CrossMark

Exploiting label semantic relatedness for unsupervised image annotation with large free vocabularies

Luis Pellegrin[1] (ORCID) · Hugo Jair Escalante[2] · Manuel Montes-y-Gómez[2] ·
Fabio A. González[3]

Abstract

Automatic Image Annotation (AIA) is the task of assigning keywords to images, with the aim to describe their visual content. Recently, an unsupervised approach has been used to tackle this task. Unsupervised AIA (UAIA) methods use reference collections that consist of the textual documents containing images. The aim of the UAIA methods is to extract words from the reference collection to be assigned to images. In this regard, by using an unsupervised approach it is possible to include large vocabularies because any word could be extracted from the reference collection. However, having a greater diversity of words for labeling entails to deal with a larger number of wrong annotations, due to the increasing difficulty for assigning a correct relevance to the labels. With this problem in mind, this paper presents a general strategy for UAIA methods that reranks assigned labels. The proposed method exploits the semantic-relatedness information among labels in order to assign them an appropriate relevance for describing images. Experimental results in different benchmark datasets show the flexibility of our method to deal with assignments from free-vocabularies, and its effectiveness to improve the initial annotation performance for different UAIA methods. Moreover, we found that (1) when considering the semantic-relatedness information among the assigned labels, the initial ranking provided by a UAIA method is improved in most of the cases; and (2) the robustness of the proposed method to be applied on different UAIA methods, will allow extending capabilities of state-of-the-art UAIA methods.

Keywords Unsupervised image annotation · Label relevance · Semantic-relatedness estimation

1 Introduction

In the last decade, technology development has facilitated the massive up-growth of digital information allocated in the Web, e.g. images, texts, audio, video, etc. Referring to images, automatic image annotation (AIA) is the task dedicated to study and develop methods

✉ Luis Pellegrin
luis.pellegrin@uabc.edu.mx

Extended author information available on the last page of the article.

Springer

capable of assigning keywords to images. Thereby, the images can be indexed by their visual contents. Traditionally, AIA methods follow a supervised approach [11, 33], i.e., they use a training dataset of labeled images in order to build label-predictive models. Then, learned models serve to assign labels to new images. However, using a supervised approach requires huge datasets with labeled images, in this regard the dataset construction could be very expensive, e.g. see ImageNet [4, 33].

UAIA methods do not require of labeled images for the training phase, instead they use reference collections of documents containing images, which can be gathered from diverse sources, e.g. the Web, news, articles, etc. Therefore, building reference collections is a cheaper process than constructing datasets with labeled images, e.g. see [1, 24, 36]. Although, these images lack of assigned labels, in most cases there is text around them that is related, and therefore some words can be considered as soft-labels (e.g. see [8, 24, 26, 43]). UAIA methods rely heavily on text mining techniques that are applied to the textual part of the documents in order to extract words for image annotation. Any extracted word can be used for annotation under this setting and, according to its relevance (generally using a score), for describing the visual content of the image it is assigned.

Under UAIA approach, different strategies have been adopted for annotating images by means of associating the text to the visual information, showing competitiveness against the supervised approach [28]. A common approach for UAIA uses the image to be annotated as a query, compare it with images from the reference collection, and then process the text information associated with the retrieved images [2, 10, 21, 28, 35, 39, 44, 46]. Alternative methods process the whole reference collection offline in order to identify visual-textual associations and then use these associations for performing the annotation [28].

Nevertheless, due to the fact that potentially any word extracted from the reference collection can be used as label, the actual UAIA methods present limitations when working with large vocabularies. On the one hand, the text in reference collections is noisy, in the sense that text is not explicitly associated with images, e.g. consider a blog's entry with the text 'vacations on the beach...' accompanied by an image with people in a hotel's front door but no a beach (see Fig. 1). Also, text is not restricted to a topic, thus, words can refer to anything, including information that is not related to visual content of images, e.g. consider a blog's entry 'whip an egg and cover the pastry...' accompanied by an image with a cake (see Fig. 2). Hence, the diversity of labels in UAIA increases (when compared to supervised variants), but at the same time, it also increases the complexity of determining the relevance of arbitrary labels for images, causing a larger number of inconsistent annotations that can be considered as wrong labeling.

With the aim of reducing the number of wrong annotations and mitigating the inadequate relevance scores during the label assignment, we propose a reranking method that

Fig. 1 Blog's entry: vacations in the beach...

Fig. 2 Blog's entry: whip an egg
and cover the pastry....

improves the initial UAIA annotation by including the semantic-relatedness information
among labels. The idea of this method is based on the assumption that the images are anno-
tated with the main topic that is defined by a consensus among a subset of candidate labels.
Consequently, the labels are reranked by updating their relevance scores with respect to the
topic. When an image is annotated, a list of candidate labels is generated where the positions
of the labels indicate their relevance. For each pair of candidate labels, there is a relationship
that can be defined by its semantic relatedness, where labels with similar meanings or that
appear in similar contexts have a strong relationship. Our method exploits this semantic-
relatedness information performing an enhancement on the initial ranking, in order to favor
those labels that are *consistent*[1] for the annotation. In this way, our method could be seen as
a module that post-processes the initial labels by assigning an adequate relevance score to
candidate labels, even when large vocabularies are used.

For example, in Fig. 3 a scene in the sea is considered, some of the assigned labels are
'canker', 'turtle', 'globule', 'angelfish', 'reef', 'coral', 'nebular', 'anchovies', 'ocean', etc.
After applying our method, the most relevant[2] labels are positioned at the top. Besides,
labels that have a soft semantic-relatedness with the rest, no matter how high is its initial
rank will be considered inconsistent and therefore moved to the bottom.

1.1 Research aims and contributions

The aim of this work is to refine the initial labeling provided by UAIA methods by means
of exploiting semantic-relatedness information among labels. With this aim in mind, we
propose a methodology that takes as input the labels provided by an UAIA method, and
updates the score assigned to each label, based on two aspects: the initial score provided
by the UAIA method and the semantic-relatedness estimated from contextual information
among the other labels assigned to the same image.

The main contributions of this work are as follows:

---

[1]By *consistent* labels, we mean those which have high semantic-relatedness with the majority of the
annotated labels.

[2]Where the relevance is guided by the labels that have the highest cohesion on their semantic-relatedness
information.

Labels assigned
1. canker
2. turtle
3. globule
4. angelfish
5. reef
6. coral
7. nebular
8. anchovies
9. ocean
10. lachrymose
11. starfish
12. seabed
13. sauce
14. sky
15. rock

after of exploiting semantic-relatedness information >>

Final annotation
1. turtle
2. angelfish
3. reef
4. coral
5. starfish
6. ocean
7. anchovies
8. seabed
9. rock
10. nebular
11. globule
12. canker
13. lachrymose
14. sky
15. sauce

Fig. 3 Illustrating an image, labels annotated by UAIA method before and after considering the semantic-relatedness information

1. We introduce a novel and effective strategy for reranking the initial labeling provided by a UAIA method based on the semantic-relatedness information among labels. We consider an annotation scenario with free annotation vocabularies (of thousands of label terms), showing the effectiveness of the method for improving initial annotations by leveraging semantic-relatedness information among labels.

2. We provide an extensive experimental evaluation of our reranking strategy applied to UAIA methods based on different annotation strategies. In this regard, experimental results are reported on two different benchmark datasets showing that our reranking method is robust and flexible to extend UAIA methods and improve their annotation performance in most cases, regardless of the annotation strategy and considered dataset.

The implications of our findings are: (1) at the theoretical level, we critically reviewed UAIA methods, exposed some issues in the label assignment process, and explored different ways for measuring semantic information; and (2) at the practical level, we aimed to accurately identify the impact of an adequate label assignment. In this regard, we proposed a methodology and evaluated it on real world images and using the state-of-the-art UAIA methods.

