Information Sciences 502 (2019) 59–75

Contents lists available at ScienceDirect

Information Sciences

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

Result diversification in image retrieval based on semantic distance

Wei Lu[a,b], Mengqi Luo[a,b,*], Zhenyu Zhang[a,b,d], Guobiao Zhang[a,b], Heng Ding[a,b], Haihua Chen[c], Jiangping Chen[c]

[a] School of Information Management, Wuhan University, Wuhan, HuBei, China
[b] Information Retrieval and Knowledge Mining Laboratory, Wuhan University, Wuhan, HuBei, China
[c] Department of Information Science, University of North Texas, USA
[d] Tencent, Shenzhen, Guangdong, China

ARTICLE INFO

Article history:
Received 30 March 2018
Revised 1 June 2019
Accepted 8 June 2019
Available online 11 June 2019

Keywords:
Image retrieval
Result diversification
Semantic distance algorithm
Social tag
Re-ranking algorithm

ABSTRACT

User requirements for result diversification in image retrieval have been increasing with the explosion of image resources. Result diversification requires that image retrieval systems are made capable of handling semantic gaps between image visual features and semantic concepts, and providing both relevant and diversified image results. Context information, such as captions, descriptions, and tags, provides opportunities for image retrieval systems to improve their result diversification. This study explores a mechanism for improving result diversification using the semantic distance of image social tags. We design and compare nine strategies that combine three different semantic distance algorithms (WordNet, Google Distance, and Explicit Semantic Analysis) with three re-ranking algorithms (MMR, xQuAD, and Score Difference) for result diversification. In order to better prove the effectiveness of our strategy of applying semantic information, we also make use of visual features of images for result diversification experiment and make comparison. Our data for experimentation were extracted from 269,648 images selected from the NUS-WIDE datasets with manually annotated subtopics. Experimental results affirm the effectiveness of applying semantic information for improving result diversification in image retrieval. In particular, WordNet-based semantic distance combined with the Score Difference (WordNet-DivScore) outperformed other strategies in diversifying image retrieval results.

1. Introduction

The evolution of image sharing websites has led to an explosion of image resources on the Internet. Consequently, it has become more challenging to effectively manage these image resources and to satisfy user needs for high-quality image retrieval. In the early stages of image retrieval research, most studies concentrated on finding relevant images for user queries [36]. An image retrieval system may use visual features and context information to locate relevant images. Visual features refer to the visual contents of digital images, such as color, texture, and shape. As a result, many images that were similar in color, shape, or size were retrieved and presented to users. For instance, Flickr Distance [44] is an algorithm which

* Corresponding author at: School of Information Management, Wuhan University, Wuhan, HuBei, China.
E-mail address: lakeygtgz@163.com (M. Luo).

https://doi.org/10.1016/j.ins.2019.06.020
0020-0255/

makes use of visual features and spatial information of image content to measure distance between the visual language models corresponding to the concepts. However, users often want to retrieve images that are semantically relevant to their queries regardless of the images' shapes, colors, and textures. For example, users submitting a query term, such as "apple", into a system may want to retrieve more diverse search results; that is, images of apples with different shapes, colors, and textures. To achieve this, image retrieval systems need to diversify search results and return images relevant to the query term but visually different from each other; in the case of "apple", the system should return both sliced and whole apples with different kinds of shapes, colors, and textures.

Our study is about image retrieval result diversification. It was motivated by several problems and gaps in current image information retrieval. Result diversification in image retrieval has attracted more and more attention in recent years. However, it has been challenging due to the semantic gap between visual features and the high level semantic concepts present in context information, which includes texts associated with images, such as captions, descriptions, and tags. Context information, if it can be obtained, provides better opportunities for result diversification. Luckily, many images are tagged by experts or by social users, which allows researchers to explore the roles of image context information for result diversification. A few studies have been conducted in this area. Some researchers have made use of the difference of image labels to diversify the image retrieval results, but these studies have been just preliminary attempts [25]. There is room for performance improvement by investigating new approaches. Especially since fewer studies have discussed the influence of the semantic distance among image tags on image retrieval [25]. Semantic distance is one concept derived from semantic similarity. Semantic similarity represents the likeness between terms according to their meanings or concepts. In image retrieval, semantic distance is defined as the distance between high level features, which reflect image context including keywords, tags, captions, subtitles, labels, concepts, and so on; in essence, they are text information. Thus, calculating semantic distance would make use of the text associated with the image. In our study, we define semantic distance between images as the distance between textual information; in a nutshell, image social tags. These tags come from social users with their tagging behaviors. Compared to visual information, semantic information involves image context, background knowledge, and logical information. Visual information could not represent the mentioned deep semantic information of images. Hence, low level features of images cannot sufficiently reflect user cognition and perception, while high level features would reduce the semantic gap better than low level features in image retrieval. As social user tags reflect different user interpretations of images, they should be effective when used to assist in reducing the semantic gap. We use social user tags for semantic distance calculation, on the one hand, to make full use of human cognition and perception; on the other hand, to overcome the semantic gap. Further, the performance of image semantic distance in different algorithms of result diversification in image retrieval has not been systematically investigated.

Our study enriches the literature by exploring result diversification in image retrieval based on context information. Our purpose is to investigate the effects of different semantic distance algorithms on image result diversification. The contributions of this study include:

(1) We explored a new direction for image result diversification applying semantic distance in image retrieval. Our study compared three semantic distance algorithms: Google similarity distance, WordNet-based distance, and ESA-based distance algorithms for calculating semantic distance between both social tags and visual tags of images, and we also compared them with visual features-based distance. Although Google similarity distance has been used in previous research, we analyzed and discussed its effectiveness more extensively than previous studies. In addition, we investigated two other representative semantic distance algorithms.
(2) Our experimental results demonstrate the effectiveness of applying semantic distance algorithms in result diversification in image retrieval, which confirms the validity and generalizability of the combination of the semantic distance algorithm and the results re-ranking algorithm. Our experimental results also indicate that using semantic information for diversification would have a better performance than using visual information. Although previous research of result diversification in image retrieval has involved applying semantic distance, its effectiveness has not been discussed in depth. In this study, we combined semantic distance algorithms and the visual features-based distance algorithm with different re-ranking algorithms for image result diversification. Their performances were compared and analyzed. We, therefore, identified the most effective combination of algorithms for image result diversification.
(3) We refined the categories of image retrieval terms and expanded the application scenarios of result diversification in image retrieval. In this paper, we studied the scenarios of result diversification in image retrieval by using different query terms. In this area, researchers often pay more attention to the diversification of retrieval results, while ignoring the analysis of retrieval terms. In our work, we classified the query terms into *scene* query, *object* query, and *other* query, and analyzed the effectiveness of different semantic distance algorithms under different types of retrieval terms. Our work provides a valuable reference for refining the scenes of result diversification in image retrieval.

In summary, the approach for result diversification in image retrieval proposed in this research can provide users as much information as possible with less energy in this age of information overload, and meet the information needs of users for result diversification in image retrieval.

The rest of the paper is organized as follows. Section 2 reviews related studies on information retrieval result diversification, especially image retrieval result diversification. Section 3 describes the use of three semantic distance algorithms, the visual features-based distance algorithm, and three re-ranking algorithms for result diversification. Section 4 presents our experimental settings, including the approaches used for the dataset selection, data annotation, retrieval system con-

figuration, evaluation measures, and results analysis. Section 5 gives details about our experimental results and evaluation. Section 6 discusses the findings and contributions of our study. Finally, Section 7 summarizes the paper and outlines future research.

2. Related studies

2.1. Diversification strategies

'Diversification' is a well-identified challenge for image retrieval. It requires that the retrieved images are both relevant to the query and different from each other. Results should contain different aspects or subtopics as related to that query. Image retrieval systems usually implement result diversification in two steps: (1) the system retrieves relevant images, and then (2) re-ranks the results [12]. Re-ranking is a process that combines the semantic and visual information of an image to obtain diversification after retrieval. Along with result diversification techniques, query expansion techniques have also been used in many retrieval systems to improve retrieval performance [19].

2.2. Diversification approaches in general

There has been considerable research on diversification retrieval, especially in text retrieval. The earliest study of diversification ranking conducted by Carbonell in 1998 applied the MMR (Maximal Marginal Relevance) algorithm [6], which was widely used for result diversification in subsequent years. For example, Xia et al. [46] proposed a diverse ranking model based on the MMR strategy. Other approaches included the Score Difference method [24], the learning-to-rank approach [34], and a hierarchical structure that considered the user's intentions [20].

