Georeferencing Flickr photos using language models at different levels of granularity: an evidence based approach

Olivier Van Laere[a], Steven Schockaert[b], Bart Dhoedt[a]

[a]*Department of Information Technology, Ghent University, IBBT, Belgium*
[b]*School of Computer Science & Informatics, Cardiff University, United Kingdom*

Abstract

The topic of automatically assigning geographic coordinates to Web 2.0 resources based on their tags has recently gained considerable attention. However, the coordinates that are produced by automated techniques are necessarily variable, since not all resources are described by tags that are sufficiently descriptive. Thus there is a need for adaptive techniques that assign locations to photos at the right level of granularity, or, in some cases, even refrain from making any estimations regarding location at all. To this end, we consider the idea of training language models at different levels of granularity, and combining the evidence provided by these language models using Dempster and Shafer's theory of evidence. We provide experimental results which clearly confirm that the increased spatial awareness that is thus gained allows us to make better informed decisions, and moreover increases the overall accuracy of the individual language models.

Keywords: Dempster-Shafer evidence theory, Language models, Georeferencing, Web 2.0, Geographic information retrieval

1. Introduction

In addition to topical relevance, the geographic scope of a web resource is often paramount for assessing its relevance. Inspired by this observation, geographic information retrieval (GIR) systems attempt to identify spatial constraints in queries, and to determine which web resources satisfy them [1, 2]. This requires appropriate, structured geographic background information, which is available in the form of gazetteers. However, as gazetteers are often restricted to administrative places or are otherwise incomplete, many of the names people use to refer to places (i.e. vernacular place names) are not recognized. Moreover, in determining the geographic scope of a web resource, other terms than toponyms may play a key role (e.g. the names of local events). As a result, there has been a recent interest in the automated acquisition of geographic knowledge from online resources which are already georeferenced, e.g. utilizing information provided by users in tagging-based systems such as Flickr [3, 4, 5, 6], other types of social websites [7, 8], or even local business directories such as Yahoo! local [9]. What is common to these approaches is that they rely on resources containing both geographic coordinates and textual descriptions (typically in the form of tags) to find correlations between locations and linguistic descriptions. These correlations are then used to obtain *geographic information* in the sense of [11, 12, 13], i.e. tuples of the form $< x, y, z, t, U >$ where U represents a 'thing' which was present at location (x, y, z) at time t. Note that in the aforementioned works U is referred to by some web object; e.g. a Flickr photo or Twitter post refers to the presence of a user at a particular location.

Given this importance of large-scale repositories of georeferenced resources, it is of interest to increase the number of resources for which appropriate geo-annotations exist. In the case of Flickr, for instance, coordinates are only available for a small fraction[1]. A number of recent research efforts have been directed towards automatically finding (approximate) coordinates of Flickr photos [14, 2, 16]. The importance of this task is twofold. On one hand, it shows how we may directly georeference online resources, without the intermediate construction of a gazetteer or other forms of explicit spatial semantics of toponyms. On the other hand, it allows to make a larger number of georeferenced Flickr photos available, which is interesting per se (e.g. to allow spatial browsing by displaying them on a map). Note that the idea of using Flickr tags to derive geo-annotations, as a form of semantic information about a photo, fits within a broader trend to use Web 2.0 data sources to bootstrap the semantic web. For example [17] suggests building *collective knowledge systems* by integrating user-contributed content from the Social Web and machine-gathered (semantic) data. Taking this idea

Email addresses: `olivier.vanlaere@intec.ugent.be` (Olivier Van Laere), `S.Schockaert@cs.cardiff.ac.uk` (Steven Schockaert), `bart.dhoedt@intec.ugent.be` (Bart Dhoedt)

[1] `http://www.flickr.com/map/` shows that around 168M photos are geotagged of over 6.46 billion photos (`http://www.flickr.com/explore`) on Flickr. Accessed on December 6th, 2011.

one step further, the *DBPedia Mobile* client proposed in [18] allows a user to browse location related information and semantically interlinked data sources, but at the same time also to contribute to the overall geospatial semantic web by publishing content that is linked with nearby DBPedia resources.

Existing work indicates that language models are particularly suitable for the task of assigning coordinates to Flickr photos [14, 19]. The geographic space is then discretized into a set of disjoint areas. After training a language model for each of these areas, we may determine which one is most likely to contain the true location of a given photo. A drawback of this approach is that it must be decided a priori what is the most suitable granularity at which the location of each photo should be determined. Clearly, such a view is at odds with the observation that the tags of some photos are more indicative of a specific place (e.g. *Central Park, New York*) than others (e.g. *picnic*).

The solution we propose in this paper is to train language models at different levels of granularity, and subsequently decide the most appropriate granularity level for each individual photo. Although we then still need to choose a specific number of clusters for each granularity level, this avoids having to fix the overall scale at which each photo should be georeferenced. In this decision, there is a trade-off between accuracy and informativeness. Essentially, we choose the finest granularity at which the most likely area is sufficiently probable. In contrast to standard language modeling approaches, the actual probabilities that come out of the language models thus become important, rather than only the ranking that is imposed by them. Since such probabilities are known to be poorly calibrated, in this paper, we study the effect of two forms of post-processing that are applied to these probabilities. First, we consider a standard approach for calibrating classifier probabilities, based on the well-known PAV (pair-adjacent violators) algorithm. The second form of post-processing relies on the spatial dimension of the problem setting. In particular, we propose an approach based on Dempster and Shafer's theory of evidence [20, 21], which allows us to deal with probabilistic information at different levels of granularity in a natural way. Moreover, the theory dictates how evidence coming from different sources — in this case the language models of areas at different granularity levels — can be combined.

The paper is structured as follows. First, in Section 2 we explain how our training and test data was selected, what relevant meta-data is available for Flickr photos, and which preprocessing we have performed. Next, Section 3 recalls the basic approach to georeferencing Flickr photos based on language models, and it explains how the resulting probabilities can be calibrated. The core of our approach is presented in Section 4, where we show how the probabilities produced by language models may be encoded as belief functions in the sense of Shafer, and how these belief functions may be combined with each other to arrive at a single belief function capturing all available evidence. Section 5 then explains how we may use belief functions in practice. Subsequently, Section 6 presents our experimental findings. Finally, we provide an overview of related work and conclude.

This paper is a substantially revised and extended version of [22]; the main extensions are as follows. First, the belief functions are now built from calibrated language model probabilities, whereas we used the raw probabilities in [22]. Second, we now consider more combination operators, and a different decision rule based on pignistic probability. Furthermore, to have a better mapping among different granularity levels, we now use one hierarchical clustering, rather an independent flat clustering for each level. Finally, the experimental results have been significantly extended, using a more representative data set.

2. Data acquisition and preprocessing

For each photo that is uploaded to its website, Flickr maintains several types of meta-data, which can be obtained via its publicly available API. In this paper, two types of meta-data will be relevant: descriptive tags that have been provided by the photo owners, and for some photos, information about where they were taken. The location information includes a geographical coordinate (latitude and longitude), and information about the accuracy of the location, encoded as a number between 1 (world-level) and 16 (street-level).

The data set we have used consists of two parts. The first part contains the 3 185 343 photos that were provided to the participants of the 2010 MediaEval Placing Task[2], a recent benchmarking initiative on the topic of automatically georeferencing Flick videos. In July 2010, we crawled Flickr in order to expand this initial data set. The query used for this additional crawl constrained the resulting photos to those with an accuracy of at least 12, to ensure that all coordinates were meaningful w.r.t. within-city location. Once retrieved, photos that did not contain any tags or whose coordinates were not valid were removed from the collection. As a result, we obtained an additional data set containing the 5 500 368 most recently georeferenced images (at that time). Combining these two sets resulted in a data set consisting of 8 685 711 georeferenced photos covering more or less the entire world.

In a preprocessing phase, we removed duplicates, i.e. photos of the same user that have an identical tag set (to reduce the impact of bulk uploads [14]). Once filtered, the remaining data set of 3 265 331 photos was divided into a training set of 2 176 719 photos ($2/3^{rd}$), a separate training set of 1 038 612 photos ($1/3^{rd}$ - 50K) that will be used for calibration of the probabilities, and a test set

[2] `http://www.multimediaeval.org/mediaeval2010/placing/index.html`

Table 1: Size of the considered data sets

Training set	2 176 719 photos
Calibration set	1 038 612 photos
Test set	50 000 photos
Total	3 265 331 photos

Table 2: Mean and standard deviation for the number of tags per photo in each data set.

Data set	Mean	Standard deviation
training set	9.34	8.24
calibration set	9.27	8.07
test set	9.20	7.95

of 50 000 photos. When separating training data from calibration and test data, we ensured that all photos from the same user were either in the training set, or in the calibration and test sets (to avoid an unfair exploitation of user-specific tags [23]). Tables 1 and 2 provide some characteristics of the different data sets. A plot of the coordinates of the photos from the training set is shown in Figure 1.

The task of estimating the location where a photo was taken can be seen as a classification problem: for each unseen photo t from the test set, we then determine which area a from a given set of areas $\mathcal{A}$ is most likely to contain this location. To create this set of areas $\mathcal{A}$, a k-medoids clustering algorithm (PAM - Partitioning Around Medoids) with geodesic distance was used to cluster the locations of the photos in the training set into 2000 disjoint areas. Note that the k-medoids algorithm was preferred over k-means as it handles the occurrence of outliers better. Among all coordinates, the initial k medoids are randomly chosen. In a subsequent step, initial clusters are obtained by associating the remaining coordinates to the closest medoid (in terms of geodesic distance). Next, for each cluster C, the new medoid is chosen as the element $c \in C$ minimizing

$$\sum_{c' \in C} d(c, c')$$

where the cluster C is identified with its set of coordinates, and d refers to geodesic distance. New clusters can then be obtained from these medoids by again assigning each coordinate to the closest medoid. This process is repeated until the cluster configuration does not change anymore. In this paper, we will consider different levels of granularity, with 2000 areas being the finest level. To obtain coarser granularity levels, we subsequently used agglomerative hierarchical clustering on this initial clustering, leading to clusterings into 1000, 500, 250 and 50 areas. This step of agglomerative clustering was accomplished by repeatedly merging those two clusters whose medoids were closest to each other w.r.t. geodesic distance. Note that each cluster at one of the coarser granularity levels then exactly corresponds to the union of one or more of the areas of the finest clustering. Note that alternative clustering algorithms, such as a grid based approach [46], mean shift clustering [14] or even a classification based on administrative boundaries can be used for this task; all we require is that each cluster from a coarser granularity level can be seen as the union of one or more clusters from the finest clustering. Examples of our clusterings are shown in Figure 2 and Figure 3, showing only the clusters located in Europe for clarity. To illustrate the characteristics of the different granularity levels, Table 3 provides the mean and standard deviation of the size of the clusters, where the size of a cluster C is taken to be the maximal distance between the medoid and any other member of the cluster.

Table 3: Mean and standard deviation of the size of the clusters in terms of kilometers.

Granularity	Mean (km)	Standard deviation (km)
50	529.91	457.74
250	177.84	180.62
500	113.44	117.97
1000	68.58	70.76
2000	39.68	41.82

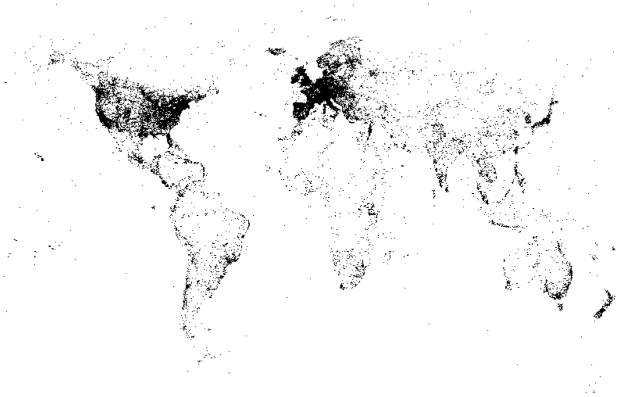

Figure 1: Plot of the training set

Next, a vocabulary V consisting of 'interesting' tags is compiled, which are tags that are likely to be indicative of geographic location. We used χ^2 feature selection to determine for each area in $\mathcal{A}$ the m most important tags[3]. The vocabulary V was then obtained by taking for each area a, the m tags with highest χ^2 value. The m values which we have used are 62 500 for the coarsest clustering, 12 500, 2 500, 500 for the intermediate resolutions and 100 for the finest clustering level. This choice of features ensures that the language models, introduced next, require approximately the same amount of memory space for each clustering level[4].

[3] Initial experiments have shown χ^2 feature selection to perform slightly better than mutual information on this task.

