content visualization and management of geo-located image databases. *CHI*, pp. 2823–2828, 2008.

[6] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1977.

[7] D.-P. Deng, T.-R. Chuang, and R. Lemmens. Conceptualization of place via spatial clustering and co-occurrence analysis. *LBSN*, pp. 49–56, 2009.

[8] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, volume 1996, pp. 226–231. AAAI Press, 1996.

[9] V. Estivill-Castro and I. Lee. AUTOCLUST: automatic clustering via boundary extraction for mining massive point-data sets. *GeoComputation*, 2001.

[10] J. Flusser, T. Suk, and B. Zitova. *Moments and moment invariants in pattern recognition*, chapter 2. John Wiley & Sons, Ltd., 2009.

[11] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pp. 1189–1232, 2001.

[12] T. Gruber. Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems*, 3(1):1–11, 2007.

[13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10, 2009.

[14] C. Hauff and G.-J. Houben. Geo-location estimation of Flickr images: social web based enrichment. *ECIR*, pp. 85–96, 2012.

[15] H. Hotta and M. Hagiwara. A neural-network-based geographic tendency visualization. *WI-IAT*, pp. 817–823, 2008.

[16] H. Hotta and M. Hagiwara. Online geovisualization with fast kernel density estimator. *WI-IAT*, pp. 622–625, 2009.

[17] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of georeferenced photographs. *MIR*, pp. 89–98, 2006.

[18] C. B. Jones, R. S. Purves, P. D. Clough, and H. Joho. Modelling vague places with knowledge from the Web. *IJGIS*, 22 (10):1045–1065, 2008.

[19] L. S. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. *WWW*, pp. 297–306, 2008.

[20] C. C. Kling, J. Kunegis, S. Sizov, and S. Staab. Detecting non-gaussian geographical topics in tagged photo collections. In *WSDM*, pp. 603–612, 2014.

[21] T. Knerr. Tagging Ontology - Towards a Common Ontology for Folksonomies. 2006.

[22] F. Limpens, F. Gandon, and M. Buffa. Bridging Ontologies and Folksonomies to leverage knowledge sharing on the social web: a brief survey. In *Automated Software Engineering-Workshops*, pp. 13–18. IEEE, 2008.