2 Related work

Given an input image, the goal of UAIA is to assign labels to describe its visual content. UAIA methods rely on a reference collection of textual documents containing images, where, usually, a piece of text is associated with each image. Due to the fact that labels are extracted from the collection's textual information, they are not restricted to a closed set (i.e., any word as label can be used). Under these conditions, traditional UAIA methods assign scores to every word, where these scores determine their relevance to describe its visual content.

One of the first UAIA approaches was based on the local paradigm [28]. Given an image to annotate, a Content-Based Image Retrieval (CBIR) method is used to find, in the reference collection, those images with similar visual content. In a second stage, a text mining technique is applied to the text associated with the retrieved images in order to determine the set of labels that are used to describe the visual content of the image. In the CBIR stage, each image is represented by visual descriptors, where each image is defined by a vector that concentrates its distinctive characteristics [9, 25]. Several publications have found that images with similar visual content, tend to contain related information in their associated texts [8, 16, 26, 38, 43, 51]. Whereas in the second stage, each text is represented by its content, e.g. using bag-of-words [34]. Methods adopting this formulation can be found in [2, 10, 21, 28, 35, 39, 44, 46].

On the other hand, a global method for a UAIA task has been recently proposed with the aim of exploiting the entire reference collection [28]. This method is based on multimodal co-occurrences that are produced between visual descriptors and textual-words. The idea is to build a visual prototype for each word extracted from the textual information in the reference collection. In this way, a visual representation of a query image can be compared directly with the words' visual prototypes. The words associated with visual prototypes with the highest similarity to the query image are used for labeling images. The visual prototypes are obtained through a multimodal indexing. The multimodal indexing is calculated by an efficient matrix product that uses two views of the documents in the reference collection: documents represented by bag-of-words, and documents represented by bag-of-visual-words. See [28] for more details.

For dealing with the diversity of words that can be extracted and used for labeling images, some UAIA methods restrict annotations to a reduced set of labels. The idea is to annotate the images by considering a limited number of labels but using the whole textual information that can be extracted. In this sense, the method in [46] uses a strategy based on co-occurrences between labels from the reduced set and the free vocabulary; those words from the vocabulary that mostly co-occur with the labels help to aggregate its relevance score to describe the image. In the same direction, several works have used lexical resources such as WordNet [23][3] to group words with similar meanings; then a label that subsumes all others can be used to describe the content of an image [2, 35, 39, 44]. Besides, some methods just assign the most frequent terms obtaining competitive results [21, 28]. Also, a weighted combination has been explored that takes into account the absence/presence of the labels to assign, e.g. see [10].

The previous works exploit the diversity of potential words that can be annotated in order to determine the relevance of the set of labels that describe the content of images [2, 35, 39, 44, 46]. In contrast, the proposed method searches to improve the performance of the UAIA methods by taking advantage of the diversity of words, incorporating a reranking strategy for assigning an adequate relevance to the labels.

As far as we know, none of the UAIA methods in the state of the art have explicitly made the use of semantic-relatedness information among labels in order to refine the image annotation. However, similar strategies have been used previously for other tasks, such as information retrieval [12–15, 17, 26, 27, 30, 32, 45], where query expansion has been widely explored in order to diversify the retrieval or integrate semantic information; and supervised image annotation [3, 5, 7, 19, 20, 31, 40–42], where visual information from images

---

[3]A lexical database for English where the words are interlinked to each others by specific senses as conceptual-semantic and lexical relations.

and textual information from labels are used as complementary to each other. However, under a supervised scenario, annotating images is limited to a number of possible labels, which defines a controlled scenario to assign relevance to labels. In contrast, the reranking method was defined to work especially with unrestricted free vocabularies for performing annotation.

3 Exploiting semantic-relatedness information in UAIA

In this section, we introduce our reranking method called $+\mathcal{SR}$ that stands for 'plus semantic-relatedness', exemplifying its capabilities in order to improve the image annotation. As we previously mentioned, UAIA methods use words extracted from reference collections for labeling, facing the difficulty of assigning a correct relevance to labels. In this regard, the $+\mathcal{SR}$ method can be seen as an additional module in the annotation process. This is illustrated in Fig. 4, where it takes a ranked list of the (visually relevant) candidate labels, and reranks this list by considering their semantic-relatedness information.

The aim of the $+\mathcal{SR}$ method is to provide an adequate relevance that favors consistent labels and reduces the relevance of inconsistent labels, provided all the considered labels have some visual relevance for the image to be labeled. The idea behind this method is to allow related labels to help each other, reassigning them a higher score. Besides, labels that have no relationship will become inconsistent, reassigning them lower scores. For instance, consider an ordered list of words that includes 'screwdriver', 'monkey', 'lion' and 'elephant'. In this case, words referring to animals help each other reaching top positions, and 'screwdriver' goes down to the last position.
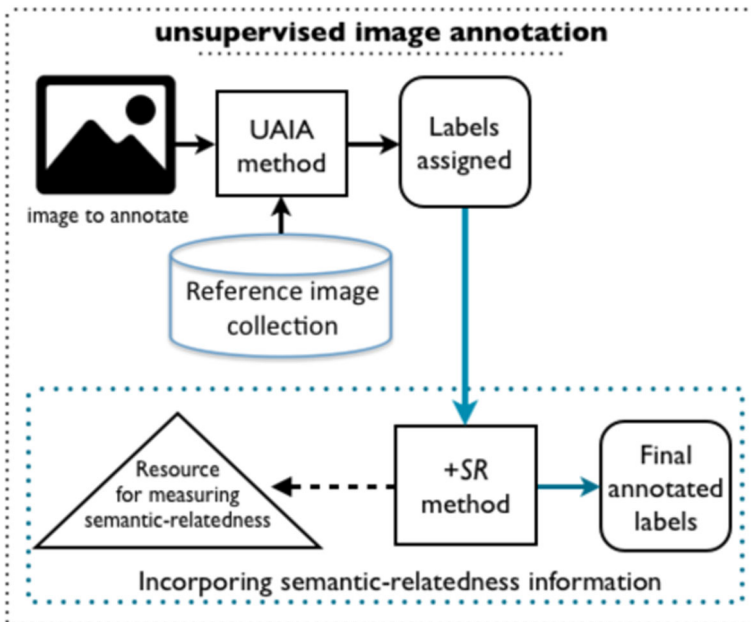

Fig. 4 General architecture of the UAIA task integrating the $+\mathcal{SR}$ method

In the next subsection we describe how the $+\mathcal{SR}$ method incorporates the semantic-relatedness information among labels in order to provide a better relevance in the scores of the assigned labels.

3.1 The $+\mathcal{SR}$ method

The input of the $+\mathcal{SR}$ method is a list of n labels with a determined rank that is induced by their assigned score, and its output is a list with the same number of elements but with a different rank. For producing this new ranking, each label is re-scored, where its new score depends on a combination of its original score and its semantic-relatedness with a subset of candidate labels from the same list. The subset of candidate labels consists of top λ labels from the list, or simply using the entire list. Intuitively, if only few candidate labels are considered, it is assumed that these were correctly assigned by the UAIA method, then the re-scoring is influenced by them.[4] Thus, when all labels are used, it follows the idea that the image has been annotated with a topic. In this case, the topic is defined by the highest semantic cohesion[5] among labels. Formally, the semantic-relatedness information used by the $+\mathcal{SR}$ method is defined by means of word similarities. Given a list of labels $\{w_1, w_2, \ldots, w_n\}$, for each label w_i its relevance is reassigned by:

$$+\mathcal{SR}(w_i) = (1 - \lambda)score(w_i) + \lambda \left[\frac{1}{\gamma} \sum_{j=1}^{\gamma} (srs(w_i, w_j) \cdot score(w_i)) \right] \qquad (1)$$

where λ is a parameter that regulates the importance between the initial annotation (denoted by $score(w_i)$) and its semantic-relatedness with a subset of candidate labels. Besides, the γ parameter regulates the number of candidate labels that are considered for calculating the semantic-relatedness. Finally, a function $srs(w_i, w_j)$ defines the semantic-relatedness similarity between labels w_i and w_j, this estimate can be obtained by at least two different approaches:

– **Knowledge-based**. Under this approach, the semantic-relatedness of two words (labels) is given by measuring their proximity using a linguistic resource (e.g, a semantic network). Usually, these measurements are estimated according to lexical resources, such as WordNet [23]. WuP [50] is a classic measure under this approach, this calculates relatedness using the depths of two words by counting their number of nodes, along with the depth of the least common subsumer (lcs):

$$WuP(w_i, w_j) = \frac{2 * depth(lcs(w_i, w_j))}{len(w_i, w_j) + 2 * depth(lcs(w_i, w_j))}$$

where w_i and w_j are two synsets in WordNet, $len(w_i, w_j)$ is the length of the shortest path from synset w_i to synset w_j, $lcs(w_i, w_j)$ is the most specific concept which is an ancestor of w_i and w_j, and $depth(w_i)$ is the length of the path to synset w_i from the global root entity where $depth(root) = 1$. Other measurements can be revised in [37].
– **Statistics-based**. In this approach large-document collections are exploited in order to embed words in a vector (semantic) space where similarities between words can be estimated. For instance, Word2vec [22] uses either of two model architectures to produce

[4]Similar criterion as the one used in blind relevance feedback, in which the n top documents are considered to be relevant to the query.
[5]Referring to those labels that share strong semantic-relatedness information.

a distributed representation of words. Both architectures can be used to produce word embeddings by reconstructing linguistic contexts of words. As input a large corpus (e.g., Wikipedia) is taken, and as output a vector for each word is produced. The idea is to represent through similar vectors those words that have a similar meaning because they share a similar context. Once the words are represented by vectors, a common way to calculate the semantic relatedness between any two word vectors is by their cosine similarity:

$$cosine(w_i, w_j) = \frac{\mathbf{w}_i \cdot \mathbf{w}_j}{||\mathbf{w}_i||\,||\mathbf{w}_j||}$$

The values of $srs(w_i, w_j)$ require to be normalized as they account for similarities among words that must be comparable to each other. However, the values of $score(w_i)$ are not required to be normalized, this is because it appears in both terms of (1). Besides, since $score(w_i)$ is in fact the output of the UAIA system, we do not impose any constrain on the values it may take.

Note that in the application of $+\mathcal{SR}$, the most expensive process is the calculation of the distances among words. However, this process is performed offline once. Hence, performing the $+\mathcal{SR}$ does not consume a significant amount of time. The complexity of estimating distances is of $\Theta(n^2)$ where n is number of words whereas the complexity of applying the $+\mathcal{SR}$ is $\Theta(n)$. In our experiments it takes around 5 minutes to test all possible combinations of hyperparameters when using 1,000 images as input, about 3 seconds per configuration. In the following by means of an example, we analyze the label positions before and after applying the $+\mathcal{SR}$ method.

3.2 Running example

In this example, the image to be annotated is shown in Fig. 5. Then using the global method proposed in [28], 50 labels have been annotated using a free-vocabulary, see Fig. 6.

Fig. 5 Image example

Ground truth: *daytime, grass, motorcycle, outdoor, person, road, sport, vehicle.*
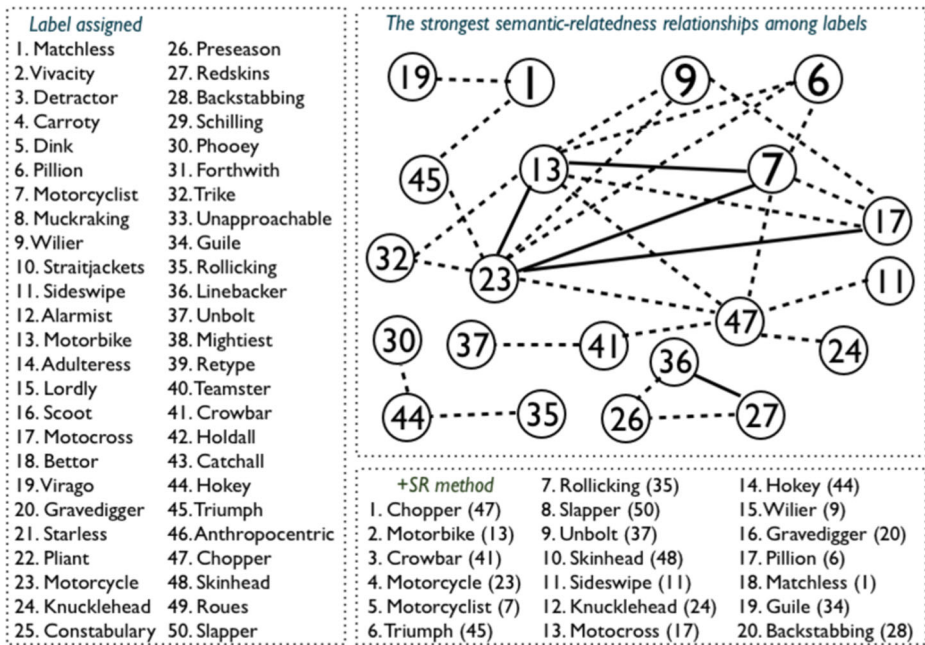
Fig. 6 The left plot shows the list of the 50 keywords provided as initial annotation by the UAIA method. The right-topmost plot shows the graph of similarities, whereas the right bottom plot shows the new rank obtained by the $+\mathcal{SR}$ method

The $+\mathcal{SR}$ method takes the list of labels assigned by the global method for generating a new ranking. We set $\gamma = 50$ and $\lambda = 1.0$ in the $+\mathcal{SR}$ method, that is, 50 labels from the list are considered for updating the rank, and the reranking of each word is estimated by the semantic-relatedness between this word with the rest weighted by the confidence of its initial assignment.

For the purpose of analyzing the reason of the new rank produced by the $+\mathcal{SR}$ method, Fig. 6 presents a graph of similarities from the initial list of labels. We have focused on only the strongest relationships among labels, where dashed lines are those that surpass a threshold of 0.5 in a scale of 0 to 1 in similarity, as estimated using the formula above, and solid lines when semantic-relatedness surpasses similarity 0.7.

Finally, the $+\mathcal{SR}$ method uses the strongest semantic-relatedness relationships to update the ranking of assigned labels, where numbers between parentheses indicate the initial position of labels. For instance, we can see that 'chopper' initially ranked in the 47 position, improves considerably its position due to gather many relationships with other labels, e.g. labels 41, 24, 11, 7, 13 and 23.

From Fig. 6, we can see that after applying the $+\mathcal{SR}$ method it has been possible to estimate a better relevance into the assigned labels, showing more related words with the visual content. In this case, the $+\mathcal{SR}$ method has been capable of bringing up relevant labels from the last positions of the initial list.

4 Experimental settings

This section introduces the datasets, baseline UAIA methods and evaluation measures to be used in Section 5.

4.1 Datasets

We have considered test sets from two different benchmarks: ImageCLEF 2013 [48] and SAIAPR TC-12 [6] datasets. The first dataset was created with the aim of being exclusively used for UAIA methods, it provides a reference collection with 250,000 documents. Each document is composed of an image, and a web page that provides textual information. The test set was recently released [47], it contains the same 1,000 images evaluated in the development set plus 1,000 new images annotated with at least one of 116 different concepts. The latter dataset, has a test set formed by 4,000 images that are annotated with at least one of 276 different concepts. In order to reduce redundancy in this dataset, we have considered 256 of the 276, where we have pruned those that are redundant e.g. the concept 'vehicle-with-tires' is subsumed by 'vehicle'.