[4] Space requirements increase quadratically with the number of clusters.

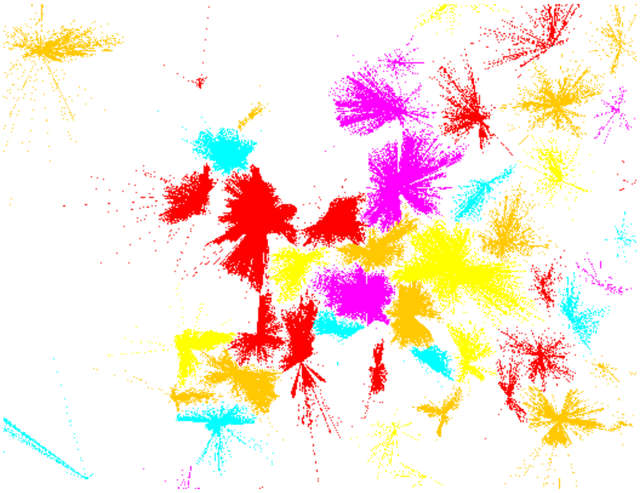

Figure 2: Coarse clustering of Europe ($|\mathcal{A}| = 250$)

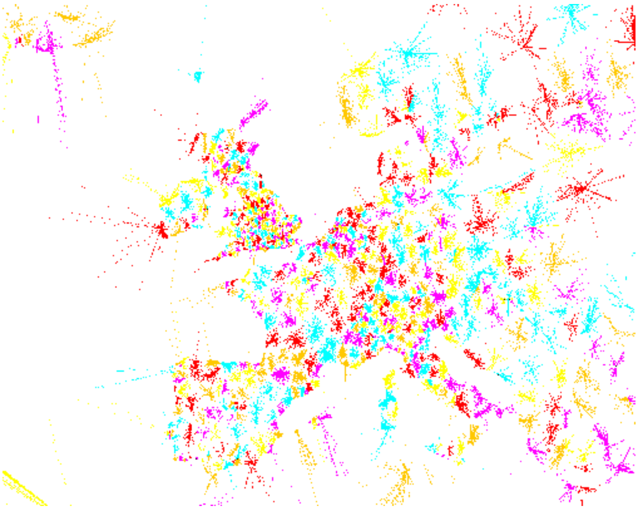

Figure 3: Fine clustering of Europe ($|\mathcal{A}| = 2000$)

3. Calibrated language models for estimating location

3.1. Language models

Let $\mathcal{A}$ be a set of (disjoint) areas, obtained by clustering the locations of the photos in our training set. For the ease of presentation, we identify an area $a \in \mathcal{A}$ with the corresponding set of photos that were taken in it. Given a previously unseen photo x, we try to determine in which area x was most likely taken by comparing its tags with those of the images in the training set. Previous work [14, 19, 2] has revealed that probabilistic (unigram) language models [24] are particularly useful to this end. The probability $p(a|x)$ that image x was taken in area a is then taken to be proportional to

$$p(a|x) \propto p(a) \cdot \prod_{t \in x} p(t|a) \tag{1}$$

which corresponds to using a multinomial Naive Bayes classifier to assign areas to photos. The prior probability $p(a)$ of area a can be estimated using the maximum likelihood method:

$$p(a) = \frac{|X_a|}{N}$$

To avoid a zero probability when x contains a tag that does not occur in area a, some form of smoothing is needed when estimating $p(t|a)$. Let $D_a(t)$ be the occurrence count of tag t in area a. The total tag occurrence count D_a of area a is then defined as follows:

$$|D_a| = \sum_{t \in V} D_a(t)$$

where V is the vocabulary that was obtained after feature selection, as explained in Section 2. One possible smoothing method is Bayesian smoothing with Dirichlet priors, in which case we have ($\mu > 0$):

$$p(t|a) = \frac{D_a(t) + \mu\, p(t|C)}{|D_a| + \mu}$$

in which the probabilistic model of the collection $p(t|C)$ is defined using maximum likelihood:

$$p(t|C) = \frac{\sum_{a' \in \mathcal{A}} D_{a'}(t)}{\sum_{a' \in \mathcal{A}} \sum_{t' \in V} D_{a'}(t')}$$

Another possibility is to use Jelinek-Mercer smoothing, in which case we have ($\lambda \in [0, 1]$):

$$p(t|a) = \lambda \frac{D_a(t)}{|D_a|} + (1 - \lambda)\, p(t|C)$$

We have experimentally found these two smoothing techniques to yield comparable results (for optimal values of the parameters $\mu = 1750$ and $\lambda = 0.80$), although Bayesian smoothing was found to be more robust w.r.t. the choice of the parameter. These findings conform to experimental results in other areas of information retrieval [25, 26], and to earlier work on georeferencing Flickr photos [14].

As we focus on the effect of different granularity levels in this paper, we restrict ourselves to a rather standard language modelling approach. Note, however, that the model presented in this section can be refined in different ways, using additional information about the owner, information from visual features, etc. For example, [15] and [44] use the home location of the user, while [54] uses information about her social network. As another form of refinement, in [14] a location-aware from of smoothing is used.

3.2. Calibration

In principle, an estimation of the actual value of $p(a|x)$, for all $a \in \mathcal{A}$, is found from (1) after normalization. However, it is well-known that Naive Bayes does not produce well-calibrated probability estimates [27]. As our approach will strongly depend on the actual values of the probability estimates, we need to apply some form of calibration. In [28], an approach called *binning* is shown to produce such well-calibrated probabilities. In [29], an extension of this method based on the PAV (pair-adjacent violators [30]) algorithm is proposed, which we have adopted in our experiments. In particular, let us write $n(a|x)$ for the normalised Naive Bayes output, i.e.:

$$n(a|x) = \frac{score(a|x)}{\sum_{a' \in \mathcal{A}} score(a'|x)} \tag{2}$$

where $score(a|x)$ denotes the estimation of the right-hand side of (1).

Some care needs to be taken to avoid underflow or a significant loss of precision, as the values $score(a|x)$ tend to be very small. As usual, these values can be calculated in log-space, i.e.

$$\log score(a|x) = \log p(a) + \sum_{t \in x} \log p(t|a)$$

The normalization cannot be carried out in log-space, so we rewrite the denominator in equation (2) in the following way:

$$\log \sum_{a' \in \mathcal{A}} score(a'|x) \tag{3}$$

$$= (\log \sum_{a' \in \mathcal{A}} \gamma \cdot score(a'|x)) - \log \gamma \tag{4}$$

$$= (\log \sum_{a' \in \mathcal{A}} \exp \log(\gamma \cdot score(a'|x)) - \log \gamma \tag{5}$$

$$= (\log \sum_{a' \in \mathcal{A}} \exp(\log(\gamma) + \log(score(a'|x))) - \log \gamma \tag{6}$$

By choosing γ sufficiently high, problems of reduced precision can be avoided; we have used

$$\log \gamma = \max_{a' \in \mathcal{A}} abs(\log(score(a'|x)))$$

In this way, $\exp(\log(\gamma) + \log(score(a'|x)))$ in equation (6) becomes $exp(0) = 1$ for the most plausible areas a', which avoids both underflow and overflow for the probability of those areas. Note that, if underflow occurs for the probability of less plausible areas, this is then because their probability is extremely small compared to the most plausible area, in which case we can safely ignore them.

The PAV algorithm is now used to map the scores $n(a|x_i)$ to accurate probability estimates, as follows [31, 32]:

- Assume that the photos $x_1, ..., x_m$ from the training set are ranked such that $n(a|x_i) \geq n(a|x_{i+1})$ for all i.
- At each stage of the algorithm, a list of bins is maintained. Let us write $B(i, j)$ for the bin that contains the images $x_i, x_{i+1}, ..., x_j$. Initially the list L contains one bin for each photo, i.e. $L = \{B(i, i) | 1 \leq i \leq m\}$. For a given bin $B^1 = B(i, j)$, we write $avg(B^1)$ for the percentage of photos in bin B^1 that actually belong to the area a.
- Let $L = (B^1, ..., B^p)$. Until it holds that $avg(B^i) \geq avg(B^{i+1})$ for all i, repeat the following
  1. Find all maximal subsequences of bins $B^i, ..., B^j$ in the list such that $avg(B^r) \leq avg(B^{r+1})$ for all $r \in \{i, i + 1, ..., j - 1\}$.
  2. Replace these subsequences in the list L by the single bin $B = B^i \cup B^{i+1} \cup ... \cup B^j$.

To ensure that meaningful probability estimates are obtained, as an additional step, we also merge each bin containing fewer than 100 items with the bin succeeding it. This is especially important for the first bin, which we otherwise found to provide an unrealistically optimistic estimation. For instance, if the highest ranked photo were correctly georeferenced, the highest bin would always be associated with a probability of 1.

Let $L = (B^1, ..., B^p)$ be the final list of bins that is obtained from this procedure. Each bin B naturally corresponds to an interval $bounds(B) = [\underline{n}, \overline{n}]$ where $\underline{n} = \min_{x \in B} n(a|x)$ and $\overline{n} = \max_{x \in B} n(a|x)$. For a given photo x from the test set, we then determine the bin B for which $bounds(B)$ contains $n(a|X)$, or whose bounds are closest to $n(a|x)$. A probability estimate $p(a|x)$ is then given by $avg(B)$. Note that $\sum_a p(a|x)$ may be different from 1. However, we refrain from normalizing these estimates at this stage, as initial experiments have shown that this may largely nullify the effect of the calibration process.

4. Combining language models of different granularity levels

The language modeling approach that was outlined in Section 3 is not spatially aware in the sense that e.g. neighboring areas are treated in the same way as areas that are located in different parts of the world. To see why this difference might be important, assume that the probability $p(.|x)$ takes a high value for two different areas a and b. If a and b are adjacent or close to each other, it makes sense to estimate the location of x at a coarser level of granularity, using an area c as result which encompasses both a and b. Indeed, the fact that all areas that are considered plausible are spatially close suggests that our estimation will be near the actual location of x, while the available information is not sufficient to distinguish reliably between a and b. In contrast, when a and b are not close, the choice between a and b is likely to be a problem of disambiguation. In such as case, it makes more sense to first determine the most likely area c at a coarser granularity level, and take a to be the result if c contains a (but not b), and b if c contains b (but not a).

Our solution uses Dempster-Shafer evidence theory [20, 21] to combine the probability distributions obtained from language models that operate at different resolutions. Based on the agreement between fine-grained models and coarse-grained models, we may then try to find the most plausible region in which a photo was taken, at the most appropriate resolution given the available information. Essentially, our approach then finds the smallest region for which all models agree (to a sufficient degree) to contain the true location with high probability.

4.1. Belief functions

Let $\{\mathcal{A}_1, ..., \mathcal{A}_k\}$ be different clusterings of the locations in the training set such that $|\mathcal{A}_1| > |\mathcal{A}_2| > ... > |\mathcal{A}_k|$, i.e. $\mathcal{A}_1$ corresponds to the finest clustering and $\mathcal{A}_k$ corresponds to the coarsest clustering. For each clustering, a language model is obtained which (after calibration) results in a probability distribution $p_i(.|x)$ in the universe $\mathcal{A}_i$ for each image x. A key observation is that the spatial extension of each area a in $\mathcal{A}_i$ corresponds to the union of the spatial extensions of a set of areas from the finest level $\mathcal{A}_1$, as the different clusterings have been obtained in a hierarchical fashion. Let us write $\mathit{areas}(a)$ for this set of areas from $\mathcal{A}_1$ that are included in a. Then, if a is the area maximizing $p(.|x)$, we can take this as evidence that the correct area, at the finest level, is among those of the set $\mathit{areas}(a)$. In other words, the probability distributions $p_2, ..., p_k$ naturally correspond to probability distributions on the power set of $\mathcal{A}_1$, i.e. to belief functions on $\mathcal{A}_1$.

Recall that a belief function [21] on a finite universe U is any $2^U \to [0,1]$ mapping m satisfying $\sum_{X \subseteq U} m(X) = 1$ and $m(\emptyset) = 0$; belief functions are also called mass assignments. Intuitively, $m(X)$ represents the amount of evidence that the correct value of some variable is among those in X. Subsets X such that $m(X) > 0$ are called focal elements. Starting from a belief function m, two measures of uncertainty are usually considered:

$$Bel(X) = \sum_{Y \subseteq X} m(Y) \qquad Pl(X) = \sum_{Y \cap X \neq \emptyset} m(Y)$$

for any $X \subseteq U$. The degree of belief $Bel(X)$ can be interpreted as a lower bound on the probability that X contains the correct value, while the degree of plausibility $Pl(X)$ is an upper bound for this probability.