2.3. Diversification in image retrieval with visual features

Most image diversification studies have focused on visual features. As notable examples, Dang-Nguyen et al. [13] performed clustering based on both textual descriptions and visual contents and Boato et al. [2] considered visual saliency as a diversification processing element. In addition, the graph clustering model was also applied to obtain diversified results by utilizing word-to-image correlation [47]. Other methods included unsupervised hierarchical clustering based on user feedback [3] and a SVM classifier-based method integrating automatization and human feedback [4]. A distance metric learning algorithm was proposed in [29], which exploited social tags and visual features of images to learn image similarity for image retrieval, and the work in this research also diversified the results. The system in [31] utilized visual information, semantic information, and social tags for the relevance and diversity of image retrieval results. These approaches generally improved diversification performance. However, it is still difficult to explain images using visual features due to the lack of semantic interpretation. To address this semantic gap, in this study, we apply the semantic information of images to improve performance of result diversification.

2.4. Diversification and relevance

A few studies on image retrieval have emphasized both relevance and diversification [24]. Wu et al. [45] utilized a non-uniform matroid constraint to assign different weights to different images according to their categories; Deselaers et al. [14] set a criterion for measurement and proposed three algorithms to strengthen both relevancy and diversity. Their results showed that a Dynamic Programming (DP)-based algorithm could better optimize the relevance and diversity of image retrieval results. Similar to DP-based programming, a dissimilarity-based greedy selection algorithm by calculating dissimilar scores is also widely applied in image diversification [5].

2.5. Data mining approaches

In addition to the methods mentioned above, some attention has also been paid to data mining approaches for image retrieval diversification, such as those that use frequent patterns [42] and the clustering of images by utilizing multiple features [12]. Kuoman et al. [26] also proposed an agglomerative hierarchical clustering (AHC) algorithm, which is based on trees of concepts by utilizing conceptual features, such as image theme and semantic words, and these features proved to be more useful than visual features in image diversification.

2.6. Social tag approaches

With the rise in popularity of social media, it has become convenient for users to describe images by using social tags. Therefore, social tags have become an important semantic and conceptual description for images. Image captions, tags, and descriptions have played key roles in image resource sharing in user-generated content [18]. They incorporate cognition from human brains and also reflect the information requirements of users [33]. Some methods were proposed to explore social tags for reducing the semantic gap. For instance, Li and Tang [30] combined social tags and visual features to deal with

the semantic gap in image tag refinement; similarly, in order to represent image data appropriately, Li et al. [28] utilized social tags of images to overcome the semantic gap. A few studies have used social tags to implement concept-based image diversification. For example, Yang et al. [48] processed diversification ranking for image retrieval results based on relevant ranking results; Kim et al. [25] applied image tags diversification for image retrieval; and Qian et al. [32] built tag graphs and processed community detection for image diversification. In addition to tag concepts, semantic concepts in image tags were also exploited [48]. Social tags as the primary image search tool are widely applied in multimedia retrieval. By using social tags, Cheng et al. [9] proposed a music retrieval model which can capture user-specific information and associate user preference and query terms. A Dual-Layer Music Preference Topic Model was proposed in [8] which exploited social tags and user listening logs for personalized music retrieval.

2.7. Social tags and visual content

Image information could be represented by both visual content and semantic tags. These two types of information are widely used in the image retrieval research area. Some of the previous research only used one type, while others combined both of them. For instance, the study in [17] Gao et al. proposed a hypergraph learning approach combining bag-of-words and bag-of-visual-words representations of images for an image retrieval task. However, a previous study concluded that, in image retrieval, tags were applied more often than content [31]. In order to recommend a social image, Zhang et al. [50] made use of user provided tags and utilized image visual content to assist in building a user tag model. Additionally, social images and tags were used by a visual-semantic embedding-based weak supervision mechanism in [49], which aimed to learn visual feature representation. Apart from these, automatic image annotation should use semantic tags in the learning stage, to build the connection between words and visual features. Tag propagation based on image neighborhoods is one of the most common image annotation approaches, which relies on social tags provided by users [41]. Also, the relation between tags and visual information has been explored in [39], which proposed metrics to quantify tag visual representativeness and emphasized the importance of tag visual representativeness. As tags come from different users, they involve not only the content of the images, but also the cognition and preference of users, thus they are important semantic representatives [50]. In conclusion, social tags as semantic information of images are superior to visual information of images in image retrieval and related work. Visual features usually represent the low level information of images. By contrast, social tags involve different aspects of a user's perspective of images, and they reflect a range of levels of image information, such as situation, background knowledge, and visual entity.

In our work, the NUS dataset contains tags from social users, and we manually annotate subtopics for the images in our experiment. In order to insure the accuracy of retrieval, we manually choose the most visually representative tag from each image as the query. For example, all images with the tag "tiger" have this animal as their main topic, although they may describe different types of tigers, so one subtopic is "white tiger". Apart from the social tags contained in the NUS image dataset, we use the Cloud Vision tool from Google Cloud to obtain visual tags for those images we used.

Web 2.0 and the advancement of image sharing websites such as Flickr allows users to share images and add tags easily. These user-generated tags can be used not only for image management but also for image result diversification. Thus, in this paper, we apply the social tags of images for both image retrieval and result diversification. For the sake of comparison, we also apply the visual tags generated by Google Cloud Vision for result diversification. Different from previous studies using social tags but only a single diversification algorithm, our approach applies Google Distance, WordNet, and ESA algorithms to calculate the semantic distance among image social tags and visual tags, and combines them with MMR, xQuAD, and Score Difference re-ranking algorithms to perform results diversification in image retrieval. The next section will describe the diversification algorithms we examined in our result diversification experiments.

3. Diversification algorithms

This study aims to examine whether the semantic distance could improve diversification performance and to compare the performance of different semantic distance algorithms on diversification. Result diversification is implemented through re-ranking of the retrieval results incorporating the calculation of semantic distances and visual-based distances. This section describes the diversification algorithms. In the following (3.1–3.2), we present the three semantic distance algorithms, visual features-based distance algorithm, and three re-ranking algorithms in detail.

3.1. Semantic distance calculation

In this study, every image is initially described by a set of social tags. We define the tag sets as $T = \{t_1,t_2,\ldots, t_n\}$. For every tag t_i (t_i is an element of T) and t_j (t_j is an element of T), the semantic similarity $sim(t_i, t_j)$ is calculated by using one of the three semantic similarity algorithms: Google Distance, WordNet, and ESA, which are described in Sections 3.1.1–3.1.3.

For each image d_i, its semantic description is a set of tags, such that:

$$t(d_i) = \{t_{i1}, t_{i2}, t_{i3}, t_{i4}, \ldots, t_{in}\}$$

where n is the number of tags associated with image d_i.

Suppose the semantic distance between every two images d_i and d_j depends on the semantic similarity between them, then the semantic similarity can be calculated by the following formula:

$$\mathrm{sim}(t(d_i), t(d_j)) = \frac{\sum_{t_m \in (t(d_i)),\ t_n \in (t(d_j))} sim(t_m, t_n)}{|t(d_i)|\left|t(d_j)\right|} \tag{1}$$

where t_m and t_n are description tags for images d_i and d_j, respectively, and $|(t(d_i)|$ and $|t(d_j)|$ indicate the size of the tags set for images d_i and d_j, respectively.

Thus, the semantic distance between these two images can be defined as:

$$diffScore\left(d_i, d_j\right) = \frac{1}{sim(t(d_i), t(d_j))} \tag{2}$$

The semantic distance $diffScore(d_i, d_j)$ between each pair of images depends on the semantic similarity of their tags, which indicates that the bigger the semantic similarity between the images, the smaller their semantic distance. In order to compare the effectiveness of different approaches, we calculated the semantic similarities by using the following three algorithms.

3.1.1. Google Distance

Google Distance is derived from the concept of frequency of two related concepts that appear in one web page [11], which is a semantic similarity calculation algorithm based on search engines. Using Google Distance, for every two words t_i and t_j, we can obtain a number of web pages retrieved by t_i and t_j, from the Google search engine. We can then calculate the semantic similarity between the two words by applying the formula derived from Google Distance:

$$sim\left(t_i, t_j\right) = \frac{max\left\{logf(t_i), logf\left(t_j\right)\right\} - logf\left(t_i, t_j\right)}{logM - min\left\{logf(t_i), logf\left(t_j\right)\right\}} \tag{3}$$

where $f(t_i)$ and $f(t_j)$ are the frequency of the tags t_i and t_j that appeared in the collection of all images, and where $f(t_i, t_j)$ indicates the frequency of the tags t_i and t_j that appeared in the same image. M is the total number of images in the data collection.