4.2 Baseline UAIA methods

We have evaluated the $+\mathcal{SR}$ method in combination with two different UAIA methods, namely the local and global approaches introduced in [28]. Mainly, we have compared their performance before and after applying the $+\mathcal{SR}$ method. Although, the global method presents a better performance than the local method, we have decided to report results with both methods. Our aim is to show the flexibility and effectiveness of the $+\mathcal{SR}$ to deal with and improve both annotation methods. Both methods can use as many labels as are available in the extracted vocabulary.

4.3 Evaluation measures

The evaluation was carried out using the evaluation framework for UAIA methods introduced in [29]. Under this evaluation framework, it is possible to evaluate any label allowing to consider free-vocabularies for annotating images. It shows some variants of classic metrics such as recall, precision, F1 and MAP that take into account the semantic similarities between the assigned labels and the labels in the ground truth. The adapted metrics rely on an α threshold that establishes how close the semantic relation is between the assigned labels and the ground truth. Thus, even though there is no exact match between words, it is possible to calculate a score that measures how strong their semantic relationship is. For instance, considering a scene with a 'cat' where the ground truth is 'cat', different methods could annotated 'feline', 'kitten', 'mouser', 'pussycat', etc. obtaining similar results as a method that annotates 'cat'. In this concern, an $\alpha = 0.6$ has been selected because it has shown a good commitment in previous works. For instance, consider the word 'orange' and its similarity with other fruits: 'apple'= 0.37, tomato= 0.24, 'pear'= 0.37, 'grapes'= 0.37, 'tangerine'= 0.9. In this case only 'tangerine' is close to 'orange', in both color and citrus family. Please note that differently from [29], in which an evaluation framework was introduced, herein we propose a novel image reranking method that, based on the semantic relatedness, improves the initial labeling provided by an UAIA.

The results are reported using two evaluation metrics: $F1_\alpha$ measure and MAP_α. The $F1_\alpha$ allows us to compare the improvements summarizing precision and recall as harmonic mean:

$$F1_\alpha = 2 \cdot \frac{R_\alpha \cdot P_\alpha}{R_\alpha + P_\alpha}$$

On the one hand, R_α is an adapted version of recall which measures the coverage. Intuitively, the measure counts each label from the ground truth that have been covered. Reaching a maximum score implies that for each word from the ground truth, there exists a matching with a word from the system output, such that this matching is greater than α:

$$R_\alpha = \frac{\sum_{w_i \in gt} 1_{(\exists w_j : w_j \in outs \wedge sim(w_i, w_j) \geq \alpha)}}{|gt|}$$

where gt refers to words from ground truth, and $outs$ are the output words given by the annotation system, $sim(w_i, w_j)$ expresses the semantic similarity between two words (this can be calculated by approaches described in Section 3.1).

On the other hand, P_α is the adapted version of precision which measures the relevance. Here, the effectiveness of the output words is given by the number of words that surpass the α threshold:

$$P_\alpha = \frac{\sum_{w_j \in outs} 1_{(\exists w_i : w_i \in gt \wedge sim(w_i, w_j) \geq \alpha)}}{|outs|}$$

In addition to the $F1_\alpha$, we consider the MAP_α. This measure allows to compare the improvements in the induced order of the labels. Similar to the $F1_\alpha$, it is an adaptation of the classic MAP measure, but it allows to consider labels without an exact match:

$$AP_\alpha = \frac{\sum_{k=1}^{|outs|} P_\alpha(k)}{|gt|}$$

where k is the word to evaluate and

$$P_\alpha(k) = \frac{\sum_{j=1}^{k} 1_{(\exists w_i : w_i \in gt \wedge sim(w_i, w_j) \geq \alpha)}}{|k|}$$

For computing the word similarities used by these measures, we applied the statistics-based approach described in Section 3.1. For more details, we refer the readers to [29].

4.4 Additional considerations

– Experimental results reported on the ImageCLEF 2013 dataset use as training set the data from the same dataset. On the other hand, SAIAPR TC-12 only provides a training set with labeled images which cannot be used because UAIA methods were developed with the aim to work on an unsupervised scenario. Therefore, we used the training subset of the ImageCLEF 2015 [49] dataset to evaluate the test set of SAIAPR TC-12. ImageCLEF 2015 dataset was built in similar way as the ImageCLEF 2013, with the difference that the dataset from 2015 contains the double of documents.

– For the reported results on the next section, we used the statistics-based approach for applying the $+\mathcal{SR}$ method, where Wikipedia serves as resource for obtaining the learned vectors that are used to calculate the semantic-relatedness similarities. In this sense, all distances among words were normalized between [0,1] before been used by the $+\mathcal{SR}$ method. We also performed experiments with the knowledge-based approach in the Section 5.4.

5 Experimental results

For evaluating each image, we have considered as input a list of 50 labels assigned by the UAIA methods. The $+\mathcal{SR}$ method receives this list, then it produces a new list with the same number of elements but with different positions. The two lists are compared using several evaluation measures as described in Section 4.3.

5.1 Results on ImageCLEF 2013 dataset

Figure 7 shows the $F1_\alpha @n$ results, where the x-axis reports the results obtained by evaluating the performance in intervals of 5 labels in the annotation. For both UAIA methods, the solid line reports their (original) scores. Both use different strategies for labeling images. Whilst the local method only uses the information of a subset of retrieved images to perform annotations, the global method uses the whole information contained in the reference collection. In this regard, the local method tends to lower its performance when the number of labels increases. This indicates that this method annotates its best labels at the very first positions. In contrast, the global method tends to rise its performance when the number of labels increases, retrieving a larger number of relevant labels.

Besides, the box plots report the results obtained by $+\mathcal{SR}$ method considering the different parameter settings. Here, each box plot concentrates 100 results derived by all possible combinations between λ and γ, being $\lambda = 0.1, 0.2, \ldots, 1.0$ and $\gamma = 5, 10, \ldots, 50$. A large box plot indicates that the results obtained by the $+\mathcal{SR}$ method have a large variation. In general, we can see an improvement of the $+\mathcal{SR}$ method over the scores obtained by the local and global methods, being more evident the improvements achieved over the global method. This indicates that the $+\mathcal{SR}$ method brought relevant labels to the first positions. The improvements over the global method reach an upper bound reporting an annotation around 20 labels.

In order to show which values of the λ and γ parameters are the best, Fig. 8 shows a grayscale map of the improvements. A white box indicates the highest improvement, whereas a black box indicates zero improvement. For the local method it is better considering a subset of 15–25 candidate labels for calculating the new rank. In contrast, for the global method it is better to use the entire list for calculating the rank, this indicates that most labels from the list are related. On the other hand, higher values of λ are advantageous for both methods.
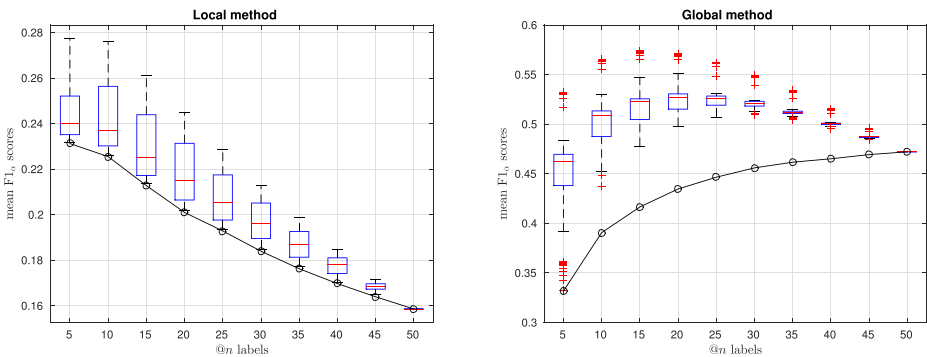