Probability distributions essentially model variability, i.e. the phenomenon that the outcome of a given experiment may not always be the same, but they lack the capability of genuinely modelling epistemic uncertainty, i.e. the uncertainty resulting from a lack of information. For example, suppose that we know with perfect certainty that the outcome of rolling a die was among the values $\{1, 2, 3\}$. In probability theory, we are left with assigning an equal probability to each of these values, i.e. $p(1) = p(2) = p(3) = \frac{1}{3}$. However, this probability distribution is not a faithful representation of the beliefs that we hold: why should we be able to infer that it is twice as likely that the outcome was odd than that the outcome was even, if all we started off with was the knowledge that the outcome was in $\{1, 2, 3\}$. Using belief functions, on the other hand, we can distinguish between the mass assignment m_1 defined by $m_1(\{1, 2, 3\}) = 1$, and the mass assignment m_2 defined by $m_2(\{1\}) = m_2(\{2\}) = m_2(\{3\}) = \frac{1}{3}$. In other words, belief functions are capable of modelling both variability and epistemic uncertainty.

Note that in the special case where all focal elements are singletons, belief functions simply correspond to probability distributions. Specifically, if we define $p(x) = m(\{x\})$, it holds that $P(X) = Bel(X) = Pl(X)$ for every $X \subseteq U$, where P is the probability measure associated with p, and Bel and Pl are the belief and plausibility measures associated with m.

Nonetheless, when it comes to making decisions based on our available beliefs, the choice between m_1 and m_2 may actually not matter. When deciding whether or not to accept a bet, for instance, all we can do is assume an equal probability for each outcome, i.e. apply the maximum entropy principle. The point of using belief functions, however, is to apply this maximum entropy principle *after* all the available evidence is combined. In other words, a difference is made between the *credal* level, which is concerned with modelling the beliefs of an agent, and the decision or *pignistic* level (from the Latin word *pignus* for bet). Specifically, when it comes to decision making based on belief functions, a mass assignement m is often converted into the associated *pignistic probability distribution* p defined by [33]

$$p(x) = \sum_{\emptyset \subset X \subseteq U, x \in X} \frac{m(X)}{|X|}$$

after which decisions may be made using standard approaches (e.g. based on maximizing expected utility).

4.2. Obtaining mass assignments

In the context of this paper, the universe U will always be the set of areas (clusters) of the most fine-grained clustering $\mathcal{A}_1$. For a given photo x, the different granularity levels lead to mass assignments $m_1, ..., m_k$ defined as follows. First, at each granularity level i, a set $\mathcal{S}_i$ containing the most likely areas from $\mathcal{A}_i$ is determined. In principle, we could take $\mathcal{A}_i = \mathcal{S}_i$, but in practice, a smaller set $\mathcal{S}_i$ is desirable to keep the approach time- and space-efficient. In our experiments, the set $\mathcal{S}_i$ was obtained by adding areas in decreasing order of likelihood until $\sum_{a \in \mathcal{S}_i} p_i(a|x) \geq \theta$ for some fixed parameter θ_i (e.g. $\theta_i = 0.95$). Recall that the probability estimates $p_i(a|x)$ are not necessarily normalized, i.e. they do not necessarily sum to 1. However, in all but a few cases we have that $\sum_{a \in \mathcal{S}_i} p_i(a|x) < 1$. Then we define:

$$m_i^x(X) = \begin{cases} p_i(a|x) & \text{if } X = \mathit{areas}(a) \text{ for } a \in \mathcal{S}_i \\ 1 - \sum_{a \in \mathcal{S}_i} p_i(a|x) & \text{if } X = \mathcal{A}_1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Note that the probability $p_i(a|x)$ is translated to the mass $m_i^x(a)$ for areas a in $\mathcal{S}_i$. The remaining mass corresponding to the areas outside $\mathcal{S}_i$, i.e. $\sum_{a \in (\mathcal{A}_i \setminus \mathcal{S}_i)} p_i(a)$ is assigned to the entire universe $\mathcal{A}_1$. This mass will be approximately equal to $1 - \theta_i$ and reflects the probability that we are ignorant about the location of x. Choosing a lower value of θ_i will thus lead to a more cautious and less informative mass assignment.

Finally, in the rare cases where $s^* = \sum_{a \in \mathcal{S}_i} p_i(a|x) \geq 1$, the probability estimates are first normalized as

$$p_i^*(a|x) = \frac{p_i(a|x)}{s^*}$$

and the mass assignment is defined as in (7), but based on the normalized estimates $p_i^*(a|x)$ instead of $p_i(a|x)$.

4.3. Combining evidence

Different belief functions may encode the evidence provided by different sources, in which case a *combination operator* may be used to obtain a single, combined belief function. In particular, given two belief functions m and m' in a universe U, Dempster [20] proposes to model the combined evidence using the mass assignment $m \oplus m'$ defined as

$$(m \oplus m')(\emptyset) = 0 \tag{8}$$

$$(m \oplus m')(X) = \frac{\sum_{Y \cap Z = X} m(Y) \cdot m'(Z)}{1 - \sum_{Y \cap Z = \emptyset} m(Y) \cdot m'(Z)} \tag{9}$$

for any subset $\emptyset \subset X \subseteq U$, and provided that

$$\sum_{Y \cap Z = \emptyset} m(Y) \cdot m'(Z) < 1$$

The denominator in (9) is a normalization factor, which corresponds to the mass that would normally be assigned to the empty set, i.e. it is a measure of the amount of conflict between m and m'. It can be shown that this combination rule is associative.

By treating the different granularity levels as independent sources, the overall evidence about the location of a photo x may thus be described by the belief function m^x:

$$m^x = m_1^x \oplus m_2^x \oplus ... \oplus m_k^x \tag{10}$$

Example 1. *Let us go back to the scenario outlined in the beginning of Section 4. In particular, assume that there are only two granularity levels, and $\mathcal{S}_1 = \{a, b\}$ and $\mathcal{S}_2 = \{u, v\}$. At the finest level, we are thus faced with the choice of a or b as the location of a given photo x. First assume that $areas(u)$ contains both a and b. In this case, the focal elements of m^x are $\{a\}$, $\{b\}$, $areas(u)$, $areas(v)$, and $\mathcal{A}_1$; we obtain*

$$\begin{aligned}
m^x(\{a\}) &= K \cdot m_1^x(\{a\}) \cdot (m_2^x(areas(u)) + m_2^x(\mathcal{A}_1)) \\
m^x(\{b\}) &= K \cdot m_1^x(\{b\}) \cdot (m_2^x(areas(u)) + m_2^x(\mathcal{A}_1)) \\
m^x(areas(u)) &= K \cdot m_1^x(\mathcal{A}_1) \cdot m_2^x(areas(u)) \\
m^x(areas(v)) &= K \cdot m_1^x(\mathcal{A}_1) \cdot m_2^x(areas(v)) \\
m^x(\mathcal{A}_1) &= K \cdot m_1^x(\mathcal{A}_1) \cdot m_2^x(\mathcal{A}_1)
\end{aligned}$$

where K is the normalization constant. Assuming that $m_1^x(\mathcal{A}_1)$ and $m_2^x(\mathcal{A}_1)$ are sufficiently small, we have

$$m^x(areas(u)) \approx m^x(areas(v)) \approx m^x(\mathcal{A}_1) \approx 0$$

and thus $K \approx m^x(\{a\}) + m^x(\{b\})$; we obtain

$$\begin{aligned}
&Bel(\{a\}) \\
&\approx \frac{m_1^x(\{a\}) \cdot m_2^x(areas(u))}{m_1^x(\{a\}) \cdot m_2^x(areas(u)) + m_1^x(\{b\}) \cdot m_2^x(areas(u))} \\
&= \frac{p_1(a|x)}{p_1(a|x) + p_1(b|x)} \\
&Bel(\{b\}) \\
&\approx \frac{m_1^x(\{b\}) \cdot m_2^x(areas(u))}{m_1^x(\{a\}) \cdot m_2^x(areas(u)) + m_1^x(\{b\}) \cdot m_2^x(areas(u))} \\
&= \frac{p_1(b|x)}{p_1(a|x) + p_1(b|x)} \\
&Bel(areas(u)) \approx 1
\end{aligned}$$

Note that because v does not overlap with any area of $\mathcal{S}_1$, most of the mass $m_2^x(areas(v))$ disappears in the normalization constant K. If u is a clear winner at the second level, i.e. $p_2(u|x) \gg p_2(v|x)$, without a clear winner at the first level, we thus obtain strong evidence that the correct location is in u, but much weaker evidence for a or b individually.

Now consider a second scenario in which $a \in areas(u)$ while $b \in areas(v)$. We then get

$$m^x(\{b\}) = K \cdot m_1^x(\{b\}) \cdot (m_2^x(areas(v)) + m_2^x(\mathcal{A}_1))$$

and $m^x(\{a\})$, $m^x(areas(u))$, $m^x(areas(v))$ and $m^x(\mathcal{A}_1)$ as before. Again assuming that $m_1^x(\mathcal{A}_1)$ and $m_2^x(\mathcal{A}_1)$ are sufficiently small, we have

$$\begin{aligned}
&Bel(\{a\}) \\
&\approx \frac{m_1^x(\{a\}) \cdot m_2^x(areas(u))}{m_1^x(\{a\}) \cdot m_2^x(areas(u)) + m_1^x(\{b\}) \cdot m_2^x(areas(v))} \\
&Bel(\{b\}) \\
&\approx \frac{m_1^x(\{b\}) \cdot m_2^x(areas(u))}{m_1^x(\{a\}) \cdot m_2^x(areas(u)) + m_1^x(\{b\}) \cdot m_2^x(areas(v))}
\end{aligned}$$

If we moreover again make the assumption that $p_2(u|x) \gg p_2(v|x)$, we get

$$\begin{aligned}
Bel(\{a\}) \approx Bel(areas(u)) \approx 1 \\
Bel(\{b\}) \approx Bel(areas(v)) \approx 0
\end{aligned}$$

Hence in this case, we do obtain strong evidence for a. Note that in the latter scenario the evidence from the second granularity level has allowed us to make a decision between a and b, while in the former scenario it has rather provided a more cautious alternative, avoiding a somewhat arbitrary choice between a and b.

The combination rule (8)–(9) is the first and most widely known combination rule, already proposed by Dempster in the 1960s. It has been argued by several authors that it constitutes the only principled way to combine independent and reliable sources in a conjunctive way [34, 35]. On the other hand, from an application perspective, when the degree of conflict $\sum_{Y \cap Z=\emptyset} m(Y) \cdot m'(Z)$ is close to 1, it is reputed to provide counterintuitive results [36]. Moreover, when the degree of conflict is equal to 1, the result of the combination is not even defined. To cope with this, when using Dempster's rule, we first apply some form of discounting, i.e. each mass assignment m is replaced by the mass assignment m_δ, defined by

$$m_\delta(A) = \delta \cdot m(A)$$

if A is different from the universe U, and

$$m_\delta(U) = \delta \cdot m(U) + (1 - \delta)$$

In our experiments, we use $\delta = 0.99$. Note that this indeed guarantees that the degree of conflict is strictly smaller than 1.

Another solution, which is adopted in the transferable belief model (TBM) of Smets [37], is to simply allow a nonzero mass for the empty set. We thus obtain the following combination operator:

$$(m_1 \odot m_2)(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \tag{11}$$

After the final mass assignment has been determined, the mass of the empty set is than added to the mass of the universe. The resulting combination operator is sometimes called Yager's rule [38] ($\emptyset \subset A \subset U$):

$$(m_1 \otimes' ... \otimes' m_k)(A) = (m_1 \odot ... \odot m_k)(A) \tag{12}$$

$$(m_1 \otimes' ... \otimes' m_k)(U) = (m_1 \odot ... \odot m_k)(U) + (m_1 \odot ... \odot m_k)(\emptyset) \tag{13}$$

$$(m_1 \otimes' ... \otimes' m_k)(\emptyset) = 0 \tag{14}$$

Note that unlike $\otimes$ and $\odot$, the operator $\otimes'$ underlying Yager's combination rule is not associative. Dubois and Prade have proposed the following alternative way of distributing the mass of the empty set [39]:

$$(m_1 \otimes'' m_2)(A) = (m_1 \odot m_2)(A) + \sum_{B \cup C = A, B \cap C = \emptyset} m_1(B) \cdot m_2(C) \tag{15}$$

The underlying intuition here is that in the presence of conflicts, we should take the point of view that one of the sources is correct, which leads to a disjunctive combination of conflicting evidence and the requirement that $B \cup C = A$ in the right-hand side of (15).