3.1.2. WordNet similarity

WordNet is the most popular English lexical dictionary database that contains English nouns, verbs, adjectives, and adverbs [15]. To improve techniques of image annotation and retrieval, some studies investigated the role of semantics with the aid of WordNet [7]. Notably, in order to expand query terms and have the image retrieval results covering a wider range of concepts, Iftene et al. [21] used WordNet for mapping queries to other Wikipedia entities, which was an instance of utilizing semantic relatedness to diversify image retrieval results.

In this study, we used the semantic relationships among words in WordNet to calculate the semantic similarity of image tags. Specifically, we used the NLTK, a tool for natural language processing in Python. NLTK provides a function called "path similarity" that outputs the similarity scores of two words, which were used in our study. WordNet, as a dictionary, provides the definition, or sense, for each word, as well as the hierarchical structure for the words. The "path similarity" function exploited the shortest path between interconnected concepts based on the sense of the words in the hierarchical structure, and gives a similarity score between two words ranging from 0 to 1. For instance, "dog" and "cat" are terms for two different animals. Their similarity score calculated by "path similarity" is 0.2. If two words belong to the same synset in WordNet, their similarity score is 1. We choose the highest similarity between those words with ambiguity to ensure diversification to a certain degree, because high similarity means low distance, which would help to avoid putting those images close to each other in the results list after re-ranking. For the compound tags, we regarded them as sentences, and compared each word one-by-one in the two sentences. For example, for the compound words "airport worker" and "airport panorama", we calculate four similarity scores: sim(airport, airport), sim(airport, panorama), sim(worker, airport) and sim(worker, panorama), and then we average them to obtain the final similarity score.

To our knowledge, this study is the first to use WordNet to calculate the semantic similarity between image tags for result diversification in image retrieval.

3.1.3. Explicit Semantic Analysis

Explicit Semantic Analysis (ESA) is a knowledge repository and rule-based approach for the semantic relationship calculation of text [16]. In a previous study, ESA was used in semantic similarity measurement [37], sentiment analysis [43], question answering systems [1], and so on. However, no prior work has yet used ESA in image retrieval; we are the first researchers to use ESA to calculate the semantic similarity of image tags. In order to calculate the semantic similarity of image tags, we used a package called "ESALib" in Java for implementation of this approach. This approach maps the input terms to Wikipedia concept vectors with TFIDF scheme-based weights, and the semantic similarity between two terms is then calculated by using the vectors.

3.1.4. VGGNet-16 and cosine distance

VGGNet is a deep convolution neural network proposed in [38]. We applied this model with 16 weight layers (VGGNet-16) for visual features extraction. In order to implement this network with moderate computing requirements, each image is re-scaled to the unique size of 224 × 224. Then we take the activations of the first fully connected layer as the feature representations with a vector of 4096 in dimension. After transforming the visual features into vectors, we applied cosine similarity to measure the distance between the visual features of images.

3.2. Re-ranking for diversification

We applied three algorithms to diversify the image retrieval results: the MMR, xQuAD, and Score Difference algorithms. The first two are iterative, while the third one is for a single time re-ranking.

3.2.1. MMR re-ranking for diversification

Maximal Marginal Relevance (MMR) was proposed in 1998 [6]. It has frequently been used in re-ranking the results in information retrieval [46]. This algorithm measures the correlation $sim(d_i, q)$ between a document and a query and the correlation $sim(d_i, d_j)$ between two documents. Based on the criteria of a linear combination of $sim(d_i, q)$ and $sim(d_i, d_j)$, it re-ranks retrieval results by constantly selecting document d_i with the maximize $sim(d_i, q)$ and the minimal $sim(d_i, d_j)$.

Since MMR has been proven to be one of the most effective diversification re-ranking algorithms if proper parameters are set [6,46], we chose it as one of the re-ranking algorithms in our experiment. Here, we set the same correlation degree function for $sim(d_i, q)$ and $sim(d_i, d_j)$. Combined with the three similarity algorithms mentioned in Section 3.2, the MMR algorithm is applied as follows:

$$d = \underset{d \in D}{argmax}\left[\lambda sim(d_i, q) - (1-\lambda)\max_{d_j \in S} sim\left(d_i, d_j\right)\right] \tag{4}$$

Among them, d is the document with highest diversification score, D is the retrieval results dataset, and S is the re-ranking results dataset.

3.2.2. xQuAD re-ranking for diversification

xQuAD is a re-ranking algorithm for search result diversification [35]. This method first discovers the subtopics of the query term, and based on the dataset of correlation results, it then combines four elements to process diversity re-ranking with multiple iterations. These include the important level of the subtopic, the coverage degree of the document on the subtopic, the novel level of the document, and the correlation degree between the retrieved document and the query term. To the best of our knowledge, this algorithm has not been used in image retrieval. In our study, we apply xQuAD as one of the three re-ranking algorithms, and we use the four elements (the important level of the subtopic; the coverage degree of the document on the subtopic; the novel level of the document; the correlation degree between the retrieved document and the query term) mentioned above and the subtopics of images labeled manually (the labeling process will be described in Section 4.2). Since this algorithm will make use of a subtopic, it mainly focuses on semantic-based re-ranking, so we can only use it for semantic distance-based re-ranking.

3.2.3. Score Difference-based re-ranking for diversification

The Score Difference method for retrieval result diversification was proposed in 2014 [24] and has been widely used in document retrieval. It calculates the difference score between the ranking values of two documents from a ranking strategy: if the different scores between two documents is higher than a difference threshold parameter, they are placed in different subtopics. In this study, we proposed a DivScore algorithm based on Score Difference as shown below:

The DivScore Algorithm.

for $1 < i \le |R(q)|$ do
  $DivScore(d_i) = (1 - \frac{i-1}{N}) \times sim(d_i, q) + \frac{i-1}{N} \times diffScore(d_i, d_{i-1})$
end for
Sort $R(q)$ on $DivScore(d_i)$

where $R(q)$ represents a set of retrieved images related to query term q; $R(q) = \{d_1, d_2, d_3, \ldots, d_i\}$; images in $R(q)$ are ranked by correlative score with query term q; $sim(d_i, q)$ represents the correlative score of image d_i related to query term q; and N is the number of images in the correlative results. In our work, we use the first ranked 100 images for experimental comparison. Thus, $N = 100$; $diffScore(d_i, d_{i-1})$ is the score the of difference between image d_i and image d_{i-1}, and it can be obtained from formula (2) in Section 3.2; and $DivScore(d_i)$ is the diversity score of image d_i related to this query term.

The DivScore Algorithm utilizes a comprehensive score for re-ranking through the linear combining similarity score and the different images. The re-ranking procedure first applies the TagIR algorithm [40] to obtain the score for correlation between the image and the query term, and then ranks the results by this score to obtain the results in dataset $R(q)$. The image with the highest correlative score would be the first one in the re-ranking results dataset, $DivScore(d_i)$ would be calculated and assigned to the remaining images starting from the second one, and these images would be re-ranked according to their $DivScore(d_i)$. Therefore, we are able to obtain a Score Difference-based re-ranked dataset of diversified images.

4. Experiments and evaluation

4.1. The research design

This study consisted of several steps, including image collection construction, relevance retrieval, semantic distance calculation and re-ranking, visual-based distance calculation and re-ranking, and result diversification evaluation. Fig. 1 depicts our research design.

4.2. The data

Our original data, the NUS-WIDE dataset, is an image database provided by the Lab for Media Search at the National University of Singapore. This dataset collected 269,648 images from Flickr. It provides the ground truth for 81 concepts by using the manual tags of images, and all of the 5018 unique tags exist in WordNet [10]. According to the level of abstraction, the NUS-WIDE dataset can be divided into two subsets: 1) the NUS-WIDE-SCENE and 2) the NUS-WIDE-OBJECT [10]. To facilitate evaluation, we chose 29 concepts as query terms which were related to less than 1000 images as the dataset of terms for our experiment. Meanwhile, we also chose images that contain both visual objects and scene descriptions. The 29 query terms belong to one of following categories: *scene, object*, or *other*. Among them, *scene category* contains 11 terms: airport, castle, frost, glacier, harbor, nighttime, rainbow, sand, town, valley, and waterfall; *object category* contains 12 terms: book, computer, elk, flags, fox, leaf, map, tattoo, tiger, vehicle, whales, and zebra; and *other category* contains 6 terms: dancing, earthquake, running, soccer, surf, and swimmers.