Fig. 7 $F1_\alpha$ results in the ImageCLEF 2013 dataset obtained by local and the global methods (in solid line) against the $+\mathcal{SR}$ method (in boxes)
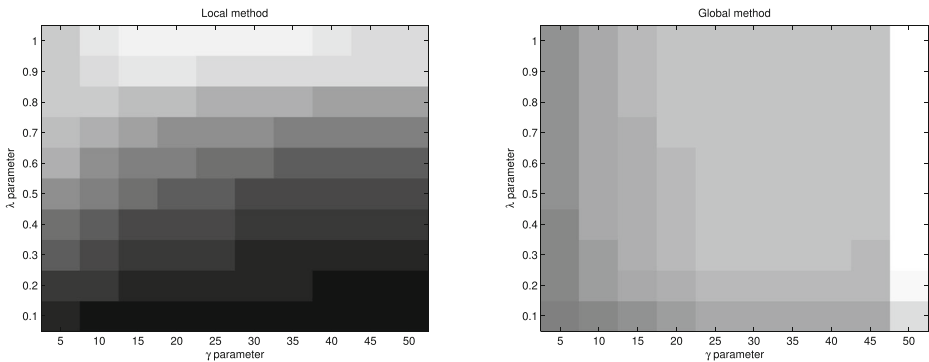
Fig. 8 Average improvements obtained in the ImageCLEF 2013 dataset by considering different parameter settings

Figure 9 reports the MAP_α results, where the performance is reported on the entire list of 50 labels. From Fig. 9, we can see that the $+\mathcal{SR}$ method achieves a better performance over the global method. For the local method only few parameter combinations improved the performance.

Finally, in order to show how well the obtained results are after applying the $+\mathcal{SR}$ method, a statistical test was performed on the MAP_α measure (the t-student test with a confidence of 99%). The improvements that are achieved over the global method are statistically significant, on the contrary, the results achieved over the local method are not statistically significant. However, in both UAIA methods improvements were obtained over the original performance.

5.2 Results on SAIAPR TC-12 dataset

The experiments reported in the previous subsection are replicated on the SAIAPR TC-12 dataset. These experiments aim to provide more evidence of the effectiveness of the $+\mathcal{SR}$ method. Note that this time, the image annotation is performed over a test set from SAIAPR TC-12 dataset, but the used training set comes from ImageCLEF-2015 collection. Therefore, different capabilities of the $+\mathcal{SR}$ method are evaluated, on the one hand, its
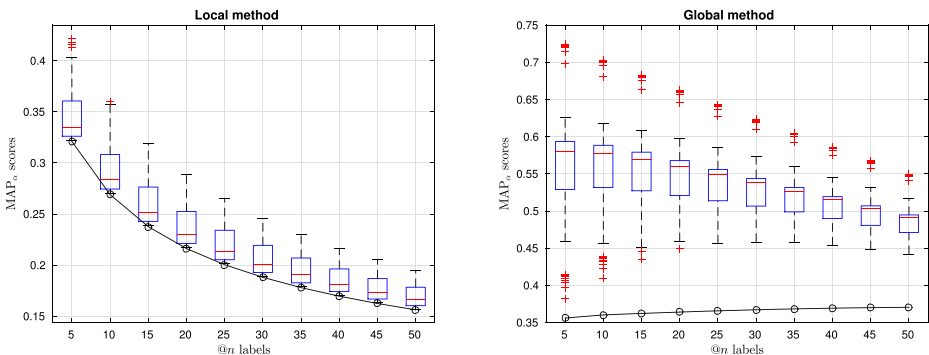

Fig. 9 MAP_α results in the ImageCLEF 2013 dataset obtained by the local and the global methods against against the $+\mathcal{SR}$ method
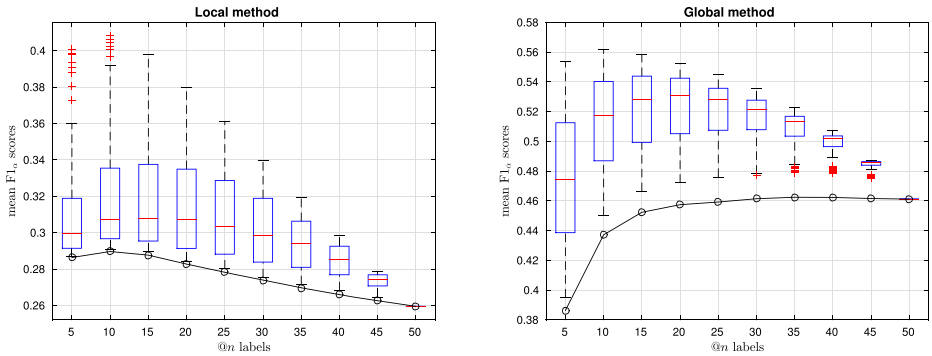
Fig. 10 $F1_\alpha$ results in the SAIAPR TC-12 dataset obtained by local and the global methods (in solid line) against the $+\mathcal{SR}$ method (in boxes)

flexibility to be adapted to different UAIA methods and also to new domains, and, on the other hand, its effectiveness to improve the results under these challenging conditions.

Figure 10 shows the results corresponding to $F1_\alpha$ measure for both annotation methods compared to the $+\mathcal{SR}$ method. Again, the results obtained by using the $+\mathcal{SR}$ method are better than those obtained by the local and global methods. The local method obtains better results considering few labels, thus the improvements obtained by the $+\mathcal{SR}$ method follow a normal distribution. Instead, the global method obtained better results when considering more labels, in this case the improvements obtained by the $+\mathcal{SR}$ method follow the same behavior reducing the variations among results when more labels are considered.

Figure 11 shows the improvements obtained by different values of the λ and γ parameters. For this case, the best improvements are achieved by using the whole list to perform reranking. Referring to the confidence between the original score and the new score, by leveraging semantic-relatedness information among labels is possible achieves the best performance.

Figure 12 presents the results of the MAP_α. Again, the $+\mathcal{SR}$ method achieves improvements against to both annotation methods. Nevertheless, the improvements achieved over the global method show larger variations among the different combinations of the λ and the
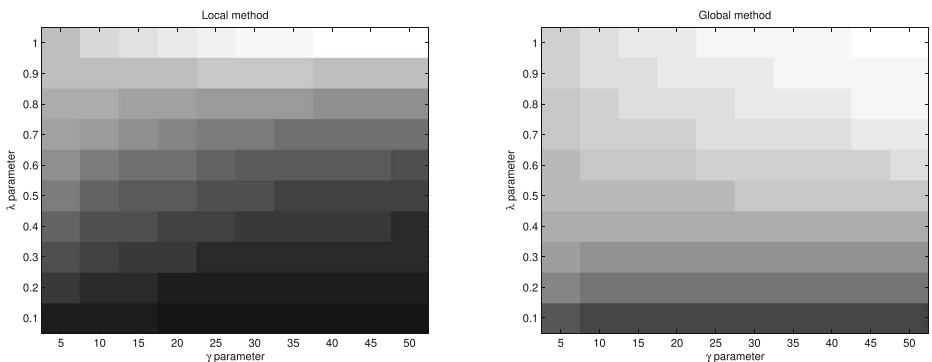

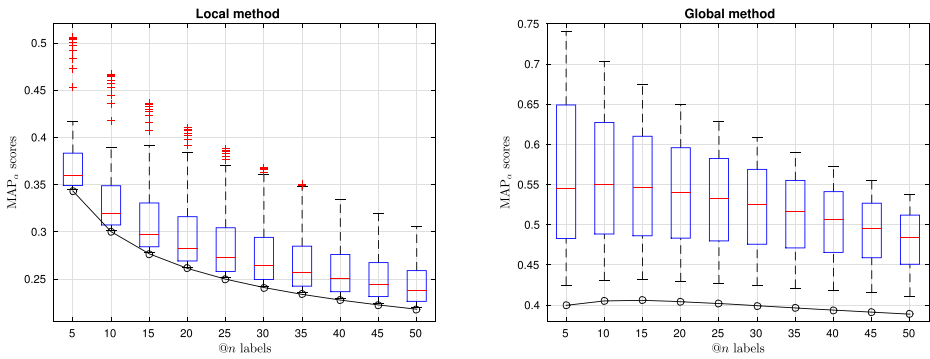**Fig. 11** Average improvements obtained in the SAIAPR TC-12 dataset denoted by percentages

Fig. 12 MAP_α results in the SAIAPR TC-12 dataset obtained by local and the global methods against the $+\mathcal{SR}$ method

γ terms. This time, most of the parameter combinations have achieved improvements over the local method.