5. Using belief functions in geographic information retrieval

By combining $m_1^x, ..., m_k^x$ using either of the combination operators, we obtain a single mass assignment m^x summarizing the available evidence about the location of x. In many cases, some post-processing of this mass assignment will be needed to obtain usable approximations of the location of x, e.g. in the form of a precise point, a precise region (i.e a polygon), or a fuzzy region (i.e. a nested set of polygons). Indeed, unlike simple representations such as points and polygons, mass assignments cannot readily be spatially indexed, which is a prerequisite if we are to use georeferencing of photos to support online location-based querying [1]. Moreover, mass assignments, unlike probability distributions and fuzzy regions, cannot be visualized in a way which is sufficiently intuitive for end users. How exactly x's location should be represented in the result depends on the precise requirements of the application context:

Supporting location-based queries Consider a user indicating an interest in photos that were taken in Manhattan. In such a case, we could simply use the mass assignment m^x of each photo x to calculate the belief or plausibility that x was taken in Manhattan, the latter being represented as a union of elements from $\mathcal{A}_1$. Similarly, if a user is interested in photos that were taken in the vicinity of a particular point-of-interest, we could determine the belief or plausibility that each photo in the collection was taken within a given radius of that point. When the mass assignments have been converted to points (the most likely location of x) or polygons (a confidence region for x) that have been spatially indexed a priori, location-based querying becomes computationally feasible.

Helping users georeference their photos When users upload a photo to Flickr, they have the option to indicate on a map where it was taken. When the user has already provided a number of tags for the photo, it makes sense to analyze these tags, and already zoom in on this map at where the photo was likely taken. In this way, less effort is required by the user, which may lead to more users georeferencing their photos, with a higher accuracy level. This application not only requires the system to determine where to center the map, but also to determine at which zoom level it should be shown. This boils down, conceptually, to finding the smallest area containing the true location of x with a given confidence level, i.e. a confidence region for x.

Visualizing plausible locations In some applications, we may simply provide the user with a visual summary of where a photo was likely taken. One of the most obvious ways to do this is by presenting a heat map, which may conceptually be seen as a mapping from locations to the unit interval [0,1], i.e. a possibility distribution [40] of locations. This requires to determine an appropriate approximation of m^x.

It seems that from an application point of view, mass assignments are mainly useful (i) to find the most likely area, at a given granularity level, in which the photo was taken, (ii) to find the most fine-grained area that contains the true location of the photo at a given confidence level, and (iii) to obtain a visual summary of the plausible locations. These three uses are discussed below.

5.1. Finding the most plausible area

The probability distribution $p_i(.|x)$ obtained by calibrating the language models of the areas in $\mathcal{A}_i$ naturally allows us to determine the most plausible area from $\mathcal{A}_i$, viz. the area a maximizing $p_i(a|x)$. The mass assignment m^x has been obtained by combining p_i with other pieces of evidence (i.e. the probability distributions over the other levels of granularity), and may thus allow us to determine the most plausible location of $\mathcal{A}_i$ in a better-informed way. In general, one could also think of combining p_i with belief functions encoding information from other sources of evidence such as gazetteers or visual feature information to obtain m^x. Obvious decision rules are choosing the area a maximizing the belief measure and choosing the area maximizing the plausibility measure:

$$\mathit{choose}_{Bel}(\mathcal{A}_i, m^x) = \arg\max_{a \in \mathcal{A}_i} Bel(\mathit{areas}(a)) \tag{16}$$

$$\mathit{choose}_{Pl}(\mathcal{A}_i, m^x) = \arg\max_{a \in \mathcal{A}_i} Pl(\mathit{areas}(a)) \tag{17}$$

A third decision rule uses the pignistic probability measure P^x induced by m:

$$\mathit{choose}_{P}(\mathcal{A}_i, m^x) = \arg\max_{a \in \mathcal{A}_i} P^x(\mathit{areas}(a)) \tag{18}$$

5.2. Determining confidence regions

Rather than first fixing the granularity level and then determining the most plausible area, it often makes sense to look for the smallest area that contains a given photo x with some predefined confidence level, where *confidence* may be measured in terms of belief, plausibility or pignistic probability. An important question is which areas are to be considered for the result. Either we may restrict ourselves to the areas in $\bigcup_i \mathcal{A}_i$, or we may allow arbitrary subsets of $\mathcal{A}_1$, possibly with the restriction that the chosen subset constitutes a connected (or even convex) region. The solution which we have adopted in the experiments is based on the former choice, which is considerably easier from a computational point of view. Moreover, as the areas in $\bigcup_i \mathcal{A}_i$ have all been obtained from clustering the training data, they likely correspond to meaningful geographic entities. For instance, if all of the most plausible areas from $\mathcal{A}_1$ are in Manhattan, it often makes more sense to use the entire region of Manhattan as result, rather than the disjoint union of these specific areas within Manhattan. The situation where available information is ambiguous forms an exception to this view: if all we know is that a photo was taken in Washington, it makes sense to represent the result e.g. as the union of Washington D.C. and Washington state.

The procedure to determine a confidence region then becomes the following. First, we check whether our confidence in the most likely area a from $\mathcal{A}_1$ — determined e.g. using choose_P, choose_{Bel} or choose_{Pl} — is sufficiently high. This confidence could again be measured in terms of pignistic probability, belief or plausibility. If this is the case, region a is taken as the result. Otherwise, we check whether our confidence in the most likely area from $\mathcal{A}_2$ is sufficiently high, etc. If even our confidence in the most likely area from $\mathcal{A}_k$ is too low, it seems reasonable to acknowledge that no reliable location could be determined for the corresponding photo.

5.3. Approximation of mass assignments

Mass assignments have the disadvantage that they are difficult to visualize, and they may require considerable amounts of storage space (which may become problematic at the scale of billions of Flickr images). Therefore, there is an interest in approximating the mass assignments m^x in a way that alleviates these issues, without losing too much relevant information. Ideally, we want an approximation of the mass assignment as a mapping from $\mathcal{A}_1$ to $[0, 1]$ (or some other scale), as such mappings are easy to visualize, e.g. as a heat map. An obvious candidate would be to use the pignistic probability. However, this still has the disadvantage that a value must be stored for each element from $\mathcal{A}_1$. Here we present an alternative solution, which uses possibility theory [40].

The main idea is to determine a nested family of areas $B_1 \subseteq B_2 \subseteq ... \subseteq B_l \subseteq \mathcal{A}_1$, such that $B_1, ..., B_l$ correspond to increasingly more cautious approximations of the location of the photo x. They can be obtained by applying the procedure from Section 5.2, using a (fixed) set of different thresholds on the required confidence. In this way, all we have to store are the l regions and the corresponding confidence values. To visualize the mass assignment, we can then simply plot these areas, using gray-scale values that depend on the confidence levels. Moreover, the use of a small number of confidence regions also means that standard spatial indexing methods can be used, e.g. when implementing a system that needs to be able to retrieve all photos that are located in a given area with a predefined confidence.

Note that the nested family $B_1 \subseteq ... \subseteq B_l$ can be seen as a mapping π from $\mathcal{A}_1$ to $[0, 1]$:

$$\pi(a) = \max_{i=1}^{l} \min\left(B_{l+1-i}(a), \frac{i}{l}\right) \tag{19}$$

where we identify the sets B_i with their characteristic mapping for the ease of presentation, i.e. we have $B_i(a) = 1$ iff $a \in B_i$ and $B_i(a) = 0$ otherwise. The mapping π is called a possibility distribution [40], and $\pi(a)$ the degree of possibility that the correct area is a. Where probability distribtutions can model variability but not epistemic uncertainty, possibility distributions can model epistemic uncertainty but not variability. A situation of complete ignorance can be modeled as $\pi(a) = 1$ for all $a \in \mathcal{A}_1$, whereas in a completely informed situation we have $\pi(a) = 1$ for exactly one $a \in \mathcal{A}_1$ and $\pi(a) = 0$ for all other areas. In general, the degree $\pi(a)$ is interpreted as the degree to which one would be surprised to learn that a is the real value of the considered variable, an interpretation which at least goes back to the work of Shackle [41].

Like probability distributions, possibility distributions also correspond to a special case of belief functions. To clarify this link, first note that with each possibility distribution π, two uncertainty measures Π and N can be associated, defined (in a universe U) by

$$\Pi(X) = \sup_{u \in U} \pi(u)$$
$$N(X) = 1 - \Pi(U \setminus X)$$

Intuitively, $\Pi(X)$ corresponds to degree to which it is consistent with our beliefs to assume that the correct value is among those in X, whereas $N(X)$ corresponds to the degree to which this is implied by our beliefs. Now, let m be a mass assignment whose focal elements constitute a nested family of sets: $\emptyset \subset X_1 \subset X_2 \subset ... \subset X_l \subseteq U$. With the mass assignment m we can associate the possibility distribution π defined by $\pi(x) = \sum_{x \in X_i} m(X_i) = Pl(\{x\})$ [42]. Then we have that for any $X \subseteq U$, it holds that $Bel(X) = N(X)$ and $Pl(X) = \Pi(X)$. In general, a mass assignment m can be approximated by a possibility distribution in different ways. One approach is to still define $\pi(x) = Pl(\{x\})$, in which case π is called the contour function of m. A second approach is to use a predefined family of nested sets, as we did in (19).

Possibility distributions are not only useful for visualization. Their graded nature makes them suitable representations for modeling the boundaries of vague vernacular geographic regions [9, 7, 10]. Such flexible boundaries could be obtained by georeferencing a "virtual photo" whose tags are the name of the region, and the city and country in which it occurs. In fact, similar ideas have already been proposed, but without making the links with possibility theory explicit. For instance, [43] proposes a method in which spatial terms occurring on a web page are converted into polygons and the overall relevance of that web page w.r.t. a given location is calculated based on the number of polygons in which that location appears. However, seeing these polygons as the focal elements of a mass assignment, this corresponds exactly to determining the degree of plausibility of the considered location.

6. Evaluation

As the baseline of our experiments, we will consider the raw probabilities that are produced by the language models (i.e. the right-hand side of (1)). This baseline technique has been the basis of a system with which we participated in the 2010 and 2011 editions of the MediaEval Placing Task competition, where it was shown to compare favorably against other georeferencing techniques [16, 44]. This result confirms and strengthens earlier support for using language models in this task [14].

The techniques that we propose in this paper aim at improving the baseline in two different ways. First, by combining evidence from different granularity levels, we can hope that better informed decisions can be made about which is the most likely area at a given granularity level (as was illustrated in Example 1). This means that the Dempster-Shafer based techniques should allow us to obtain a higher overall accuracy. Second, by calibrating the probabilities and by combining evidence from different granularity levels, we can also hope that more reliable confidence estimates are obtained. Here, we are not interested in improving the overall accuracy, but in determining which of the photos we can georeference in an accurate way. This is important from an application point of view, as clearly not all photos have sufficiently descriptive tags to allow meaningful coordinates to be found. What we need then, is a way of selecting a maximal set of photos such that at least, say 95% of these photos is correctly georeferenced. Both goals are more or less independent, in the sense that techniques which succeed in improving the overall accuracy may not necessarily be best suited to determine photos that are likely to be georeferenced correctly. In the following, we analyze both goals.

6.1. *Overall accuracy*

Considering the first goal, Table 4 summarizes the overall accuracies that are obtained at each of the 5 considered granularity levels, for each of the considered methods. The line *Probability - Raw* contains the results that are obtained when using the raw probabilities provided by the language models, and the line *Probability – Calibrated* contains the results of using the PAV algorithm to calibrate these probabilities as explained in Section 3.2. As can be seen from the table, calibration leads to a minor (but consistent) improvement in accuracy. This is somewhat surprising, as the aim of calibration was not to obtain better predictions but to obtain better confidence scores (in relation to the second goal). It should be emphasized here that we used a separate set for calibrating the probabilities, which did neither overlap with the test set nor with the training set that was used for training the language models. As such, in applying the PAV algorithm, we may implicitly take the observation into account that the probabilities for some areas are systematically too large or too small, and thus influence which area is considered to be the most plausible one for a given photo.

Nonetheless, the improvement in accuracy that is witnessed by applying the PAV algorithm is rather small. One of the main reasons for applying this technique was that accurate probability estimates were needed by the Dempster-Shafer method, to compare the probabilities from language

Table 4: Accuracy of the predictions at each of the five considered granularity levels.