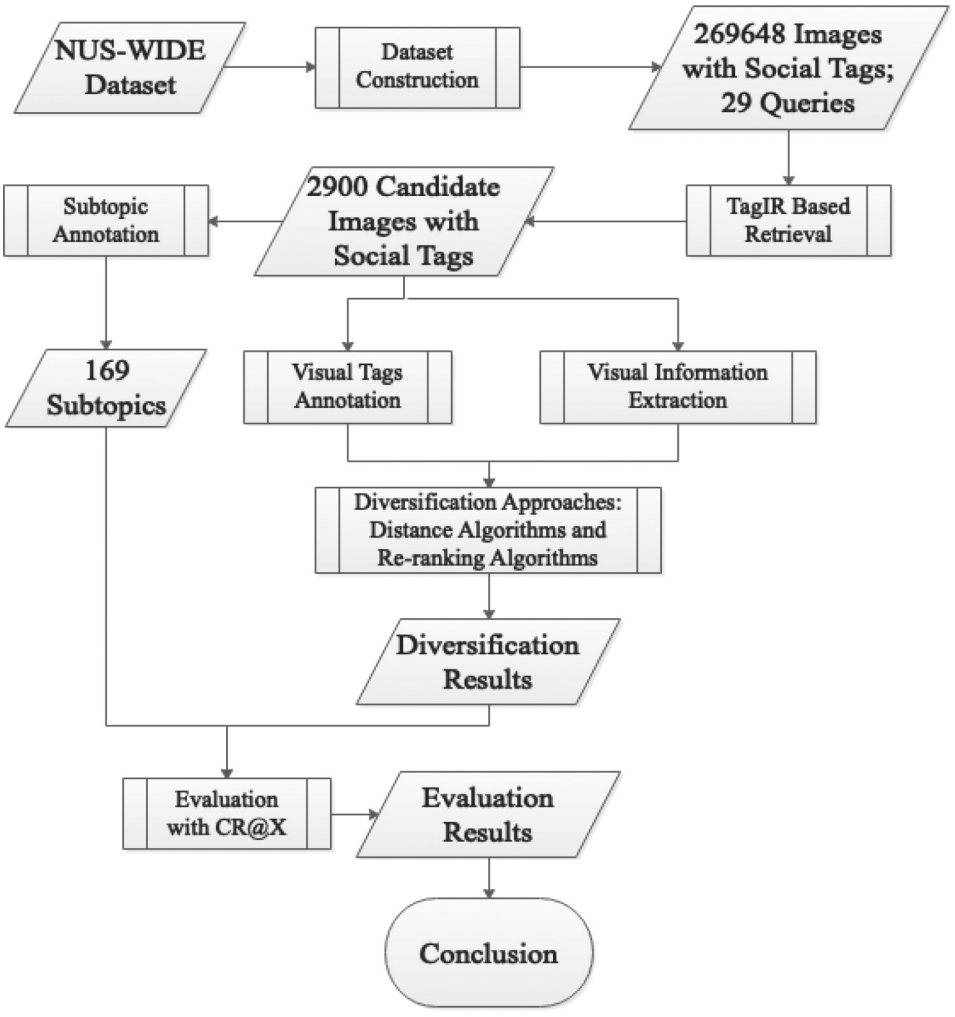

Fig. 1. Research design.

Table 1
Query terms and their subtopics.

Query Term	Subtopics
airport	airport hall, airport panorama, airport worker, civil airport, military airport
book	book, book design, bookshelf, notebook, reading book
castle	castle courtyard, castle design, castle gate, castle hill, castle moat, castle night, castle tower, castle wall, ruined castle
computer	computer business, computer poster, computer room, desktop computer, notebook computer, palm computer
dancing	cheerleading dancing, dancing animals, dancing designs, dancing kids, dancing movement, dancing party, dancing show, folk dancing, light dancing, street dancing
earthquake	earthquake destroy, earthquake rescue, earthquake victim
elk	dead elk, elk cartoon, elk mating, female elk, male elk
flags	colored flag, flag pattern, national flag, team flag
fox	baby fox, fox animal, fox painting, fox pet, snow fox
frost	frost flower, frost weather, rime, snow frost
glacier	enjoy glacier, glacier adventure, glacier animals, glacier melting, glacier scene
harbor	civil harbor, harbor bridge, harbor city, harbor fishing, military harbor
leaf	autumn leaf, falling leaf, frost leaf, green leaf, leaf animals, leaf art, leaf flowers, leaf vein, lotus leaf, red leaf
map	digital map, hand drawn map, paper map, traffic map
nighttime	bridge nighttime, building site nighttime, city nighttime, light painting nighttime, lonely night, moon nighttime, near night, night traffic flow, rainy nighttime, wild nighttime, winter nighttime
rainbow	rainbow, rainbow drawing
running	race, running animals, running away, running exercises, running people, running play, running rain, running to work
sand	sand art, sand beach, sand desert, sand dune, sand print, sand stone, shadow on sand
soccer	amateur soccer, professional soccer, soccer fans, soccer field, soccer poster, soccer referee, soccer shooting, soccer star
surf	prone surfing, standing surfing
swimmers	child swimmers, professional swimmers, swimmer diving, swimmers, swimmers under water, water ballet swimmers, winter swimmers
tattoo	tattoo, back tattoo, leg tattoo, tattoo friend, tattoo girl, tattoo man, tattoo work
tiger	tiger animal, tiger people, tiger pictures, white tiger
town	activity town, beauty town, sunset town, town night, town painting, town street, town traffic, town winter
valley	valley life, valley scene, valley town, valley traveler
vehicle	air vehicle, destroyed vehicle, land vehicle, military vehicle, two-wheel vehicle, water vehicle
waterfall	waterfall scene, waterfall shower, waterfall travel
whales	watch whales, whales animal, whales painting, whales performance, white whales
zebra	zebra animals, zebra baby, zebra close-up, zebra crossing, zebra painting, zebra wild

4.3. Subtopic annotation for evaluation

In order to evaluate the effect of our proposed approaches on image result diversification, we applied the subtopic idea assuming that images associated with the same tag may have different subtopics representing different semantic concepts. By annotating subtopics for each concept, we could evaluate diversification by checking whether the algorithm could come up with similar results as could be achieved with human annotation. This approach is similar to system-oriented information retrieval where human judgment is necessary for evaluating information retrieval performance.

Subtopic annotation mainly relies on the visual differences among images. We took a similar approach as in [23]. Specifically, we invited five volunteers to take part in the annotation. Among them, two performed the first step that manually annotated all the images related to each query term with subtopics. In the second step, the third volunteer reviewed the annotation from the first step and unified all the subtopics for each query term. In the third step, the remaining two volunteers judged the correctness of the annotated subtopics independently. The subtopics annotated by the volunteers has been validated by the consistency test. It is worth mentioning that we only use those subtopics appeared in WordNet.

Finally, we obtained 169 subtopics, with an average of 5.8 subtopics for each query term. Table 1 presents the query terms and their subtopics.

4.4. Data preprocessing for retrieval

Image semantic distance calculation mainly relies on descriptive information for images. The dataset of social image tags in the NUS-WIDE originates from Flickr where users originate from all over the world. These users may use different languages to tag the images, therefore, the tags need to be cleaned and reformatted to facilitate further processing.

Preprocessing included language transformation, word form restoration, and duplicate semantic words removal. Language transformation uses a translator to translate words from one language into another language. As most of the social tags in Flickr are in English, we chose English as the semantic distance calculation language. We thus translated all others language tags into English tags. We also performed lemmatization to unify the descriptions in English and to remove duplicate tags.

Similarly, to reduce ambiguity, we transformed social tags into consistent forms, especially for proper nouns and phrases. For instance, “Des Moines” is a place name and shouldn’t be regarded as two words. We therefore used “_” to connect them as a single term (i.e., “Des_Moines”).

4.5. Relevant results retrieval

We adopted an algorithm called TagIR proposed by Sun et al. [40] to obtain the relevance image results for each query term. TagIR calculates relevance based on tag relatedness, tag discrimination, tag length normalization, tag-query matching, and the query itself. It is consistent with our proposed data selection and data processing approach.

4.6. Re-ranking experiments and evaluation

Combining the three semantic distance calculation algorithms and the three image diversity re-ranking algorithms presented in Section 3, we conducted nine groups of experiments: Google-MMR, WordNet-MMR, ESA-MMR, Google-xQuAD, WordNet-xQuAD, ESA-xQuAD, Google-DivScore, WordNet-DivScore, and ESA-DivScore, which will be used to represent their respective experiments in the remaining texts. For examples, WordNet-MMR, ESA-MMR, and Google-MMR refer to the diversification experiments using the three different semantic distance algorithms in combination with the MMR re-ranking algorithm. For the aims of comparison, the nine groups of experiments are conducted by using both social tags and visual tags of images; and we also conducted experiments of the visual features distance-based re-ranking algorithm; the one is Visual-MMR and the other one is Visual-DivScore.

We performed the result diversification based on the relevant image results from the TagIR method. By horizontally comparing the experimental results, we aim to explore the effectiveness of different semantic distance algorithms under the same diversification re-ranking ideology. Similarly, by vertically comparing the experimental results, we try to quantify the effectiveness of different diversity re-ranking ideologies under the same semantic distance algorithm.