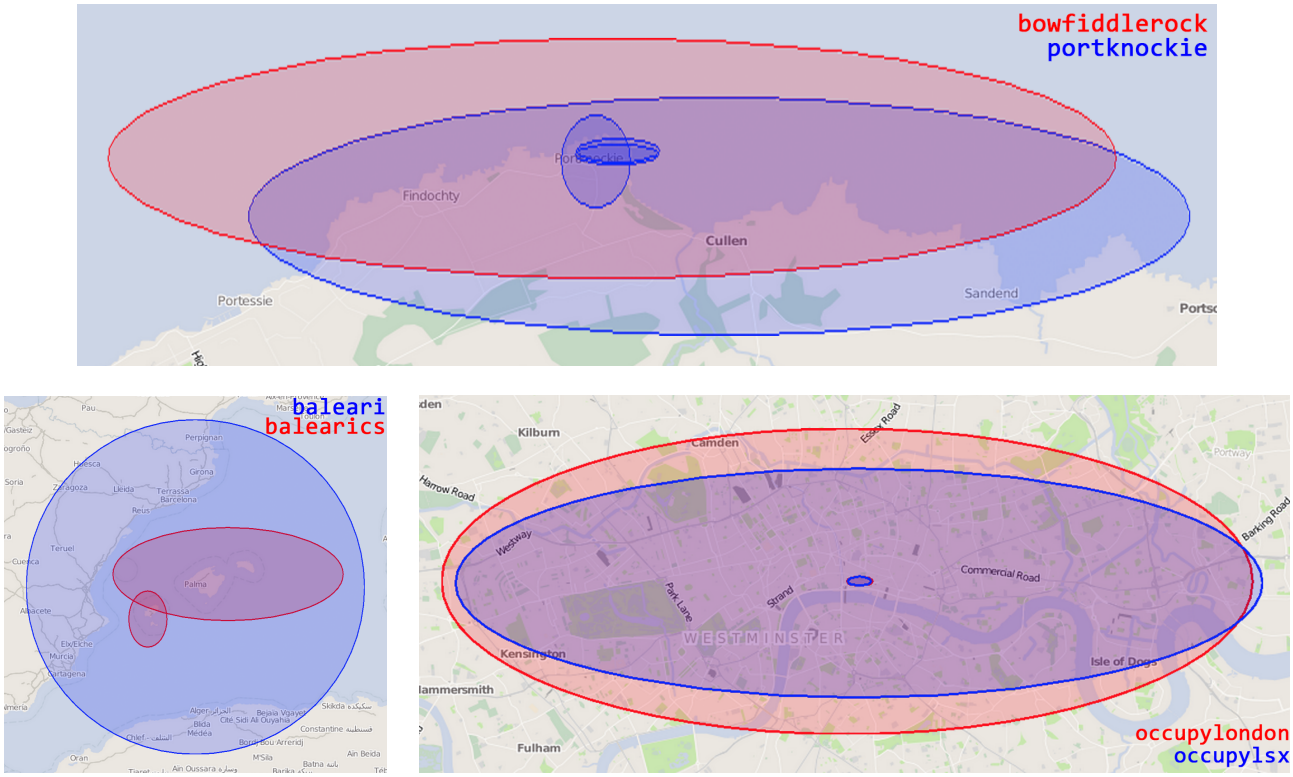

Figure 6: Example overlaps between the tag clusters of `portknockie` – `bowfiddlerock` (top), `baleari` – `balearics` (bottom-left) and `occupylondon` – `occupylsx` (bottom-right). The densities within the clusters and overlaps are not shown, but in all three examples the peaks of the densities within the clusters are located in close proximity of each other. The density peaks of `baleari` and `balearics` are principally located on the islands, whereas those of `occupylondon` and `occupylsx` are centered near St. Paul's Cathedral where one of the protester camps of the Occupy London movement was based.

[23] M. Lux and G. Dosinger. From folksonomies to ontologies: employing wisdom of the crowds to serve learning purposes. *International Journal of Knowledge and Learning*, 3(4):515–528, 2007.

[24] J. MacQueen. Some methods for classification and analysis of multivariate observations. *5th Berkeley symposium on mathematical statistics and probability*, 1967.

[25] A. Maedche and S. Staab. Mining ontologies from text. In *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*, pp. 189–202. Springer, 2000.

[26] J. S. Marron and M. P. Wand. Exact Mean Integrated Squared Error. *The Annals of Statistics*, 20(2):712–736, 1992.

[27] P. Mika. Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(1):5–15, 2007.

[28] R. Navigli and S. Ponzetto. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, 2012.

[29] A. Passant. Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs. In *ICWSM*, 2007.

[30] A. Passant. Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In *LDOW workshop at WWW*. Citeseer, 2008.

[31] C. Rasmussen. The infinite Gaussian mixture model. *NIPS*, 1999.

[32] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from Flickr tags. In *SIGIR*, pp. 103–110, 2007.

[33] R. Redner and H. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*, 1984.

[34] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 1956.

[35] S. Rudinac, A. Hanjalic, and M. Larson. Finding representative and diverse community contributed images to create visual summaries of geographic areas. *ICMR*, pp. 1109–1112, 2011.

[36] P. Schmitz. Inducing ontology from flickr tags. In *Collaborative Web Tagging Workshop at WWW*, volume 50, 2006.

[37] G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 1978.

[38] P. Serdyukov, V. Murdock, and R. van Zwol. Placing Flickr photos on a map. *SIGIR*, pp. 484–491, 2009.

[39] B. Sigurbjornsson and R. van Zwol. Flickr Tag Recommendation based on Collective Knowledge. *WWW*, pp. 327–336, 2008.

[40] R. Sinha. A cognitive analysis of tagging, 2005.

[41] S. Sizov. GeoFolk: Latent Spatial Semantics in Web 2.0 Social Media. In *WSDM*, pp. 281–290, Feb. 2010.

[42] O. Van Laere, S. Schockaert, and B. Dhoedt. Finding locations of Flickr resources using language models and similarity search. *ICMR*, p. 48, 2011.

[43] T. Vander Wal. Folksonomy Coinage and Definition, 2007.

[44] K. Yanai, H. Kawakubo, and B. Qiu. A visual analysis of the relationship between word concepts and geographical locations. *CIVR*, p. 13, 2009.

[45] H. Zhang, M. Korayem, E. You, and D. J. Crandall. Beyond co-occurrence: discovering and visualizing tag relationships from geo-spatial and temporal similarities. In *WSDM*, p. 33, 2012.