	Accuracy				
	50	250	500	1000	2000
Probability – Raw	82.08	67.43	61.90	57.46	51.14
Probability – Calibrated	82.65	68.14	62.56	58.02	51.97
Belief – Dempster	84.30	72.38	67.95	63.29	**53.28**
Plausibility – Dempster	84.30	72.41	67.91	62.90	52.66
Pign. Prob. – Dempster	**84.33**	**72.44**	**68.20**	**63.41**	53.27
Belief – Yager	84.30	72.38	67.95	63.29	**53.28**
Plausibility – Yager	84.30	72.41	67.91	62.90	52.66
Pign. Prob. – Yager	82.62	71.50	67.54	63.25	53.27
Belief – Dubois-Prade	84.17	71.89	67.52	62.97	53.11
Plausibility – Dubois-Prade	84.15	72.03	67.44	62.38	52.05
Pign. Prob. – Dubois-Prade	83.67	71.17	66.87	62.29	52.90

Table 5: Percentage of photos for which the found location was within 1km, 5km, 10km, 50km, 100km and 1000 km of the true location, and the median distance on the error (in kilometers), when using the raw probabilities (full test set).

Gran.	1	5	10	50	100	1000	10000	Median
50	00.15	00.54	00.89	03.13	05.49	62.50	97.37	732.80
250	01.10	06.48	09.53	20.69	31.03	78.02	96.76	188.97
500	02.39	11.34	16.54	33.58	47.66	76.97	96.49	110.46
1000	04.60	17.69	24.04	47.39	56.90	76.52	96.41	59.34
2000	09.91	25.35	32.98	52.47	59.56	76.28	96.30	40.61

Table 6: Percentage of photos for which the found location was within 1km, 5km, 10km, 50km, 100km and 1000 km of the true location, and the median distance on the error (in kilometers), when using pignistic probabilities obtained from Dempster's combination rule (full test set).

Gran.	1	5	10	50	100	1000	10000	Median
50	00.15	00.54	00.93	03.20	05.59	63.27	97.57	728.47
250	01.18	06.84	10.04	21.86	32.65	80.89	97.43	174.62
500	02.56	12.22	17.84	36.29	51.49	80.91	97.39	94.70
1000	04.91	18.90	25.77	51.68	61.66	80.66	97.20	45.81
2000	09.95	25.43	33.16	53.65	61.46	79.49	96.70	37.62

models at different granularity levels. Table 4 shows the results that were obtained using three different combination rules (Dempster (8)–(9), Yager (12)–(14), and Dubois-Prade (15)), each time considering three different decision rules (based on belief (16), plausibility (17) and pignistic probability (18)). For each of these 9 configurations, a clear improvement is found over the results of the (calibrated) language model probabilities. The difference is most pronounced at the intermediate granularity levels. It appears that the language models' results for the coarsest granularity level are difficult to improve, as (i) most of the incorrectly georeferenced photos are simply not tagged in a sufficiently descriptive way (i.e. the language model probabilities are nearly optimal), and (ii) there is little evidence to be found at the finer granularity levels to help make a decision at the coarsest level. Note that, at the coarsest level, there are only 50 clusters for the entire world, hence classification here basically amounts to finding the right country for a photo. Conversely, the results for the finest granularity level are also difficult to improve, which may be due to the same two reasons. While many photos contain tags that allow us to pinpoint the right city, finer predictions can often not be made. Moreover, evidence from the coarser granularity levels is usually not sufficiently specific to help make this decision. For the three intermediate granularity levels, larger improvements are obtained.

Comparing the three combination rules in Table 4, we notice that Dempster and Yager produce identical results when either belief or plausibility is used as the decision rule. This was to be expected, since Dempster's and Yager's rules only differ in how the mass of the empty set is redistributed. As a result, the ranking of areas according to their degree of belief or degree of plausibility is unaltered. When using pignistic probability, however, some changes may occur. Similarly, when using Dubois and Prade's combination rule, additional focal elements are introduced, which may affect which area is considered to be the most plausible one at a given granularity level. While Dubois and Prade's rule leads to similar results as Dempster's and Yager's, results of the latter combination rules are slightly better. Concerning the decision rule, pignistic probability was found to be slightly better when using Dempster's rule, while belief was slightly better in combination with Dubois and Prade's rule. In most cases, using belief was also the best choice in combination with Yager's rule.

Tables 5 and 6 provide an overview of these results in terms of error distance between the estimated location for a photo and its true location. These tables confirm the main conclusion from Table 4: the use of Dempster-Shafer theory leads to a moderate, but consistent improvement over the baseline, with larger gains to be found at the intermediate levels. Tables 7 and 8 provide an overview of the results for the same two methods in terms of accuracy at a city level, local administrative unit (LAU) level and country level. The ground truth information for this evaluation was obtained by feeding the real coordinates to the Google Geocoding API [45]. The "Admin" category in the tables corresponds to the *administrative_area_level_1* information provided by Google (i.e. the first-level administrative divisions in a country, such as provinces or states). As the administrative information could not be determined for several photos in the test set and the medoids of several clusters, for the evaluation in Tables 7 and 8, we have excluded all photos from the test set for which we could not determine the relevant information, as well as all photos which were assigned to a cluster, by any of the methods at any of the granularity levels, with a medoid for which we could not determine the relevant information. This has led to a reduced test set of 32 748 test items (65.49 % of the original test set). The results in Table 7 and 8 are thus mainly meaningful relative to each other.

Table 7: Percentage of photos for which the found location was within the correct city, administrative region and country, when using the raw probabilities. (restricted test set)

Granularity	City	Admin	Country
50	01.29	14.09	48.16
250	12.36	39.44	76.75
500	21.73	52.83	81.93
1000	27.38	59.48	82.45
2000	32.36	63.97	81.41

Table 8: Percentage of photos for which the found location was within the correct city, administrative region and country, when using pignistic probabilities obtained from Dempster's combination rule (restricted test set).

Granularity	City	Admin	Country
50	01.32	14.34	48.64
250	13.04	41.15	77.33
500	23.48	56.49	84.88
1000	29.26	63.77	86.15
2000	32.61	65.85	86.01

To gain a better insight into why the use of Dempster-Shafer theory leads to improved results, we discuss two concrete examples of photos in the test set, where it was needed to look at evidence from other granularity levels to find the correct location. Consider the upper example in Table 9. All the tags mentioned in the example were retained at the coarsest granularity level (50 areas). Using the raw probability, this photo was estimated to be in a cluster that represents the North-East of the US, whereas using the pignistic probability correctly assigned it to a cluster in Western Europe. To find the location of this photos, mainly the tags *Colchester* and *zoo* are important, as they clearly suggest that the photo was taken in *Colchester zoo* in the UK. However, at the coarse granularity level of 50 areas, the tag *zoo* will have very little discriminative power, as most of the 50 clusters will contain the location of several zoos. The term *Colchester*, however, will help to find the right cluster, although it leads to an ambiguity: the area containing the UK will

Table 9: Example assignments of test photos by using *Probability – Raw* and *Pign. Prob. – Dempster.*

animal zoo wildlife straw colchester mandrill forage foraging

	True location		Estimated location (50 areas)	
Probability – Raw	51.8619	0.8267	40.9441	78.9678
Pign. Prob. – Dempster	51.8619	0.8267	51.2189	4.4012

sandals korea toji

	True location		Estimated location (2000 areas)	
Probability – Raw	35.1293	127.7567	30.0665	-51.2359
Pign. Prob. – Dempster	35.1293	127.7567	35.2601	128.7594

definitely contain several occurrences of this tag, but this is also true for the cluster containing the North-East of the US (which contains places called Colchester in VT, CT, NY and IL). Without any further help to make the decision, the baseline system incorrectly assigned it the photo to the US. When looking at the granularity level of 2000 levels, on the other hand, the location becomes obvious: there is only one cluster which a substantial number of occurrences of both *zoo* and *Colchester* (none of the places called Colchester in the North-East of the US has a zoo). The Dempster-Shafer based methods are able to use this evidence from the 2000 area level to find the correct cluster at the 50 area level.

The lower example in Table 9 is an illustration of the opposite case, where coarser levels can help us to correctly assign a photo to a cluster at the finer-grained levels. The example concerns a photo taken in South-Korea, which was mistakenly estimated to be in southern Brazil by the baseline, despite the occurrence of the toponym tag *korea* and the apparent lack of ambiguity. After inspecting the training data, we found that the error was due to one cluster (at the 2000 area level) in Brazil with a disproportionate number of occurrences of the tag *toji*, caused by a large number of photos of one user's cat named toji. This tag turned out to be more discriminative than the term *korea* (which occurs in several clusters within Korea), leading to an incorrect decision. At the coarser levels, however, the tag *korea* becomes very discriminative while the tag *toji* loses its importance. In this way, the Dempster-Shafer based methods can using the evidence from the coarser levels to avoid making the mistake at the finest level.

6.2. *Confidence score reliability*

We now turn to the second goal of trying to identify those photos for which the predicted area is most likely to be correct. Being able to identify the "easy" cases from the "hard" cases assists an application in determining the action to be taken: if the application has high confidence in its estimation, it will georeference the photo at hand. Else, when confidence is low, the application does not suggest the location of the photo. Preferably, we want to have a system that is highly accurate in recognizing the "easy" cases. Another way of viewing this task is that we should determine for each photo individually, at which granularity level it is best classified (cfr. the use cases that were outlined in Sections 5.2 and 5.3). To illustrate this idea, consider the following examples: In the case of a photo tagged with *water wales boats bay cardiff cardiffbay barrage*, the tags unambiguously identify a specific location at a fine granularity level, hence the system should be quite confident in georeferencing such a photo. Secondly, a photo tagged with *france* will not yield a likely locations at the finest granularity level, but at a coarser level of granularity (say, a level at the scale of the European countries), it should become very confident that the photo was taken in the area covering France. Lastly, a photo tagged only with *birthday abby* clearly is a hard case, which is impossible to georeference even for a human assessor. To determine about which photos' predictions we are confident enough, we can put some threshold on the considered confidence scores. These confidence scores may be probabilities (raw or calibrated), degrees of belief, degrees of plausibility, and pignistic probabilities. In the last three cases, the confidence scores may be evaluated w.r.t. the combined mass assignments resulting from either of the three considered combination rules. The choice of the threshold value allows us to tune the trade-off between having a higher accuracy and having more photos georeferenced.

To assess which method provides the most useful confidence scores, in Tables 10–13 we show how many photos can be georeferenced when a given level of accuracy is imposed. Comparing the performance of the raw and calibrated probabilities in Table 10, we can see that the calibrated probabilities perform consistently better, with the improvement being largest for the finest granularity levels and the highest accuracy thresholds. For instance, at the finest granularity level (2000 clusters), 24% of the photos can be georeferenced with 95% accuracy using the calibrated probabilities, as opposed to only 14% when using the raw probabilities. This means that e.g. if we allow the pignistic probability method to choose 24% of the pho-

Table 10: Percentage of photos that can be classified at each level of granularity when a fixed accuracy level is imposed (using the probabilities from the language models).

	Acc. (%)	Percentage of photos				
		50	250	500	1000	2000
Probability – Raw	75	100	94	78	72	62
	80	100	78	70	64	52
	85	94	72	62	56	42
	90	84	62	52	44	30
	95	72	46	34	28	14
Probability – Calibrated	75	100	88	78	72	62
	80	100	80	72	66	54
	85	94	72	64	58	46
	90	84	62	54	48	36
	95	74	44	36	32	24

Table 11: Percentage of photos that can be classified at each level of granularity when a fixed accuracy level is imposed (using Dempster's rule of combination to combine evidence from different granularity levels).

	Acc. (%)	Percentage of photos				
		50	250	500	1000	2000
Plausbility – Dempster	75	100	96	88	80	60
	80	100	88	80	74	52
	85	98	80	74	66	44
	90	88	72	66	58	34
	95	78	62	56	46	0
Belief – Dempster	75	100	96	88	82	66
	80	100	88	82	76	56
	85	98	80	74	68	48
	90	88	72	66	60	36
	95	78	62	56	46	22
Pign. Prob. – Dempster	75	100	96	90	82	66
	80	100	88	82	76	58
	85	98	80	74	68	48
	90	88	72	66	60	36
	95	78	62	56	46	22

tos, which it thinks are easiest to georeference, and only require it to georeference these 24%, it will assign a correct cluster to 95% of them. To interpret the meaning of these results, consider an application which suggests a location to users uploading and tagging photos on Flickr, as a way to encourage more people to reveal location-based information, e.g. by showing a map of where the system think the photo was taken. As users will be annoyed if the system is wrong too often, we may need to get it right in, say, 95% in the cases. As this is not possible, by any method, in general (due to there being too many photos with tags that do not reveal any location at all), we can only accomplish this by only making a suggestion to the user when we are confident enough that it is correct. So, given the results in Table 10, and assuming that we want 95% of the suggestions we make to be correct, we can only suggest a location in 14% of the cases when using raw probabilities, while we can do it in 24% of the cases using calibrated probabilities. Note that this improvement is mainly due to the better capabilities of the latter method of distinguishing easy cases from hard cases, rather than being (much) better at the actual task of georeferencing.