4.6.1. The diversification measure

We use CR@X (Cluster-Recall at X, $X = 5$,10, 20, 30, or 40) [27] to measure the performance of our approach for diversification in image retrieval. CR@X provides criteria for the diversity degree of image retrieval by calculating the number of different subtopics in the top X retrieved images. Under this indicator, results with a higher value of CR@X would be considered to have a better effect on diversification in image retrieval. It can be calculated as follows:

$$\mathrm{CR@}X = \frac{\left|\cup_{i=1}^{K} subtopics(D_i)\right|}{N_t} \tag{5}$$

where K is the number of relevant images in the result dataset; D_i is the image ranked i in the result dataset; $subtopics(D_i)$ is the subtopic corresponding to D_i; Nt is the total number of subtopics related to the query term; and X represents the total number of retrieved images. Since this formula is used to calculate CR@X for one query, the results reported in the next section for CR@X is the average over the 29 queries.

ImageCLEF (2009) used CR@10 as an official evaluation metric in the task of image retrieval, while MediaEval (2014) used CR@20 as an official evaluation metric in the task of diversification in image retrieval [22]. In our work, we have a complete analysis on the proposed approach. To do this, we calculated CR@X when $X = 10$, 20, 30, 40 of CR@X for each experiment.

4.6.2. Analysis of experimental results

We analyzed the experiments from three dimensions:

(1) *The effects of semantic distance algorithm on result diversification.* We chose WordNet, ESA, and Google Distance as our semantic distance algorithms, and apply them by using both social tags and visual tags. In order to make a comparison of semantic distance and visual distance, we also use visual features extracted by the VGGNet-16 model, and calculate the cosine distance between them. As every semantic algorithm has its own construction of semantic relationship, we can explore the impact of different semantic relationship constructions on diversification in image retrieval.
(2) *The effects of semantic distance and re-ranking algorithms on retrieval results.* We selected three image diversification re-ranking algorithms: MMR, xQuAD, and DivScore. To analyze the difference between different ideologies of image diversification, we use re-ranking as our second aspect of evaluation.
(3) *The effects of the semantic distances and re-ranking algorithms on different categories of query terms.* We separated image query terms into three categories: *scene* query terms, *object* query terms, and *other* query terms. From the dimension of categories of terms to evaluate, we can compare the impact on classification in image retrieval by exploring the difference among subtopics in different categories.

The evaluation results are reported in the next section.

5. Results

We evaluated the nine experiments (WordNet-MMR, ESA-MMR, and Google-MMR; WordNet-xQuAD, ESA-xQuAD, and Google-xQuAD; WordNet-DivScore, ESA-DivScore, and Google-DivScore) by using CR@X with varying X values (i.e., X belongs to {10, 20, 30, 40}) as measures on both social tags and visual tags of images. Visual-MMR and Visual-DivScore is performed as the comparison method. TagIR is performed as the baseline method. In the following tables, the best results for each

Table 2
The original and diversification results for the query "airport".

ID	Tags	Semantic distance	Re-ranked ID	Tags
90077	plane airplane airport aircraft aviation	–	90077	plane airplane airport aircraft aviation
89924	night plane airplane us airport force aircraft aviation air	0.221	57542	Amsterdam plane airplane airport wings gates aircraft air tail jets airplanes wing jet cockpit best motors engines planes airliner airliners rudder aircrafts flaps fuselage jetliner jetliners airlines
89767	sky plane airplane fly us flying airport force aircraft aviation air flight airforce	0.219	2714	voyage travel sky plane airplane flying airport wings cabin Europe aircraft aviation tail jets airplanes wing jet cockpit aerial motors engines planes takeoff touring airliner airliners rudder aircrafts flaps fuselage jetliner jetliners airplanes
3581	plane airplane airport aircraft aviation Des_Moines	0.226	57760	voyage travel plane airplane airport wings gate gates aircraft tail jets airplanes wing jet cockpit motors engines planes a touring airliner airliners rudder a aircrafts flaps fuselage jetliner jetliners airlines
90082	plane airplane flying airport aircraft aviation	0.287	57667	travel plane airplane flying airport wings aircraft tail jets airplanes wing jet cockpit motors landing Malaysia planes airliner airliners rudder aircrafts flaps fuselage jetliner jetliners airlines
137956	airplane airport aviation jet	0.318	57679	travel Canada plane airplane flying airport wings aircraft aviation tail jets airplanes wing jet cockpit motors landing engines landing gear planes arrival airports runway spotting airliner airliners rudder aircrafts flaps fuselage jetliner stopping jetliners airplanes airline
89963	plane airplane airport aircraft aviation	0.385	89959	plane airplane airport force aircraft aviation air airforce USAF Des Moines DesMoines
89905	plane airplane flying airport aircraft aviation	0.314	89704	blue plane airplane fly us flying airport aircraft aviation military flight afterburner
89876	plane airplane airport aircraft aviation	0.314	116478	California mountain vertical flying airport wings cobra fighter aircraft aviation smoke airplanes helmet delta cargo diamond airshow hornet mustang canopy blue angels Mather pilot warbird warthog aerobatics ejection seat afterburner
90092	plane airplane airport aircraft aviation Moines DesMoines	0.339	89820	USA plane airplane fly us flying airport wake force aircraft aviation air flight wash airforce Moines DesMoines afterburner onlythebestare

indicator are marked in bold, and all indicators are averaged in the same way as the baseline. Table 2 presents an example using the query term "airport". The first column shows the IDs of the original image retrieval result list. The results were ordered by their relevance to the query term. The most relevant one is presented as the first in the table. The second column shows the tags associated with the image in the first column. The third column presents the semantic distance scores calculated using the WordNet-based semantic distance algorithm and the DivScore re-ranking algorithm. The fourth column shows the image IDs of the new result list based on that score. The tags of the respective result images of the new list are presented in the fifth column.

5.1. Effects of different semantic distance algorithms

We first calculated CR@X under every image diversification algorithm for the 29 terms by using social tags, and obtained the result of CR@X for every algorithm by taking the average scores from the results of CR@10, CR@20, CR@30, and CR @40. Results show that all of the three semantic distance algorithms bring some degree of improvement on retrieval results. Fig. 2 depicts the average diversification rates among the three diversification algorithms. Among them, the improvement results from the WordNet-based semantic distance algorithm outperforms all the other algorithms.

5.2. Effects of semantic distance algorithms combined with different diversification algorithms

Table 3 shows the average CR@X scores of the nine proposed methods as well as the baseline method, Visual-MMR and Visual-DivScore method on 29 queries. Among them, the nine proposed methods are processed by using both social tags and visual tags. In the table, 'ST' represents social tags, 'VT' represents visual tags. From the results, we made the following observations:

First, after re-ranking through semantic distance and diversification algorithms, the diversification performances in image retrieval all improve over the baseline method from 0.1% to 9.6% on CR@X by using social tags. This improvement shows us that the proposed re-ranking approaches generate image retrieval output lists that are ranked better. By comparison, the diversification results produced by visual tags do not improve as much as those produced by social tags. Although on

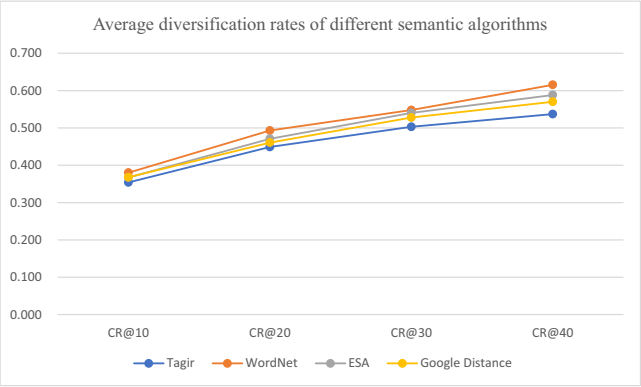

Fig. 2. Average diversification rates of different semantic algorithms.

Table 3
Evaluation results for the experiments.