A statistical test on the MAP_α measure was performed (the t-student test with a confidence of 99%). The improvements achieved over the global method are statistically significant, meanwhile, only the improvements with a $\lambda \geq 0.6$ for any γ value over the local method are statistically significant.

5.3 Discussion

Summarizing the obtained results, we have shown that the $+\mathcal{SR}$ method is able to improve the annotation performance of different UAIA methods by considering the semantic-relatedness information among labels, which allows to assign a better relevance to the labels. We have evaluated two different paradigms of UAIA, from the obtained results it is possible to observe two different kinds of behaviors: (1) considering F1, both UAIA methods obtained improvements by the reranking approach. This was expected because when considering a large number of labels, the correct ones tend to support each other and, therefore, to increase their final ranking; (2) considering the MAP, we found that redundancy among the labels is helpful, e.g. in Figs. 9 and 12, the global method obtains considerable improvements; in contrast, the local method, which usually gathers the correct annotation(s) at the top of the list, shows a decrease in annotation performance when considering a larger number of words.

On the other hand, the correct tuning of α and γ hyperparameters has an associated computational cost. Nevertheless, this time is negligible: it takes around 5 minutes to test all of the possible combinations of hyperparameters when using 1,000 images as input. In addition, we showed that for different collections, the explored models exhibit a good performance for a wide range of combinations of these parameters. Our observation is that these hyperparameters, once they are adjusted for a method given a dataset, they can be used in other datasets obtaining acceptable performance, although not necessarily the optimal.

5.4 Robustness under different approaches for calculating semantic-relatedness

Note that so far, the $+\mathcal{SR}$ method and the evaluation metrics use the same approach for calculating semantic-relatedness similarities (i.e. statistics-based approach in Section 5). In this

subsection, we ask ourselves the question *'is it better to use a statistics-based approach or a knowledge-based approach in the $+\mathcal{SR}$ method?'*. On the one hand, a knowledge-based approach uses a resource manually built, which although is limited to certain number of words, these words have carefully assigned different senses, e.g. 'bank' could refer to financial institution or sloping land. On the other hand, the statistics-based approach considers a greater number of words, but during the representation of these words, all senses are combined.

Figure 13 puts in comparison the statistics-based approach and the knowledge-based approach by switching both approaches in the srs function used the $+\mathcal{SR}$ method. The improvements achieved by the $+\mathcal{SR}$ method over the both annotation methods are reported. The results reported were obtained by considering a $\gamma = 50$, and the best parameter in λ when is applied the $+\mathcal{SR}$ for both methods. The stacked bars represent the improvements achieved by the $+\mathcal{SR}$ method over the scores obtained by the local and the global methods (denoted by the black bars). In gray bars are shown the improvements obtained by using a knowledge-based approach. After, white bars indicate the improvements obtained by using a statistics-based approach.

In general, we can see that the obtained improvements are consistent, and they are not affected by the approach used for calculating semantic-relatedness similarities. We observed discrepancies when statistics-based and knowledge-based approaches are compared, the reason is that the statistics-based approach offers a wider coverage of labels which allow to measure any relationship. Whilst under the knowledge-based approach those labels without relationship in WordNet has zero similarity.

5.5 Stressing the matching among labels

In order to show if the $+\mathcal{SR}$ method can be useful for supervised evaluations. We have replicated the experiments of the Section 5.1 using an $\alpha = 1.0$ for reporting MAP_α (please note that this time are only considered as good annotations, those with an exact matching). Nevertheless, we still consider to annotate using free-vocabulary. In these experiments, we take into consideration only the ImageCLEF 2013 dataset, thus the training and test data come from the same source, which increases the possibility to annotate correct labels showing an exact match.
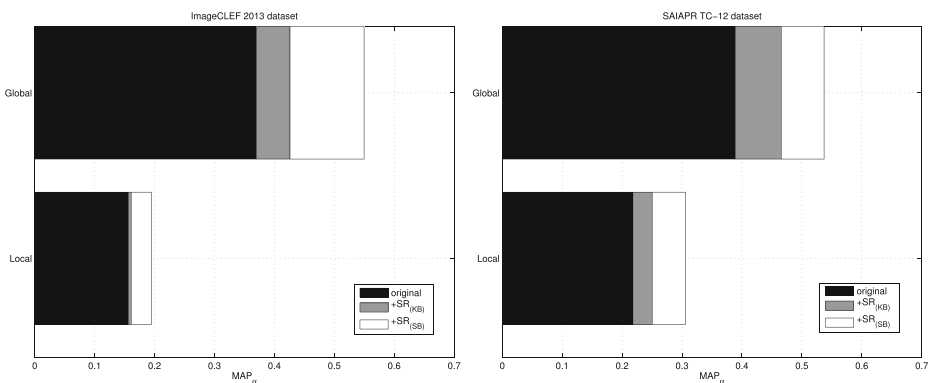

Fig. 13 Comparing performance by using a knowledge-based (KB) approach vs a statistics-based (SB) approach in the $+\mathcal{SR}$ method
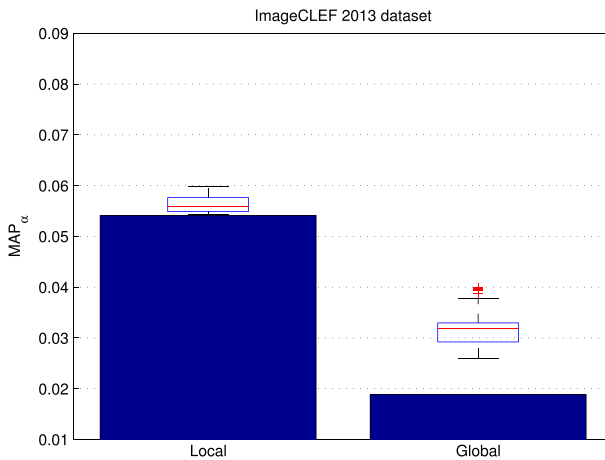
Fig. 14 MAP_α results in the ImageCLEF 2013 dataset for the local and the global methods against the $+\mathcal{SR}$ method, considering an $\alpha = 1$

Figure 14 shows the obtained result. We can see lower performance, in both, the original scores and the new scores obtained after applying the $+\mathcal{SR}$ method. The reason is that the evaluation is more rigid, thus there is almost null matching between the labels assigned and the those in the test set. However, we can see that the $+\mathcal{SR}$ method helps to improve the annotation bringing up the labels with an exact match to the top positions.

6 Conclusions

UAIA methods are attractive due to their capability of annotating images by using unrestricted vocabularies, and also to the easiness of being trained without the need of manually labeled images. By using large vocabularies to perform image annotation the diversity increases, however, at the same time it also increases the number of inconsistent annotations that are considered as wrong.

In this work we introduced the $+\mathcal{SR}$ method that aims to move correct labels toward top positions by means of assigning an adequate relevance. The $+\mathcal{SR}$ method uses the semantic-relatedness information among assigned labels, in order to identify consistent and inconsistent labels. The concept of semantics has different connotations in different domains. In this work, we associate semantics to the meaning of words in terms of the context in which they are used. This is actually the interpretation of semantics in the context of word embeddings (distributed representations). A promising direction of research is considering visual or multimodal semantics for performing the reranking. That is, the semantic relatedness of terms can be estimated by accounting for word-usage in images. An interesting path to explore is to research if it is possible to perform a reranking of the assigned labels by using their representations but in a visual context (e.g. see [18, 52]).