In Table 11, the results of using Dempster's combination rule are presented. A marked improvement over the results from Table 10 can be seen, which is largest at the intermediary granularity levels and the higher accuracy thresholds. For instance, at the third granularity level (500 clusters), using Dempster's combination rule and the pignistic probability decision rule, 56% of the photos can be georeferenced with 95% accuracy, as opposed to only 36% for the calibrated probabilities and 34% for the raw probabilities. The best results are found when using pignistic probabilities, although the results for degrees of belief are almost identical and the results for degrees of plausibility are similar in most of the cases. Tables 12 and 13 show the results for respectively Yager's rule and Dubois and Prade's rule. Overall, we may conclude that Dempster's rule provides the best results, followed by Dubois and Prade's rule, and then Yager's rule.

A graphical view on the relation between the number

Table 12: Percentage of photos that can be classified at each level of granularity when a fixed accuracy level is imposed (using Yager's rule of combination to combine evidence from different granularity levels).

	Acc. (%)	Percentage of photos				
		50	250	500	1000	2000
Plausbility – Yager	75	100	96	88	80	56
	80	100	88	80	72	48
	85	94	80	74	64	42
	90	84	72	64	56	34
	95	74	62	54	46	0
Belief – Yager	75	100	94	88	82	64
	80	100	86	80	74	56
	85	96	78	74	68	46
	90	84	70	64	58	36
	95	64	58	54	46	24
Pign. Prob. – Yager	75	100	94	86	82	64
	80	100	86	80	74	56
	85	94	78	72	66	46
	90	84	70	64	58	36
	95	72	58	54	46	24

of photos that can be georeferenced and the resulting level of accuracy is provided in Figures 4–13. These figures provide a clear view of the trade-off in applications between georeferencing a larger percentage of all photos and maintaining a higher accuracy. All the photos in the test set are ranked according to their confidence score (i.e. pignistic probability, belief, or plausibility). As mentioned in the introduction of Section 6.2, all the photos whose confidence scores are above a certain threshold would be considered as the "easy" cases. Specifically, for each number of photos n on the X-axis, the accuracy of the n photos with the highest values for this confidence score is reported. First, Figures 4–8 compare the performance of the three combination rules (using pignistic probabilities), each time also displaying the results for raw and calibrated probabilities. What is particularly noticeable is that the use of calibrated probabilites does not improve the raw probabilities at all for the coarser granularity levels, while at the finest granularity level (Figure 8), the calibrated probabilities are essentially as good as the outcome of the Dempster-Shafer based approaches. Overall, we can also see that the combination operator being used does not affect the performance in a crucial way. Figures 9–13 compare the performance of the three decision rules (using Dempster's rule of combination). Here we can clearly see that using degrees of belief or using pignistic probabilities does not substantially change the result. Regarding degrees of plausibility, the results are somewhat mixed. At the finer granularity levels and the left-most part of the graphs, plausibility degrees perform even worse than the baseline. In some sense, this is not surprising, as the idea of plausibility degrees is somewhat at odds with the task of finding those photos for which sufficient location evidence can be found. Indeed, plausibility degrees reflect the compatibility of a given element with available evidence, rather than an amount of support.

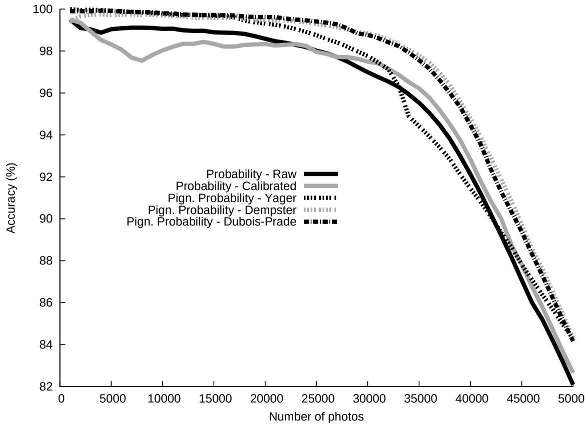

Figure 4: Comparing the trade-off between number of georeferenced photos and accuracy for different combination rules, using pignistic probability and 50 clusters.

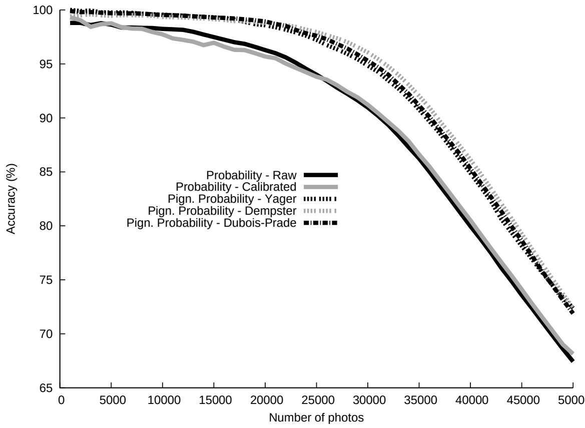

Figure 5: Comparing the trade-off between number of georeferenced photos and accuracy for different combination rules, using pignistic probability and 250 clusters.

Table 13: Percentage of photos that can be classified at each level of granularity when a fixed accuracy level is imposed (using Dubois and Prade's rule of combination to combine evidence from different granularity levels).

	Acc. (%)	Percentage of photos				
		50	250	500	1000	2000
Plausbility – Dubois-Prade	75	100	94	88	80	58
	80	100	86	80	72	50
	85	98	80	72	66	42
	90	88	72	64	56	34
	95	78	62	54	46	0
Belief – Dubois-Prade	75	100	94	88	82	64
	80	100	86	80	74	56
	85	98	80	74	68	48
	90	88	72	66	58	36
	95	78	60	54	46	22
Pign. Prob. – Dubois-Prade	75	100	94	86	80	64
	80	100	86	80	74	56
	85	96	78	72	66	48
	90	88	70	64	56	36
	95	76	58	52	44	22

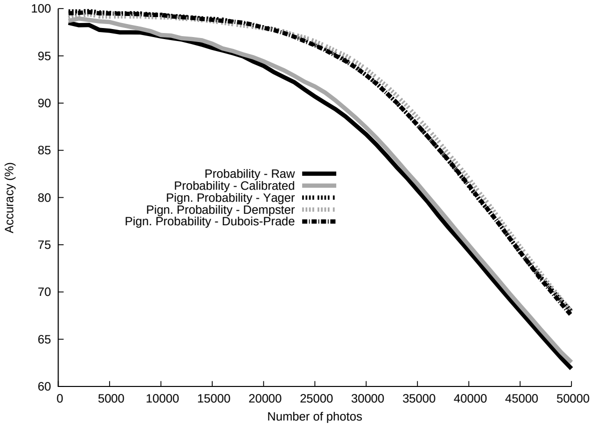

Figure 6: Comparing the trade-off between number of georeferenced photos and accuracy for different combination rules, using pignistic probability and 500 clusters.

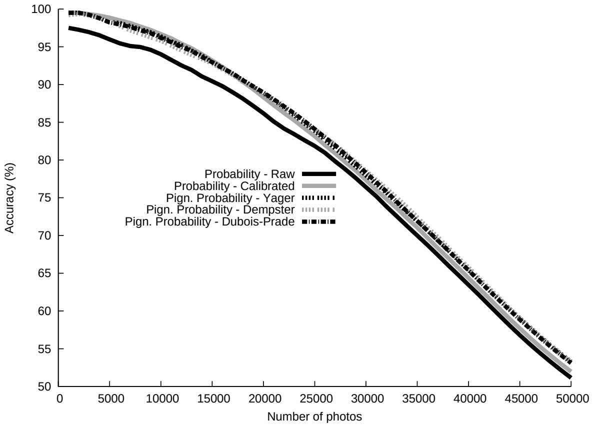

Figure 8: Comparing the trade-off between number of georeferenced photos and accuracy for different combination rules, using pignistic probability and 2000 clusters.

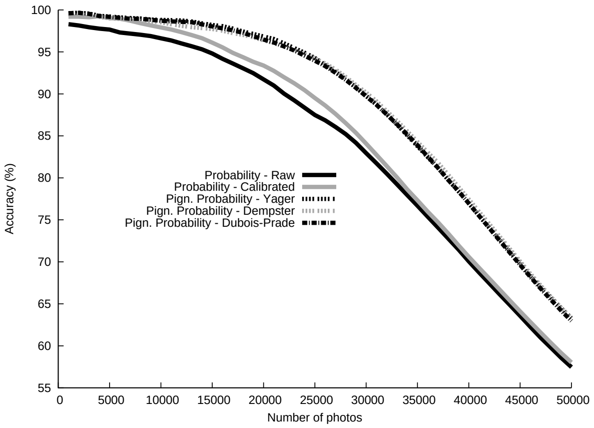

Figure 7: Comparing the trade-off between number of georeferenced photos and accuracy for different combination rules, using pignistic probability and 1000 clusters.

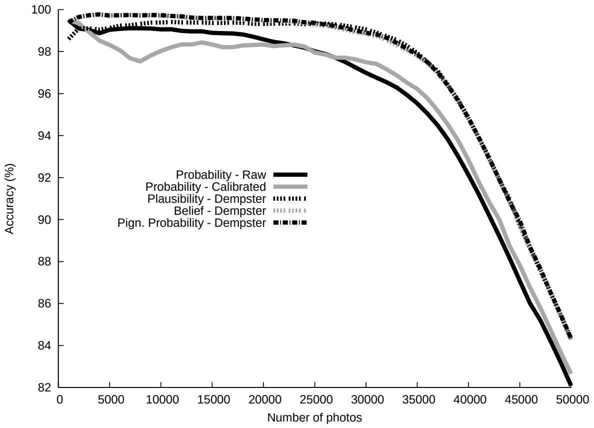

Figure 9: Comparing the trade-off between number of georeferenced photos and accuracy for different decision rules, using Dempster's combination rule and 50 clusters.

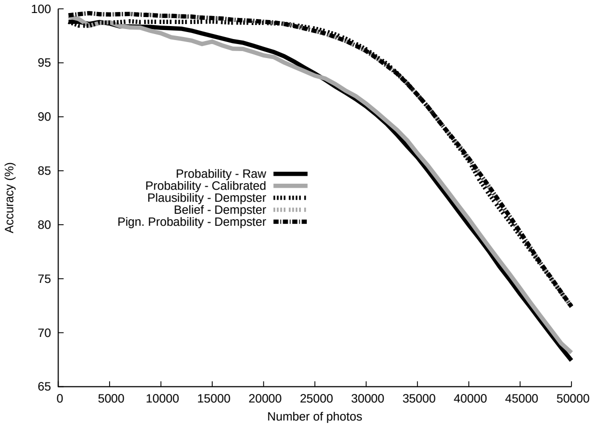

Figure 10: Comparing the trade-off between number of georeferenced photos and accuracy for different decision rules, using Dempster's combination rule and 250 clusters.

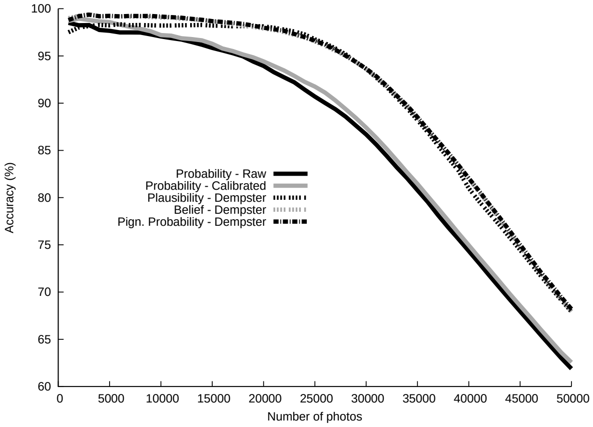

Figure 11: Comparing the trade-off between number of georeferenced photos and accuracy for different decision rules, using Dempster's combination rule and 500 clusters.