Algorithms	CR@10		CR@20		CR@30		CR@40	
Baseline method								
TagIR	0.354		0.449		0.503		0.537	
Different semantic distance algorithms with MMR								
WordNet-	ST	VT	ST	VT	ST	VT	ST	VT
MMR	0.364	0.349	0.482	0.474	0.531	0.543	0.611	0.600
ESA-	ST	VT	ST	VT	ST	VT	ST	VT
MMR	0.361	0.421	0.471	0.497	0.544	0.529	0.622	0.565
Google-	ST	VT	ST	VT	ST	VT	ST	VT
MMR	0.366	0.435	0.478	0.484	0.534	0.557	0.581	0.593
Visual-MMR	0.354		0.451		0.554		0.627	
Different semantic distance algorithms with xQuAD								
WordNet-	ST	VT	ST	VT	ST	VT	ST	VT
xQuAD	0.357	0.351	0.480	0.457	0.533	0.492	0.602	0.546
ESA-	ST	VT	ST	VT	ST	VT	ST	VT
xQuAD	0.355	0.404	0.477	0.483	0.536	0.522	0.574	0.582
Google-	ST	VT	ST	VT	ST	VT	ST	VT
xQuAD	0.365	0.411	0.454	0.504	0.512	0.557	0.548	0.589
Different semantic distance algorithms with DivScore								
WordNet-	ST	VT	ST	VT	ST	VT	ST	VT
DivScore	**0.420**	0.347	**0.517**	0.444	**0.580**	0.499	**0.633**	0.543
ESA-	ST	VT	ST	VT	ST	VT	ST	VT
DivScore	0.385	0.354	0.465	0.449	0.541	0.504	0.569	0.533
Google-	ST	VT	ST	VT	ST	VT	ST	VT
DivScore	0.373	0.354	0.451	0.449	0.538	0.503	0.582	0.537
Visual-DivScore	0.354		0.449		0.503		0.537	

some indicators the diversification performance of visual tags beat the performance of social tags, on other indicators there appear to be decreases of diversification degree by comparing to TagIR. In general, as the visual tags only reflect the visual information of images, their performance for diversification was poorer than those social tags which reflect the semantic information of images.

Second, WordNet-DivScore with social tags significantly outperforms other approaches including the baseline, Visual-MMR, Visual-DivScore, and all other approaches based on visual tags on any indicators. It improves the baseline by 18.6% on CR@10, by 15.1% on CR@20, by 15.3% on CR@30, and by 17.9% on CR@40, respectively.

Although Visual-MMR and Visual-DivScore also bring a degree of improvement on diversification, the effectiveness of them was not as good as the results from semantic distance, and this is especially so when using social tags. In terms of which semantic distance algorithms performs the best under a certain diversification algorithm and which diversification algorithm contributes the most when using a certain semantic distance algorithm, we find that when using MMR or xQuAD as the diversification algorithm, which semantic distance algorithm performs better depends on the indicators. However, when using DivScore as the diversification algorithm, WordNet always performs the best no matter which indicators are used. Similar results appear when studying the influence of different diversification algorithms, except for DivScore, which

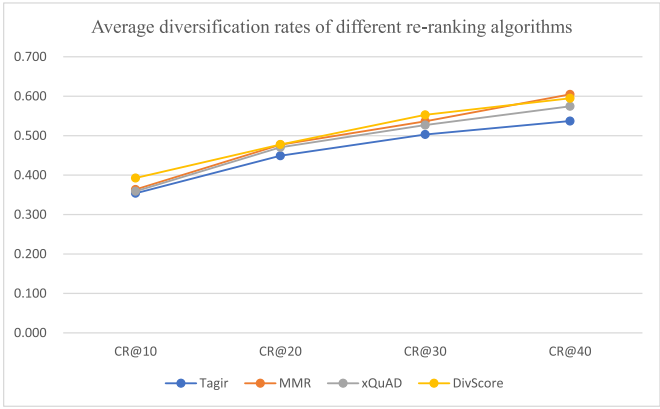

Fig. 3. Average diversification rates of different re-ranking algorithms.

Table 4
Evaluation results on different types of queries.

Algorithms	CR@10			CR@20			CR@30			CR@40		
	scene	object	other	scene	object	other	scene	object	other	scene	object	other
Baseline method												
TagIR	0.260	0.332	**0.488**	0.370	0.464	0.563	0.401	0.555	0.587	0.429	0.612	0.587
Different semantic distance algorithms with MMR												
WordNet-MMR	0.398	0.382	0.374	0.442	0.515	0.490	0.450	0.608	0.527	0.491	**0.713**	0.627
ESA-MMR	0.382	0.322	0.395	0.422	0.507	0.488	0.450	0.589	**0.624**	0.503	0.695	**0.693**
Google-MMR	0.353	0.336	0.391	0.411	0.524	0.508	0.445	0.589	0.587	0.473	0.666	0.608
Different semantic distance algorithms with xQuAD												
WordNet-xQuAD	0.351	0.315	0.374	0.411	**0.529**	0.511	0.462	0.601	0.527	0.501	0.692	0.607
ESA-xQuAD	0.333	0.336	0.371	0.404	0.526	0.512	0.417	0.610	0.607	0.456	0.653	0.630
Google-xQuAD	0.333	0.339	0.375	0.380	**0.529**	0.441	0.414	0.575	0.565	0.445	0.633	0.565
Different semantic distance algorithms with DivScore												
WordNet-DivScore	**0.449**	**0.496**	0.477	**0.536**	0.477	**0.562**	**0.572**	0.574	0.604	**0.619**	0.625	0.676
ESA-DivScore	0.368	0.385	**0.488**	0.399	0.487	0.542	0.454	**0.610**	0.563	0.481	0.644	0.584
Google-DivScore	0.387	0.372	0.434	0.399	0.468	0.512	0.445	0.586	0.611	0.520	0.594	0.669

contributes the most on all indicators while different semantic distance algorithms are used, while the others all depend on using specific indicators or the semantic distance algorithm.

Fig. 3 visually depicts the results on the re-ranking algorithms. It shows that the averaged DivScore beats others. Thus, when setting WordNet as the semantic distance algorithm and DivScore as the diversification algorithm, the corresponding performances would be significantly better. This is confirmed by the results presented in Table 3.

5.3. Results on different types of queries

We next investigated the performance on different types of queries. The results are presented in Table 4. It shows that all re-ranking approaches outperform the baseline significantly on scene category queries for CR@ {10, 20, 30, 40}, and WordNet-DivScore beats the rest of the algorithms on all the indicators.

However, in terms of object category queries, although all the re-ranking approaches outperform the baseline to some extent, some variations occur, and different algorithms achieve the best results on different indicators: WordNet-DivScore, WordNet-xQuAD (Google-xQuAD performs the same), ESA-DivScore, and WordNet-MMR, where they improved the corresponding baselines by 49.4%, 14.0%, 9.9%, and 18.1% on CR@ {10, 20, 30, 40}, respectively.

To our surprise, the baseline method achieves the best performance on *other* category queries when measured by CR@10; it also outperforms some re-ranking approaches on CR@ {20, 30, 40} for that category.

6. Discussion

6.1. Performance of different semantic distance algorithms

Our experimental results showed that WordNet had the best performance among three semantic distance algorithms, affirming the significant positive effect of WordNet as a semantic distance algorithm for image diversification.

Generally, comparing to the semantic distance algorithm, the Visual-MMR and Visual-DivScore algorithms did not perform so well. This is because of the simplification of visual features only describing the objects of images; by contrast, semantic tags describe the context of images from different perspectives, which would be helpful in result diversification. In semantic distance algorithms, using social tags achieved better diversification results than using machine-generated visual tags. The reason for this is that visual tags only represent visual information in images, such as color, shape, and object, and they do not describe the context of images.

Different semantic distance algorithms for diversification lead to large differences in diversification results. This may be caused by both the construction of the semantic network, and information noise.

The construction of semantic networks in WordNet considers hyponym, synonym, antonym, and partitive relationships among different words. Our study considered subtopics relative to images, as well as the division of subtopics (also referred to as hyponymy relations). This might explain why WordNet-based approaches achieved a better performance. In contrast, even though Google Distance-based and ESA-based algorithms improved diversification in some of our experiments, they were not as good as the WordNet-based approach because neither of these two methods contain rich semantic relationships.

The other reason for differences in diversification performance is the impact of information noise. The ontology dictionary adopted in the WordNet algorithm is manually built and has accurate semantic concept relationships. However, there is more information noise in the Wikipedia knowledge repository adopted in the ESA algorithm and in the resources from the Internet adopted in Google Distance. Thus, we consider that WordNet has an advantage compared to Google Distance and ESA in diversification.

6.2. Results as measures by CR@X

Results of our experiment showed that performance under CR@10 and CR@20 are better than others. This finding is consistent with the evaluation metrics applied by ImageCLEF 2009 (CR@20) and MediaEval 2014 (CR@20) [22]. This indicates that when more retrieved images are added to the result list, the diversification algorithms are less effective.

6.3. Effects of results diversification on different types of queries

As our experimental results showed, algorithms performed better on *scene* category terms than on terms in the other two categories on all the indicators. Also, the *scene* category achieved significant improvement on CR@ {10, 20, 30, 40}. However, improvement of the diversification effect in the *object* category terms seems to be more stable as a whole. Compared with the former two types, performance on the *other* category queries is worsened in eight of the nine experiments at the CR@10 level.