Through exhaustive experimentation, it has been shown the effectiveness of the $+\mathcal{SR}$ method for improving image annotation, and its flexibility to extend main annotation methods improving their performance. From the experiments using different approaches to compute the semantic relatedness of words, we have found that the statistics-based approach allows to achieve higher improvements. However, this does not imply improvements in all

cases, i.e. when the initial assignment has many wrong annotations. For these cases, we recommend to use small values for γ. In addition, we have analyzed the possibility of using the $+\mathcal{SR}$ method under supervised approaches, where it has shown encouraging results.

Publisher's note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Ahmed Z, Zeeshan S, Dandekar T (2016) Mining biomedical images towards valuable information retrieval in biomedical and life sciences. Database 2016. https://doi.org/10.1093/database/baw118
2. Budíková P, Botorek J, Batko M, Zezula P (2014) DISA at imageclef 2014 revised: search-based image annotation with decaf features. arXiv:1409.4627
3. Chavez-Garcia RO, Montes M, Sucar L (2010) Image re-ranking based on relevance feedback combining internal and external similarities. In: Proceedings of the twenty-third international florida artificial intelligence research society conference (FLAIRS 2010), pp 140–141
4. Deng J, Dong W, Socher R, Li LJ, Li K, Li FF (2009) Imagenet: a large-scale hierarchical image database. In: CVPR09. IEEE Computer Society, pp 248–255
5. Escalante HJ, Montes y Gomez M, Sucar LE (2007) Word co-occurrence and markov random fields for improving automatic image annotation. In: Proceedings of the British machine vision conference, vol 2. BMVA Press, Warwick, pp 600–609. https://doi.org/10.5244/C.21.60
6. Escalante HJ, Hernández C, Gonzalez J, López-López A, Montes M, Morales E, Sucar LE, Villaseñor L, Grubinger M (2010) The segmented and annotated {IAPR} tc-12 benchmark. Comput Vis Image Underst 114(4):419–428. https://doi.org/10.1016/j.cviu.2009.03.008. Special issue on image and video retrieval evaluation
7. Escalante HJ, Montes M, Sucar E (2011) An energy-based model for region labeling. Comput Vis Image Underst 115(6):787–803
8. Feng Y, Lapata M (2008) Automatic image annotation using auxiliary text information. In: Proceedings of ACL-08: HLT. Association for Computational Linguistics, Columbus, pp 272–280
9. Grauman K, Leibe B (2011) Visual object recognition. Synthesis Lectures on Artificial Intelligence and Machine Learning 5:1–181. https://doi.org/10.2200/S00332ED1V01Y201103AIM011
10. Guillaumin M, Mensink T, Verbeek J, Schmid C (2009) Tagprop: discriminative metric learning in nearest neighbor models for image auto-annotation. In: International conference on computer vision (ICCV), pp 309–316. https://doi.org/10.1109/ICCV.2009.5459266
11. Hanbury A (2008) A survey of methods for image annotation. J Vis Lang Comput 19(5):617–627. https://doi.org/10.1016/j.jvlc.2008.01.002
12. He Y, Li Y, Lei J, Leung C (2016) A framework of query expansion for image retrieval based on knowledge base and concept similarity. Neurocomputing 204(C):26–32. https://doi.org/10.1016/j.neucom.2015.11.102
13. Hernández-Gracidas CA, Sucar LE, Montes-Y-Gómez M (2013) Improving image retrieval by using spatial relations. Multimedia Tools Appl 62(2):479–505. https://doi.org/10.1007/s11042-011-0911-1
14. Hoque E, Hoeber O, Gong M (2013) Cider: concept-based image diversification, exploration, and retrieval. Inf Process Manag 49(5):1122–1138. https://doi.org/10.1016/j.ipm.2012.12.001
15. Hyung Z, Park JS, Lee K (2017) Utilizing context-relevant keywords extracted from a large collection of user-generated documents for music discovery. Inf Process Manag 53(5):1185–1200. https://doi.org/10.1016/j.ipm.2017.04.006
16. Kherfi ML, Ziou D, Bernardi A (2004) Image retrieval from the world wide web: issues, techniques, and systems. ACM Comput Surv 36(1):35–67. https://doi.org/10.1145/1013208.1013210
17. Li H, Guan Y, Liu L, Wang F, Wang L (2016) Re-ranking for microblog retrieval via multiple graph model. Multimed Tools Appl 75(15):8939–8954

18. Li J, Xu C, Yang W, Sun C, Tao D (2017) Discriminative multi-view interactive image re-ranking. IEEE Trans Image Process 26(7):3113–3127. https://doi.org/10.1109/TIP.2017.2651379

19. Llorente A, Rüger S (2009) Using second order statistics to enhance automated image annotation. In: Proceedings of the 31th European conference on IR research on advances in information retrieval, ECIR '09, pp 570–577

20. Llorente A, Motta E, Rüger S (2010) Exploring the semantics behind a collection to improve automated image annotation. In: Proceedings of the 10th international conference on cross-language evaluation forum: multimedia experiments, CLEF'09, pp 307–314. https://doi.org/10.1007/978-3-642-15751-6-40

21. Makadia A, Pavlovic V, Kumar S (2010) Baselines for image annotation. Int J Comput Vis 90(1):88–105

22. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th international conference on neural information processing systems, NIPS'13. Curran Associates Inc., USA, pp 3111–3119

23. Miller GA (1995) Wordnet: a lexical database for english. Commun ACM 38(11):39–41. https://wordnet.princeton.edu/

24. Mithun NC, Panda R, Roy-Chowdhury AK (2016) Generating diverse image datasets with limited labeling. In: Proceedings of the 2016 ACM on multimedia conference, MM '16. ACM, New York, pp 566–570, https://doi.org/10.1145/2964284.2967285

25. Murray N, Perronnin F (2014) Generalized max pooling. In: Proceedings of the 2014 IEEE conference on computer vision and pattern recognition, CVPR '14. IEEE Computer Society, Washington, pp 2473–2480. https://doi.org/10.1109/CVPR.2014.317

26. Ordonez V, Han X, Kuznetsova P, Kulkarni G, Mitchell M, Yamaguchi K, Stratos K, Goyal A, Dodge J, Mensch A, Daumé H., Berg AC, Choi Y, Berg TL (2016) Large scale retrieval and generation of image descriptions. Int J Comput Vis 119(1):46–59

27. Pandey S, Khanna P, Yokota H (2016) A semantics and image retrieval system for hierarchical image databases. Inf Process Manag 52(4):571–591. https://doi.org/10.1016/j.ipm.2015.12.005

28. Pellegrin L, Escalante HJ, Montes-y Gómez M, González FA (2016) Local and global approaches for unsupervised image annotation. Multimed Tools Appl 1–26. https://doi.org/10.1007/s11042-016-3918-9

29. Pellegrin L, Escalante HJ, Montes-y Gómez M, Villegas M, González FA (2018) A flexible framework for the evaluation of unsupervised image annotation. In: Mendoza M, Velastín S (eds) Progress in pattern recognition, image analysis, computer vision, and applications. Springer International Publishing, Cham, pp 508–516

30. Rahman M, Antani S, Thoma G (2011) A query expansion framework in image retrieval domain based on local and global analysis. Inf Process Manag 47(5):676–691. https://doi.org/10.1016/j.ipm.2010.12.001

31. Reshma IA, Ullah MZ, Aono M (2014) Ontology based classification for multi-label image annotation. In: 2014 international conference of advanced informatics: concept, theory and application (ICAICTA), pp 226–231. https://doi.org/10.1109/ICAICTA.2014.7005945