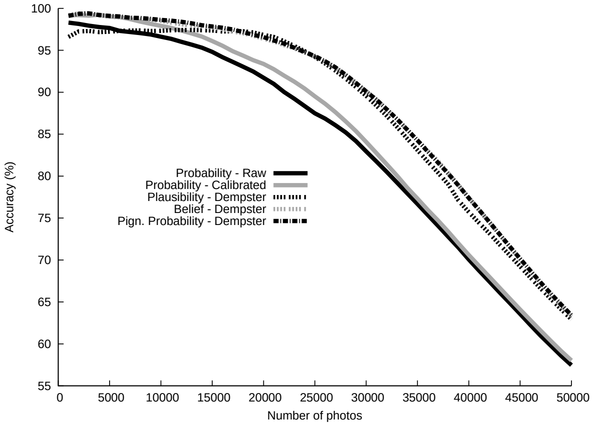

Figure 12: Comparing the trade-off between number of georeferenced photos and accuracy for different decision rules, using Dempster's combination rule and 1000 clusters.

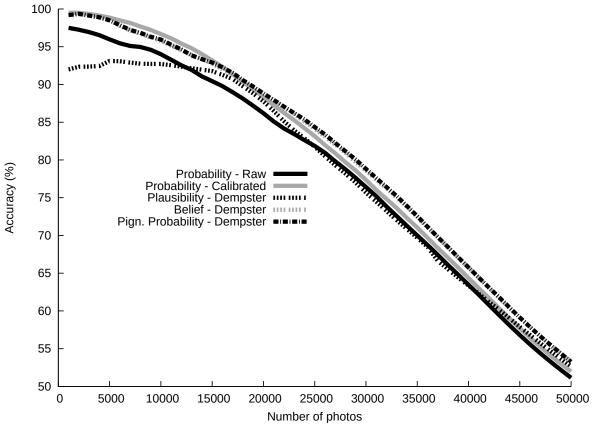

Figure 13: Comparing the trade-off between number of georeferenced photos and accuracy for different decision rules, using Dempster's combination rule and 2000 clusters.

7. Related work

7.1. Finding locations of resources

The task of deriving geographic coordinates for photos has recently gained in popularity (see e.g. [16]). However, to the best of our knowledge, the idea of combining evidence from different granularity levels and the related problem of finding the most appropriate granularity level for a given photo have not been previously considered. In the context of geographic information systems, on the other hand, it is well known that different *scales* may yield different effects on the spatial and thematic resolution of geographic data [12] (e.g. monitoring the earth's surface using satellites with different resolutions).

Most existing approaches are based on clustering, in one way or another, to convert the task into a classification problem. For instance, in [46] target locations are determined using mean shift clustering, a non-parametric clustering technique from the field of image segmentation. The advantage of this clustering method is that an optimal number of clusters is determined automatically, requiring only an estimate of the scale of interest. Specifically, to find good locations, the difference is calculated between the density of photos at a given location and a weighted mean of the densities in the area surrounding that location. To assign locations to new images, both visual (keypoints) and textual (tags) features were used. Experiments were carried out on a sample of over 30 million images, using both Bayesian classifiers and linear support vector machines, with slightly better results for the latter. Two different resolutions were considered corresponding to approximately 100 km (finding the correct metropolitan area) and 100 m (finding the correct landmark). It was found that visual features, when combined with textual features, substantially improve accuracy in the case of landmarks. In [47], an approach is presented which is based purely on visual features. For each new photo, the 120 most similar photos with known coordinates are determined. This weighted set of 120 locations is then interpreted as an estimate of a probability distribution, whose mode is determined using mean-shift clustering. The resulting value is used as prediction of the image's location.

The idea that when georeferencing images, the spatial distribution of the classes (areas) could be utilized to improve accuracy has already been suggested in [14]. Their starting point is that typically not only the correct area will receive a high probability, but also the areas surrounding the correct area. Indeed, the expected distribution of tags in these areas will typically be quite similar. Hence, if some area a receives a high score, and all of the areas surrounding a also receive a relatively high score, we can be more confident in a being approximately correct than when all the areas surrounding a receive a low score. Motivated by this intuition, [14] proposes to smooth $P(a|x)$ as follows (using a uniform prior):

$$P^*(a|x) \propto \alpha P(x|a) + (1-\alpha) \cdot \sum_{b \in neigh_d(a)} \frac{P(x|b)}{(2d+1)^2 - 1}$$

where $d > 0$ and $neigh_d(a)$ is the set of all areas that are within distance d of a.

Some Flickr tags are intuitively more important than others in determining the location of a photo. Toponyms in particular are by definition indicative of geographic location. One way of recognizing toponyms is by looking for so-called comma-groups. These are groups of words that are comma-separated, e.g *San Francisco, California, USA*. In this example, there is a clear relationship between the comma-separated values, as San Francisco is a city, located in the state of California, which is in turn one of the states of the USA. As a result, resolution of the toponyms represented by this group reveals an unambiguous geographical reference. Resolution of such comma-groups has been studied by Lieberman in [48].

In addition to georeferencing Flickr photos, several authors have recently focused on finding the location of other web resources such as Twitter posts or Wikipedia pages. For instance, in [49], a probabilistic framework based on maximum likelihood estimation was used to estimate the location of users based on the content of their tweets. In particular, a generative probabilistic model proposed in [50] is used to determine words with a geographic scope within a tweet, and a form of neighborhood smoothing is employed to refine the estimations. For 51% of the users, a location was obtained that is within a 100 mile radius of their true location. Next, [51] looked into georeferencing Wikipedia articles as well as Twitter posts. After laying out a grid over the earths surface (in a way similar to [1]), for each grid cell a generative language model is estimated. To assign a test item to a grid cell, its Kullback-Leibler divergence with the language models of each of the cells is calculated. In [52], it was shown how Wikipedia pages can be georeferenced using language models that are trained from Flickr, taking the view that the relative sparsity of georeferenced Wikipedia pages does not allow for sufficiently accurate language models to be trained, especially at finer levels of granularity.

Interestingly, some recent language modeling approaches have combined the idea of topic models with location-dependent language models. For instance, [54] proposes geographic topic models with the aim of simultaneously capturing linguistic variation across different regions and different topics.

7.2. Using locations of resources

When available, the coordinates of a photo may be used in various ways. In [55], for instance, coordinates of tagged photos are used to find representative textual descriptions of different areas of the world. These descriptions are then put on a map to assist users in finding images that were taken in a given location of interest. Their approach is based on spatially clustering a set of geotagged Flickr images, using k-means, and then relying on (an adaptation of) tf-idf weighting to find the most prominent tags of a given area. Similarly, [56] looks at the problem of suggesting useful tags, based on available coordinates. The relevance of a given tag is measured in terms of the number of users that have used it to describe photos located within a certain radius of the current photo's coordinates. A refinement of this method only looks at tags that occur with visually similar photos, which is shown to improve the quality of the proposed tags. Some authors have looked at using geographic information to help diversify image retrieval results [57, 58]. Finally, in [53] GeoSR is presented as a way of measuring the semantic relatedness of Wikipedia articles based on their geographic context, allowing users to explore information in Wikipedia that is relevant to a particular location.

Geotagged photos are also useful from a geographic perspective, to better understand how people refer to places, and overcome the limitations and/or costs of existing mapping techniques [59]. For instance, by analyzing the tags of georeferenced photos, Hollenstein [60] found that the city toponym was by far the most essential reference type for specific locations. Moreover, [60] provides evidence suggesting that the average user has a rather distinct idea of specific places, their location and extent. Despite this tagging behaviour, Hollenstein concluded that the data available in the Flickr database meets the requirements to generate spatial footprints at a sub-city level. Finding such footprints for non-administrative regions (i.e. regions without officially defined boundaries) using georeferenced resources has also been adressed in [9] and [6]. Another problem of interest is the automated discovery of which names (or tags) correspond to places. Especially for vernacular place names, which typically do not appear in gazetteers, collaborative tagging-based systems may be a rich source of information. In [61], methods based on burst-analysis are proposed for extracting place names from Flickr. Finally, note that to some extent, even without geographic coordinates, ontologies, and in particular ontologies of places may be derived from Flickr tags [62].

7.3. Evidence theory

Various authors have investigated the use of Dempster-Shafer theory for combining the results of different classifiers [63, 64, 65, 66]. However, the aim of using Dempster-Shafer theory in this context is quite different from our aim in this paper. Specifically, these methods mainly use Dempster-Shafer theory for its ability to represent partial ignorance. For instance, if a given classifier assigns a probability p_i to each class c_i, a belief function may be constructed by choosing $m(\{c_i\}) = f_i$ for some $f_i < p_i$, and $m(C) = 1 - \sum_i f_i$, for $C = \{c_1, ..., c_n\}$ the set of all classes. The value $1 - \sum_i f_i$ can then intuitively be interpreted in terms of confidence in the associated classifier. Note also that all focal elements are then either singletons or the universe, which makes Dempster-Shafer theory sufficiently scalable to deal with large numbers of classes, although sometimes focal elements of the form $C \setminus \{c_i\}$ are also used.

Dempster-Shafer theory has also been widely considered for dealing with the imperfection of real-world geographic information; [67] provides a survey on works using Dempster-Shafer theory in a GIS setting. More generally, we refer to [68] for an overview of different frameworks for handling uncertainty, applied to spatial information.

8. Conclusions

We have proposed an approach to georeferencing Flickr photos which combines the evidence provided by different language models using Dempster-Shafer evidence theory. As these language models were trained at different granularity levels, they provide complementary views on the georeferencing process, and implicitly add a spatial dimension to the language models.

The core idea of our approach is to see a probability distribution over coarse areas as a probability distribution over sets of fine-grained areas. Noting that this latter probability distribution corresponds to the notion of a mass assignment from Dempster-Shafer theory, we can connect to the vast amount of work that has already been done on combining evidence. In particular, we have experimented with three well-known combination rules, due to Dempster, Yager, and Dubois and Prade respectively.

After the evidence from the language models has been combined, we end up with a mass assignment that summarizes all available evidence about the location of a given photo. This mass assignment may then be used in different ways: we may use it to select the most likely area at a given granularity level, we may determine the smallest area that contains the true location of the photo with a predefined certainty, or we may simply visualize the evidence after approximating the mass assignment to a possibility distribution. In our experiments, we have focused on the first two of these tasks, as the quality of visual representations is difficult to quantify. In both cases, we have found that our evidence-based approach considerably improves the performance of individual language models.

References

[1] C. B. Jones, A. I. Abdelmoty, D. Finch, G. Fu, S. Vaid, The SPIRIT spatial search engine: Architecture, ontologies and spatial indexing, in: Proceedings of the Third International Conference on Geographic Information Science, 2004, pp. 125–139.

[2] O. Van Laere, S. Schockaert, B. Dhoedt, Towards automated georeferencing of flickr photos, in: Proceedings of the 6th Workshop on Geographic Information Retrieval, 2010, pp. 5:1–5:7.

[3] L. Hollenstein, R. Purves, Exploring place through user-generated content: Using Flickr to describe city cores, Journal of Spatial Information Science 1 (1) (2010) 21–48.

[4] A. Popescu, G. Grefenstette, H. Bouamor, Mining a multilingual geographical gazetteer from the web, in: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, 2009, pp. 58–65.

[5] C. Keßler, P. Maué, J. Heuer, T. Bartoschek, Bottom-up gazetteers: Learning from the implicit semantics of geotags, in: Proceedings of the 3rd International Conference on Geospatial Semantics, 2009, pp. 83–102.

[6] F. Wilske, Approximation of neighborhood boundaries using collaborative tagging systems, in: Proceedings of the GI-Days, 2008, pp. 179–187.

[7] F. A. Twaroch, C. B. Jones, A. I. Abdelmoty, Acquisition of a vernacular gazetteer from web sources, in: Proceedings of the First International Workshop on Location and the Web, 2008, pp. 61–64.

[8] I. Holt, J. Green, Social networks as a future geographical data source, in: Proceedings of the W3C Workshop on the Future of Social Networking, 2009.

[9] S. Schockaert, M. De Cock, Neighborhood restrictions in geographic IR, in: Proceedings of the 30th Annual International ACM SIGIR Conference, 2007, pp. 167–174.

[10] C. B. Jones, R. S. Purves, P. D. Clough, H. Joho, Modelling vague places with knowledge from the web, International Journal of Geographical Information Science 22 (2008) 1045–1065.

[11] M. F. Goodchild, M. J. Egenhofer, K. K. Kemp, D. M. Mark, E. Sheppard, Introduction to the Varenius project, International Journal of Geographical Information Science 13 (8) (1999) 731–745.

[12] M. F. Goodchild, A geographer looks at spatial information theory, in: Proceedings of the International Conference on Spatial Information Theory, Springer-Verlag, 2001, pp. 1–13.

[13] P. A. Longley, M. F. Goodchild, D. J. Maguire, D. W. Rhind, Geographic Information Systems and Science, John Wiley & Sons, 2005.