We suspect features of images in these three categories lead to the result mentioned above. The *scene* category, especially landscape images, according to our subtopic annotation method, can have multiple types of aspects, such as time, weather, seasons, or some other events, and there is a large difference between each aspect. For instance, subtopics for the query term “nighttime” could be “rainy nighttime” about weather; “nighttime moon” about scenery; “winter nighttime” about season; “city in the nighttime” about object; and even “light painting” about art. These diverse subtopics may cause the high diversification rate in the *scene* category.

As for the *object* category, due to the explicit image content, there is little difference among subtopics, and their aspects are not as rich. In other words, users cannot distinguish between them using only visual aspects. For instance, for the query term “zebra”, subtopics could be “zebra crossing”, “wild zebra”, and “zebra painting”. Although these subtopics are different from each other in concept aspects, these concepts of images come from social tags, and users would like to adopt similar concepts for image tagging. Thus, the semantic distance among those subtopics may not be large enough. Compared to the *scene* category, diversification in this category is not so remarkable.

In our work, query terms besides those in the “*scene*” and “*object*” categories are allocated to the “*other*” category. The images in this category do not have a unique topic, and their subtopics do not have a uniform rule, which may result in an unstable diversification rate, or even a decrease.

7. Conclusions

In this study, we carried out experiments to evaluate three different semantic distance algorithms (WordNet, Google Distance, and ESA) combined with three re-ranking algorithms (MMR, xQuAD, and Score Difference) on image diversification retrieval based on a subset of the NUS-WIDE image dataset. Our experiments allowed a comparison of these algorithms both

on social tags and visual tags to understand their strengths and weaknesses, and a comparison of visual distance algorithms to prove the effectiveness of semantic information in result diversification. We found that (1) All three semantic distance algorithms brought a certain degree of improvement on retrieval result diversification, and semantic distance performs better than visual distance in result diversification. Moreover, the semantic distance algorithms obtained better diversification results by using social tags than using visual tags. Among them, the WordNet-based algorithm achieved the most improvement. (2) Different re-ranking algorithms brought a certain degree of improvement on average diversification rates. Among them, DivScore achieved the best performance. (3) Among the nine combinations of semantic distance and re-ranking algorithms, the WordNet-based semantic distance algorithm combined with the DivScore re-ranking algorithm performed best for improving diversification in image retrieval. The performance increased 19%, 15%, 15%, and 18% on CR@10, CR@20, CR@30, and CR@40, respectively.

The usefulness of WordNet as a widely used lexical resource has been reconfirmed. Our analysis of the experimental results indicated that WordNet is the most appropriate method for semantic distance calculation for result diversification in image retrieval. To the best of our knowledge, this is the first study that systematically analyzes and compares different semantic distance combined with re-ranking algorithms based on image tags for diversification in image retrieval. Our study specifies a new direction for diversification in image retrieval: exploring the semantic distance of tags based on high-quality linguistic resources. We believe high-quality linguistic resources, such as WordNet, have the potential to improve the diversification of retrieval results.

The significance of this study also includes the construction of a reusable dataset for diversification research evaluation in image retrieval. Based on the NUS-WIDE image dataset, we selected 29 query terms in *scene, object*, and *other* categories, and then manually annotated subtopics for images related to each query term. There are a total of 169 subtopics in our dataset, on average, and each image has 5.8 subtopics. This dataset is available to the image diversification research community upon request.

Furthermore, we proposed a multi-dimension evaluation mechanism for analyzing the results of diversification in image retrieval. Specifically, image diversification results can be analyzed according to semantic distance and visual distance among images, re-ranking algorithms, and types of query terms. This evaluation mechanism can be applied to assess other image diversification systems.

For future research, we plan to extend the scale of our dataset, enrich query terms, and further normalize the tags of images. We also want to explore new algorithms for improving relevancy and diversification in image retrieval.

Funding

This work was supported by the grant from The National Nature Science Foundation of China (71473183) 2015–2019.

This grant supported in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Declaration of Competing Interest

None.

CRediT authorship contribution statement

Wei Lu: Funding acquisition, Project administration. **Mengqi Luo:** Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Zhenyu Zhang:** Data curation, Investigation, Methodology, Resources. **Guobiao Zhang:** Formal analysis, Validation. **Heng Ding:** Methodology. **Haihua Chen:** Writing - review & editing. **Jiangping Chen:** Writing - review & editing.

Acknowledgements

This work was supported by a grant from The Natural Science Foundation of China (71473183) 2015–2019. We would like to express our sincere gratitude to Marie Bloechle at University of North Texas and Lisa Jeon at the Linguistics Department at Rice University for her editing work to this research article.

References

[1] S.A. Aroussi, N. El Habib, O. El Beqqali, Improving question answering systems by using the explicit semantic analysis method, in: Intelligent Systems: Theories and Applications (SITA), 2016 11th International Conference on, IEEE, 2016, pp. 1–6.

[2] G. Boato, D.T. Dang-Nguyen, O. Muratov, N. Alajlan, F.G.B. De Natale, Exploiting visual saliency for increasing diversity of image retrieval results, Multimed. Tools Appl. 75 (10) (2016) 5581–5602. https://doi.org/10.1007/s11042-015-2526-4.

[3] B. Boteanu, I. Mironică, B. Ionescu, Pseudo-relevance feedback diversification of social image retrieval results, Multimed. Tools Appl. 76 (9) (2016) 1–28.

[4] B. Boteanu, I. Mironică, B. Ionescu, A relevance feedback perspective to image search result diversification, in: Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on, IEEE, 2014, pp. 47–54.

[5] R.T. Calumby, R. Da Silva Torres, M.A. Goncalves, Diversity-driven learning for multimodal image retrieval with relevance feedback, in: 2014 IEEE International Conference on Image Processing, ICIP, 2014, pp. 2197–2201. https://doi.org/10.1109/ICIP.2014.7025445.

[6] J. Carbonell, J. Goldstein, The use of MMR, diversity-based reranking for reordering documents and producing summaries, in: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '98, 1998, pp. 335–336. https://doi.org/10.1145/290941.291025.

[7] S.-H. Chen, Y.-H. Chen, A content-based image retrieval method based on the Google Cloud Vision API and WordNet, in: Asian Conference on Intelligent Information and Database Systems, Cham, Springer, 2017, pp. 651–662.

[8] Z. Cheng, S. Jialie, S.C.H. Hoi, On effective personalized music retrieval by exploring online user behaviors, in: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, ACM, 2016, pp. 125–134.

[9] Z. Cheng, J. Shen, L. Nie, T.-S. Chua, M. Kankanhalli, Exploring user-specific information in music retrieval, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2017, pp. 655–664.

[10] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, Y. Zheng, NUS-WIDE: a Real-World Web Image Database from National University of Singapore, Acmmm 1 (2009). https://doi.org/10.1145/1646396.1646452.

[11] R.L. Cilibrasi, P.M.B. Vitányi, The Google similarity distance, IEEE Trans. Knowl. Data Eng. 19 (3) (2007) 370–383. https://doi.org/10.1109/TKDE.2007.48.

[12] D.T. Dang-Nguyen, L. Piras, G. Giacinto, G. Boato, F.G.B. De Natale, A hybrid approach for retrieving diverse social images of landmarks, in: Proceedings - IEEE International Conference on Multimedia and Expo (Vol. 2015–August), 2015.

[13] D.T. Dang-Nguyen, L. Piras, G. Giacinto, G. Boato, F.G.B.D. Natale, Multimodal retrieval with diversification and relevance feedback for tourist attraction images, ACM Trans. Multimed. Comput. Commun. Appl. 13 (4) (2017) 1–24.

[14] T. Deselaers, T. Gass, P. Dreuw, H. Ney, Jointly optimising relevance and diversity in image retrieval, in: Proceeding of the ACM International Conference on Image and Video Retrieval - CIVR '09, 2009, p. 1. https://doi.org/10.1145/1646396.1646443.

[15] C. Fellbaum, G. Miller, WordNet:An Electronic Lexical Database, MIT Press, 1998.

[16] E. Gabrilovich, S. Markovitch, Computing semantic relatedness using wikipedia-based explicit semantic analysis, in: IJCAI International Joint Conference on Artificial Intelligence, 2007, pp. 1606–1611. https://doi.org/10.1145/2063576.2063865.

[17] Y. Gao, M. Wang, H. Luan, J. Shen, S. Yan, D. Tao, Tag-based social image search with visual-text joint hypergraph learning, in: Proceedings of the 19th ACM international conference on Multimedia, ACM, 2011, pp. 1517–1520.

[18] Y. Gao, M. Wang, Z.J. Zha, J. Shen, Visual-textual joint relevance learning for tag-based social image search, IEEE Trans. Image Process. 22 (1) (2013) 363–376.