32. Ramírez-de-la Rosa G, Montes-y Gómez M, Solorio T, Villaseñor-Pineda L (2013) A document is known by the company it keeps: neighborhood consensus for short text categorization. Lang Resour Eval 47(1):127–149

33. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) Imagenet large scale visual recognition challenge. Int J Comput Vision 115(3):211–252. https://doi.org/10.1007/s11263-015-0816-y

34. Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Inf Process Manag 24(5):513–523. https://doi.org/10.1016/0306-4573(88)90021-0

35. Sánchez-Oro J, Montalvo S, Montemayor A, Pantrigo J, Duarte A, Fresno V, Martínez R (2013) Urjc&uned at imageclef 2013 photo annotation task. In: CLEF 2013 evaluation labs and workshop, online working notes, vol 1179. CEUR-WS.org

36. Schroff F, Criminisi A, Zisserman A (2011) Harvesting image databases from the web. IEEE Trans Pattern Anal Mach Intell 33(4):754–766. https://doi.org/10.1109/TPAMI.2010.133

37. Schwartz H, Gomez F (2011) Evaluating semantic metrics on tasks of concept similarity. In: FLAIRS conference

38. Siva P, Russell C, Xiang T, Agapito L (2013) Looking beyond the image: unsupervised learning for object saliency and detection. In: 2013 IEEE conference on computer vision and pattern recognition, pp 3238–3245. https://doi.org/10.1109/CVPR.2013.416

39. Stathopoulos S, Kalamboukis T (2014) Ipl at imageclef 2014: Scalable concept image annotation. In: CLEF 2014 evaluation labs and workshop, online working notes, vol 1180, pp 398–403. CEUR-WS.org

40. Tariq A, Foroosh H (2017) Learning semantics for image annotation. arXiv:1705.05102

41. Tian F, Shen X (2013) Annotating web images by combining label set relevance with correlation. In: Proceedings of the 14th international conference on web-age information management, WAIM'13. Springer, Berlin, pp 747–756. https://doi.org/10.1007/978-3-642-38562-9-76

42. Tian F, Shen X (2014) Learning label set relevance for search based image annotation. In: 2014 international conference on virtual reality and visualization, pp 260–265. https://doi.org/10.1109/ICVRV.2014.39
43. Torralba A, Fergus R, Freeman WT (2008) 80 million tiny images: a large data set for non-parametric object and scene recognition. IEEE Trans Pattern Anal Mach Intell 30(11):1958–1970. https://doi.org/10.1109/TPAMI.2008.128
44. Uricchio T, Bertini M, Ballan L, del Bimbo A (2013) MICC-UNIFI at ImageCLEF 2013 scalable concept image annotation. In: Working notes for CLEF 2013 conference, Valencia, Spain, September 23–26, vol 1179. CEUR-WS.org
45. Villatoro E, Juárez A, Montes M, Villaseñor L, Sucar LE (2012) Document ranking refinement using a markov random field model. Nat Lang Eng 18(2):155–185
46. Villegas M, Paredes R (2012) Overview of the imageclef 2012 scalable concept image annotation task. In: CLEF 2012 evaluation labs and workshop, online working notes, vol 1178. CEUR-WS.org
47. Villegas M, Paredes R (2013) 2013 imageclef webupv collection. https://doi.org/10.5281/zenodo.257722
48. Villegas M, Paredes R, Thomee B (2013) Overview of the imageclef 2013 scalable concept image annotation subtask. In: CLEF 2013 Evaluation labs and workshop, pp 1–19. CEUR-WS.org
49. Villegas M, Müller H, Gilbert A, Piras L, Wang J, Mikolajczyk K, de Herrera AGS, Bromuri S, Amin MA, Mohammed MK, Acar B, Uskudarli S, Marvasti NB, Aldana JF, del Mar Roldán García M (2015) General overview of ImageCLEF at the CLEF 2015 labs. In: Experimental IR meets multilinguality, multimodality, and interaction: 6th international conference of the clef association. Lecture notes in computer science, vol 9283. Springer International Publishing, Cham, pp 444–461. https://doi.org/10.1007/978-3-319-24027-5-45
50. Wu Z, Palmer M (1994) Verbs semantics and lexical selection. In: Proceedings of the 32nd annual meeting on association for computational linguistics, ACL '94. Association for Computational Linguistics, Stroudsburg, pp 133–138. https://doi.org/10.3115/981732.981751
51. Xu H, Pan P, Lu Y, Xu C, Chen D (2014) Improving automatic image annotation with google semantic link. In: 2014 10th international conference on semantics, knowledge and grids, pp 177–184. https://doi.org/10.1109/SKG.2014.12
52. Zhang L, Zhang Q, Zhang L, Tao D, Huang X, Du B (2015) Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding. Pattern Recogn 48(10):3102–3112. https://doi.org/10.1016/j.patcog.2014.12.016. Discriminative feature learning from big data for visual recognition

Luis Pellegrin was graduated from the Instituto Tecnologico Superior de Cajeme in 2006. In 2008, he earned a M.Sc. degree in Artificial Intelligence in the Universidad Veracruzana. And in 2017, he received the Ph.D. in Computer Sciences from the Instituto Nacional de Astrofisica, Optica y Electronica (INAOE). His research interests include image annotation, vision and language processing, neural networks, bio-inspired algorithms, and machine learning focus in representation, indexing and automatic analysis of multimodal data.

Hugo Jair Escalante is a part of the Laboratory of Language Technologies (LabTL) at the National Institute of Astrophysics, Optics and Electronics (INAOE) México. He received his Ph.D. and M.Sc. in Computer Science from National Institute of Astrophysics, Optics and Electronics (INAOE) in 2010 and 2006 respectively. Among his research topics of interest are machine learning, text mining, information retrieval and computer vision.

Manuel Montes-y-Gómez obtained his Ph.D. in Computer Science from the Computing Research Center of the National Polytechnic Institute of Mexico. Currently, he is a full-time lecturer at the Computational Sciences Department of the National Institute of Astrophysics, Optics and Electronics, located in Puebla, Mexico. His research is on automatic text processing. He is author of more than 100 international papers in the fields of information retrieval, question answering, information extraction and text mining. Dr. Montes has been visiting professor at the Polytechnic University of Valencia (Spain), at the University of Geneva (Italy), and from August 2010 he is visiting professor at the University of Alabama (USA). In addition, Dr. Montes is member of the Mexican National System of Researchers, the Mexican Society of Artificial Intelligence, the Mexican Association for Natural Language Processing and the International Web Intelligence Consortium.

Fabio A. González is an Associate Professor at the Department of Computer Systems and Industrial Engineering at the National University of Colombia, where he leads the Machine Learning, Perception and Discovery Lab (MindLab). He earned a Computer Systems Engineer degree and a MSc in Mathematics degree from the National University of Colombia in 1993 and 1998 respectively, and a MSc and PhD degrees in Computer Science from the University of Memphis, USA, in 2003. His research work revolves around machine learning and information retrieval, with a particular focus on the representation, indexing and automatic analysis of multimodal data (data encompassing different types of information: textual, visual, signals, etc.). Some of the applications of his research work include content-based image retrieval of biomedical and natural images, automatic image annotation and visual content mining, among others.

Affiliations

Luis Pellegrin[1] ⬛ **· Hugo Jair Escalante[2] · Manuel Montes-y-Gómez[2] ·**
Fabio A. González[3]

Hugo Jair Escalante
hugojair@inaoep.mx

Manuel Montes-y-Gómez
mmontesg@inaoep.mx

Fabio A. González
fagonzalezo@unal.edu.co

[1]   Faculty of Sciences, Universidad Autónoma de Baja California (UABC), Ensenada, Mexico

[2]   Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE),
      San Andrés Cholula, Mexico

[3]   Computing Systems and Industrial Engineering Department, Universidad Nacional
      de Colombia, Bogotá, Colombia