[14] P. Serdyukov, V. Murdock, R. van Zwol, Placing Flickr photos on a map, in: Proceedings of the 32nd Annual International ACM SIGIR Conference, 2009, pp. 484–491.

[15] M. L. et al., Automatic tagging and geotagging in video collections and communities, in: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, 2011,, pp. 51:1–51:8.

[16] M. Larson, M. Soleymani, P. Serdyukov, V. Murdock, G. Jones (Eds.), Working Notes of the MediaEval Workshop, 2010.

[17] T. Gruber, Collective knowledge systems: Where the social web meets the semantic web, Journal of Web Semantics 6 (1) (2008) 4 – 13.

[18] C. Becker, C. Bizer, Exploring the geospatial semantic web with DBpedia Mobile, Journal of Web Semantics 7 (4) (2009) 278 – 286.

[19] O. Van Laere, S. Schockaert, B. Dhoedt, Ghent university at the 2010 Placing Task, in: Working Notes of the MediaEval Workshop, 2010.

[20] A. Dempster, A Generalization of Bayesian Inference, Journal of the Royal Statistical Society. Series B (Methodological) 30 (2) (1968) 205–247.

[21] G. Shafer, A mathematical theory of evidence, Princeton University Press, 1976.

[22] O. Van Laere, S. Schockaert, B. Dhoedt, Combining multiresolution evidence for georeferencing Flickr images, in: Proceedings of the 4th International Conference on Scalable Uncertainty Management, 2010, pp. 347–360.

[23] O. Van Laere, S. Schockaert, B. Dhoedt, Finding locations of flickr resources using language models and similarity search, in: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, 2011, pp. 48:1–48:8.

[24] J. Ponte, W. Croft, A language modeling approach to information retrieval, in: Proceedings of the 21st Annual International ACM SIGIR Conference, 1998, pp. 275–281.

[25] C. Zhai, J. Lafferty, A study of smoothing methods for language models applied to information retrieval, ACM Transactions on Information Systems 22 (2) (2004) 179–214.

[26] M. D. Smucker, J. Allan, An investigation of Dirichlet prior smoothing's performance advantage, Tech. Rep. IR-445, University of Massachusetts (2005).

[27] P. Bennett, Assessing the calibration of Naive Bayes' posterior estimates, Tech. Rep. CMU-CS00-155, Carnegie Mellon (2000).

[28] B. Zadrozny, C. Elkan, Obtaining calibrated probability estimates from decision trees and NaiveBayesian classifiers, in: Proceedings of the 18th International Conference on Machine Learning, 2001, pp. 609–616.

[29] B. Zadrozny, C. Elkan, Transforming classifier scores into accurate multiclass probability estimates, in: Proceedings of the 8th ACM SIGKDD International Conference, 2002, pp. 694–699.

[30] M. Ayer, H. Brunk, G. Ewing, W. Reid, E. Silverman, An empirical distribution function for sampling with incomplete information, The Annals of Mathematical Statistics 26 (4) (1955) 641–647.

[31] W. Wilbur, L. Yeganova, W. Kim, The synergy between PAV and AdaBoost, Machine Learning 61 (2005) 71–103.

[32] T. Fawcett, A. Niculescu-Mizil, PAV and the ROC convex hull, Machine Learning 68 (2007) 97–106.

[33] P. Smets, Constructing the pignistic probability function in a context of uncertainty, in: Proceedings of the 5th Annual Conference on Uncertainty in Artificial Intelligence, 1990, pp. 29–40.

[34] D. Dubois, H. Prade, On the unicity of Dempster rule of combination, International Journal of Intelligent Systems 1 (2) (1986) 133–142.

[35] F. Klawonn, E. Schwecke, On the axiomatic justification of Dempster's rule of combination, International Journal of Intelligent Systems 7 (5) (1992) 469–478.

[36] L. A. Zadeh, A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination, AI Magazine 7 (2) (1986) 85–90.

[37] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (2) (1994) 191 – 234.

[38] R. R. Yager, On the Dempster-Shafer framework and new combination rules, Information Sciences 41 (2) (1987) 93 – 137.

[39] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Computational Intelligence 4 (3) (1988) 244–264.

[40] D. Dubois, H. Prade, Possibility theory: an approach to computerized processing of uncertainty, Plenum Press, 1988.

[41] G. Shackle, Decision, Order and Time in Human Affairs, Cambridge University Press, 1961.

[42] D. Dubois, H. Prade, Fuzzy sets, probability and measurement, European Journal of Operational Research 40 (2) (1989) 135–154.

[43] R. Larson, Geographic information retrieval and spatial browsing, GIS and Libraries: Patrons, Maps and Spatial Information (1996) 81–124.

[44] O. Van Laere, S. Schockaert, B. Dhoedt, Ghent university at the 2011 Placing Task, in: Working Notes of the MediaEval Workshop, 2011.

[45] Google Geocoding API [cited December 6th, 2011].
URL `http://code.google.com/apis/maps/documentation/geocoding/`

[46] D. J. Crandall, L. Backstrom, D. Huttenlocher, J. Kleinberg, Mapping the world's photos, in: Proceedings of the 18th International Conference on World Wide Web, 2009, pp. 761–770.

[47] J. H. Hays, A. A. Efros, IM2GPS: Estimating geographic information from a single image, in: Proceedings of the 21st IEEE Compuster Society Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[48] M. D. Lieberman, H. Samet, J. Sankaranayananan, Geotagging: using proximity, sibling, and prominence clues to understand comma groups, in: Proceedings of the 6th Workshop on Geographic Information Retrieval, 2010, pp. 6:1–6:8.

[49] Z. Cheng, J. Caverlee, K. Lee, You are where you tweet: a content-based approach to geo-locating twitter users, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010, pp. 759–768.

[50] L. Backstrom, J. Kleinberg, R. Kumar, J. Novak, Spatial variation in search engine queries, in: Proceedings of the 17th International Conference on World Wide Web, 2008, pp. 357–366.

[51] B. Wing, J. Baldridge, Simple supervised document geolocation with geodesic grids, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 955–964.

[52] C. De Rouck, O. Van Laere, S. Schockaert, B. Dhoedt, Georeferencing Wikipedia pages using language models from Flickr, in: Proceedings of the Terra Cognita 2011 Workshop, 2011, pp. 3–10.

[53] B. Hecht, M. Raubal, GeoSR: Geographically explore semantic relations in world knowledge, in: Proceedings of the 11th AGILE International Conference on Geographic Information Science, 2008, pp. 95–114.

[54] J. Eisenstein, B. O'Connor, N. A. Smith, E. P. Xing, A latent variable model for geographic lexical variation, in: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010, pp. 1277–1287.

[55] S. Ahern, M. Naaman, R. Nair, J. H.-I. Yang, World explorer: visualizing aggregate data from unstructured text in georeferenced collections, in: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, 2007, pp. 1–10.

[56] E. Moxley, J. Kleban, B. Manjunath, Spirittagger: a geo-aware tag suggestion tool mined from Flickr, in: Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, 2008, pp. 24–30.

[57] L. Kennedy, M. Naaman, Generating diverse and representative image search results for landmarks, in: Proceedings of the 17th International Conference on World Wide Web, 2008, pp. 297–306.

[58] A. Popescu, I. Kanellos, Creating visual summaries for geographic regions, in: IR+SN Workshop (at ECIR), 2009.

[59] M. Goodchild, Citizens as sensors: the world of volunteered geography, GeoJournal 69 (2007) 211–221.

[60] L. Hollenstein, Capturing vernacular geography from georeferenced tags, Master's thesis, University of Zurich (2008).

[61] T. Rattenbury, M. Naaman, Methods for extracting place semantics from Flickr tags, ACM Transactions on the Web 3 (1) (2009) 1–30.

[62] P. Schmitz, Inducing ontology from Flickr tags, in: Proceedings of the Collaborative Web Tagging Workshop, 2006, pp. 210–214.

[63] A. Al-Ani, M. Deriche, A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence, Journal of Artificial Intelligence Research 17 (1) (2002) 333–361.

[64] T. Denœux, A k-nearest neighbor classification rule based on Dempster-Shafer theory, IEEE Transactions on Systems, Man, and Cybernetics 25 (5) (1995) 804–813.

[65] G. Rogova, Combining the results of several neural network classifiers, Neural Networks 7 (5) (1994) 777–781.

[66] L. Xu, C. Suen, Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on Systems, Man, and Cybernetics 22 (3) (1992) 418–435.

[67] J. Malpica, M. Alonso, M. Sanz, Dempster-Shafer theory in geographic information systems: a survey, Expert Systems with Applications 32 (1) (2007) 47 – 55.

[68] R. Jeansoulin, O. Papini, H. Prade, S. Schockaert (Eds.), Methods for Handling Imperfect Spatial Information, Studies in Fuzziness and Soft Computing, Springer, 2010.

WEB OF KNOWLEDGE℠ DISCOVERY STARTS HERE THOMSON REUTERS

Sign In | Marked List (0) | My EndNote Web | My ResearcherID | My Citation Alerts | My Saved Searches | Log Out | Help

All Databases | Select a Database | Web of Science | Additional Resources

Search | Author Search | Cited Reference Search | Advanced Search | Search History

Web of Science®

<< Back to results list ◀ Record 12 of 80 ▶ Record from Web of Science®

S·F·X (0) | Save to: ENDNOTE® WEB ENDNOTE® I Wrote These Publications R more options

Georeferencing Flickr photos using language models at different levels of granularity: An evidence based approach

Author(s): Van Laere, O (Van Laere, Olivier)[1]; Schockaert, S (Schockaert, Steven)[2]; Dhoedt, B (Dhoedt, Bart)[1]

Source: JOURNAL OF WEB SEMANTICS Volume: 16 Pages: 17-31 DOI: 10.1016/j.websem.2012.05.005 Published: NOV 2012

Times Cited: 0 (from Web of Science)

Cited References: 67 [view related records] Citation Map

Abstract: The topic of automatically assigning geographic coordinates to Web 2.0 resources based on their tags has recently gained considerable attention. However, the coordinates that are produced by automated techniques are necessarily variable, since not all resources are described by tags that are sufficiently descriptive. Thus there is a need for adaptive techniques that assign locations to photos at the right level of granularity, or, in some cases, even refrain from making any estimations regarding location at all. To this end, we consider the idea of training language models at different levels of granularity, and combining the evidence provided by these language models using Dempster and Shafer's theory of evidence. We provide experimental results which clearly confirm that the increased spatial awareness that is thus gained allows us to make better informed decisions, and moreover increases the overall accuracy of the individual language models. (C) 2012 Elsevier B.V. All rights reserved.

Accession Number: WOS:000314420200002

Document Type: Article

Language: English

Author Keywords: Dempster-Shafer evidence theory; Language models; Georeferencing; Web 2.0; Geographic information retrieval

KeyWords Plus: DEMPSTER-SHAFER THEORY; SEMANTIC WEB; INFORMATION; COMBINATION; CLASSIFIERS; KNOWLEDGE; SYSTEMS; RULE; PAV

Reprint Address: Van Laere, O (reprint author)
⊞ Univ Ghent, Dept Informat Technol, IBBT, Ghent, Belgium.

Addresses:
⊞ [1] Univ Ghent, Dept Informat Technol, IBBT, Ghent, Belgium
⊞ [2] Cardiff Univ, Sch Comp Sci & Informat, Cardiff CF10 3AX, S Glam, Wales

E-mail Addresses: olivier.vanlaere@intec.ugent.be; S.Schockaert@cs.cardiff.ac.uk; bart.dhoedt@intec.ugent.be

Publisher: ELSEVIER SCIENCE BV, PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS

Web of Science Categories: Computer Science, Artificial Intelligence; Computer Science,

Times Cited: 0

Create Citation Alert

This article has been cited 0 times in Web of Knowledge.

Related Records:

Find similar Web of Knowledge records based on shared references.

[view related records]

Cited References: 67

View the bibliography of this record (from Web of Science®).

Citation Map

Additional information

- View the journal's impact factor (in Journal Citation Reports®)

Suggest a correction

If you would like to improve the quality of the data in this record, please suggest a correction.

Information Systems; Computer Science, Software Engineering

Research Areas: Computer Science

IDS Number: 082VJ

ISSN: 1570-8268

<< Back to results list ◀ Record 12 of 80 ▶ Record from Web of Science®

Output Record

Step 1: Select content.

○ Authors, Title, Source
☑ Abstract
● Full Record
□ Cited References

Step 2: Select destination. [Learn about saving to bibliographic software]

ENDNOTE® WEB Save to: ENDNOTE®
I Wrote These Publications R
Save to other Reference Software Save
(0)

View in: | 体中 | 繁體中 | English | 日本語 |