[19] E. Hoque, O. Hoeber, M. Gong, CIDER: concept-based image diversification, exploration, and retrieval, Inf. Process. Manage. 49 (5) (2013) 1122–1138. https://doi.org/10.1016/j.ipm.2012.12.001.

[20] S. Hu, Z. Dou, X. Wang, T. Sakai, J.-R. Wen, Search result diversification based on hierarchical intents, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, ACM, 2015, pp. 63–72.

[21] A. Iftene, B. Alexandra-Mihaela, Using Semantic Resources in Image Retrieval, Procedia Comput. Sci. 96 (2016) 436–445.

[22] B. Ionescu, Retrieving diverse social images at MediaEval 2014: challenge, dataset and evaluation, MediaEval (2014) 1263.

[23] B. Ionescu, A. Popescu, M. Lupu, A.L. Gînsc\ua, B. Boteanu, H. Müller, Div150Cred: a social image retrieval result diversification with user tagging credibility dataset, in: Proceedings of the 6th ACM Multimedia Systems Conference, 2015, pp. 207–212. https://doi.org/10.1145/2713168.2713192.

[24] S. Kharazmi, M. Sanderson, F. Scholer, D. Vallet, Using score differences for search result diversification, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '14, 2014, pp. 1143–1146. https://doi.org/10.1145/2600428.2609530.

[25] E. Kim, T. Yamamoto, K. Tanaka, in: Computing Tag-Diversity for Social Image Search, Springer International Publishing, 2014, pp. 328–335.

[26] C. Kuoman, S. Tollari, M. Detyniecki, Using tree of concepts and hierarchical reordering for diversity in image retrieval, in: Proceedings - International Workshop on Content-Based Multimedia Indexing, 2013, pp. 251–256. https://doi.org/10.1109/CBMI.2013.6576593.

[27] M. Lestari Paramita, M. Sanderson, P. Clough, Diversity in photo retrieval: overview of the ImageCLEFPhoto task 2009, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6242, LNCS, 2010, pp. 45–59. https://doi.org/10.1007/978-3-642-15751-6_6.

[28] Z. Li, J. Liu, J. Tang, H. Lu, Robust structured subspace learning for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2085–2098.

[29] Z. Li, J. Tang, Weakly supervised deep metric learning for community-contributed image retrieval, IEEE Trans. Multimedia 17 (11) (2015) 1989–1999.

[30] Z. Li, J. Tang, Weakly supervised deep matrix factorization for social image understanding, IEEE Trans. Image Process. 26 (1) (2017) 276–288.

[31] D. Lu, X. Liu, X. Qian, Tag-based image search by social re-ranking, IEEE Trans. Multimedia 18 (8) (2016) 1628–1639.

[32] X. Qian, D. Lu, Y. Wang, L. Zhu, Y.Y. Tang, M. Wang, Image re-ranking based on topic diversity, IEEE Trans. Image Process. 26 (8) (2017) 3734–3747. https://doi.org/10.1109/TIP.2017.2699623.

[33] Y.S. Rawat, M.S. Kankanhalli, ConTagNet: exploiting user context for image tag recommendation, in: Proceedings of the 2016 ACM on Multimedia Conference, ACM, 2016, pp. 1102–1106.

[34] R.L.T. Santos, C. Macdonald, I. Ounis, Search result diversification, Found. Trends Inf. Retr. 9 (1) (2015) 1–90.

[35] R.L.T. Santos, J. Peng, C. MacDonald, I. Ounis, Explicit search result diversification through sub-queries, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 5993, LNCS, 2010, pp. 87–99. https://doi.org/10.1007/978-3-642-12275-0-11.

[36] S. Seth, P. Upadhyay, R. Shroff, R. Komatwar, Review of content based image retrieval systems, Int. J.Eng. Trends Technol. 19 (4) (2015) 178–181.

[37] M. Shirakawa, K. Nakayama, T. Hara, S. Nishio, Wikipedia-based semantic similarity measurements for noisy short texts using extended Naive Bayes, IEEE Trans. Emerg. Topics Comput. 3 (2) (2017) 205–219.

[38] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. ArXiv Preprint ArXiv:1409.1556.

[39] A. Sun, S.S. Bhowmick, Quantifying tag representativeness of visual content of social images, in: Proceedings of the 18th ACM international conference on Multimedia, ACM, 2010, pp. 471–480.

[40] A. Sun, S.S. Bhowmick, K.T. Nam Nguyen, G. Bai, Tag-based social image retrieval: an empirical evaluation, J. Am. Soc. Inf. Sci. Technol. 62 (12) (2011) 2364–2381. https://doi.org/10.1002/asi.21659.

[41] T. Uricchio, M. Bertini, L. Seidenari, A. Bimbo, Fisher encoded convolutional bag-of-windows for efficient image retrieval and social image tagging, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 9–15.

[42] W. Voravuthikunchai, B. Crémilleux, F. Jurie, Image re-ranking system based on closed frequent patterns, Int. J. Multimed. Inf. Retr. 3 (4) (2014) 231–244. https://doi.org/10.1007/s13735-014-0066-8.

[43] C.H. Wang, D. Han, Sentiment analysis of micro-blog integrated on explicit semantic analysis method, Wirel. Pers. Commun. (1079) (2018) 1–11.

[44] L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma, S. Li, Flickr distance, in: Proceedings of the 16th ACM international conference on Multimedia, ACM, 2008, pp. 31–40.

[45] L. Wu, Y. Wang, J. Shepherd, X. Zhao, Max-sum diversification on image ranking with non-uniform matroid constraints, Neurocomputing 118 (2013) 10–20. https://doi.org/10.1016/j.neucom.2013.02.008.

[46] L. Xia, J. Xu, Y. Lan, J. Guo, X. Cheng, Learning maximal marginal relevance model via directly optimizing diversity evaluation measures, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pp. 113–122.

[47] Y. Yan, G. Liu, S. Wang, J. Zhang, K. Zheng, Graph-based clustering and ranking for diversified image search, Multimed. Syst. 23 (1) (2017) 41–52.

[48] K. Yang, M. Wang, X.S. Hua, H.J. Zhang, in: Tag-Based Social Image Search: Toward Relevant and Diverse Results, Social Media Modeling and Computing. Springer, London, 2011, pp. 25–45.

[49] H. Zhang, X. Shang, H. Luan, M. Wang, T.-S. Chua, Learning from collective intelligence: feature learning using social images and tags, ACM Trans. Multimed. Comput. Commun. Appl. 13 (1) (2017) 1.

[50] J. Zhang, Y. Yang, Q. Tian, L. Zhuo, X. Liu, Personalized social image recommendation method based on user-image-tag model, IEEE Trans. Multimedia 19 (2017) 2439–2449 PP(99), 1.

Wei Lu, Wuhan University, China, Wei Lu received his PhD degree in information science from Wuhan University. He is now a professor in the School of Information Management, Wuhan University. His research interests include information retrieval, text mining, and knowledge management.

Mengqi Luo, Wuhan University, China, Mengqi Luo received her MSc degree in information technology management from Hong Kong Baptist University, Hong Kong, in 2012. She is currently a PhD candidate in the School of Information Management, Wuhan University. Before her PhD study, she held research positions in several universities in Hong Kong. Her research interests include image retrieval, natural language processing, and text mining.

Zhenyu Zhang, Product Manager at Tencent, China, Zhenyu Zhang received his BSc degree in information management and information systems from Wuhan University in 2014 and his MSc degree in management science and engineering from Wuhan University in 2017. He is currently a product manager at Tencent. He studied the semantic diversity of images during his university research.

Guobiao Zhang, Wuhan University, China, Guobiao Zhang received his MSc degree in information science from Shanxi University of Finance and Economics, 2017. He is currently a PhD student in the School of Information Management, Wuhan University. His research interests include image retrieval, deep learning, and natural language processing.

Heng Ding, Wuhan University, China, Heng Ding received the MSc degree from Wuhan University, China in 2013. He is currently a Ph.D. candidate in the School of Information Management, Wuhan University. His research interests include information retrieval, data mining and deep learning.

Haihua Chen, University of North Texas, USA, Haihua Chen received his BSc degree in information management from Central China Normal University, Wuhan, in 2014 and his MSc degree in information science from Wuhan University in 2017. He is currently a PhD student specializing in data science in the Department of Information Science at the University of North Texas. His research interests include information retrieval, digital libraries, and recommendation systems.

Jiangping Chen, University of North Texas, USA, Jiangping Chen earned her doctorate from Syracuse University. She is now a professor in the Department of Information Science at the University of North Texas. Her research interests include information retrieval, digital libraries, and data science.