REGIMvid at ImageCLEF2011: Integrating Contextual Information to Enhance Photo Annotation and Concept-based Retrieval

Amel Ksibi, Anis Ben Ammar and Chokri Ben Amar
REGIM: REsearch Group on Intelligent Machines,
University of Sfax, ENIS, BP W, 3038, Sfax, Tunisia
{amel.ksibi, anis.benammar, chokri.benamar}@ieee.org

Abstract. Recently, social tagging has become a new trend to describe and to search online resources on the Social Web. Such movement has generated an abundant volume of users' tags. One key challenge is the absence of contextual information associated with these tags. In our work for ImageCLEF 2011, we explore the correlation inter-tags by constructing two types of contextual graph: inter-concepts graph and concept-tags graph, to further improve automatic photo annotation and retrieval. The results show that our system runs reasonably but with some inherent limitations due to using only the contextual correlation and ignoring the semantic relationships for these tasks.

Keywords: context, inter-concepts graph, concept-tags graph, concept detection, automatic annotation, concept-based retrieval, Flickr-related-tags, query-to-concepts mapping

1 Introduction

Our group REGIMvid within REGIM unit research participated in the photo annotation and concept based retrieval tasks [1]. The first task aims to automatically annotate a large number of consumer photos with multiple annotations from a predefined set of keywords (the concepts). The second task deals with the concept-based retrieval by using logical connections of concepts from the annotation task.

The main challenge for these tasks is the concept detection process. Psychophysics studies have shown that the detection of a concept of interest (e.g. buildings) could provide a statistical evidence to find other concepts (e.g. street, city) which may not themselves be of interest. Such concept correlation, namely contextual information, will probably enhance the performance of the concept detection process.

Considering the concept as a specific tag, our system explores Flickr tags to extract such contextual relationships concerning tag relations: if two tags frequently co-occurred, they must have a relation between them. A high frequency of co-occurrence emphasizes the possibility of some connection between the tags.

In this paper, we attempt to apply these contextual relationships in two ways: Firstly, we attempt to extract contextual relationships between concepts to build the inter-concepts graph. Such graph will be used not only to improve the annotation by detecting implicit concepts which have relations with concepts detected explicitly, but also to expand the query in concept based retrieval by adding other implicit concepts which are in the same context with the initial selected concepts. Secondly, we propose to model each concept by its related tags as a concept-tags graph to further matching these tags with the tags of each image test for concept detection.

The rest of the paper is organized as follows. Section 2 describes our proposed contextual graphs; Section 3 details the processes of photo annotation and concept based retrieval. Section 4 details the submitted runs.

2 Modeling Contextual graphs

This section describes the two proposed contextual graphs used for annotation and retrieval tasks. The first is the inter-concepts graph and aims to model the contextual relationship inter-concepts. The second is the concept-tags graph and aims to order the Flickr's Related Tags of each concept according to their contextual information.

To define these graphs, we first introduce the following notations which will be used in the rest of the paper: Given N concepts (C_j) j=1..N , for each concept, we note its Flickr's Related Tags by $(tc_j^i)_{i=1..M}$ where M is the number of Flickr's Related tags to concept C_j . For each test image, we note its tags by $(tim^h)_{h=1..T}$ where T defines the number of tags for the image.

2.1 Modeling inter-concepts graph

Generally, concepts are correlated to each other. Such concept correlation contains useful contextual information which can contribute a lot to photo annotation and retrieval. In fact, the detection of a concept could provide statistical evidence to confirm the existence of another concept from the inter-concepts contextual relationships. For example, the concept “Bird” frequently co-occurs with concept “Sky”. Using the contextual information from Sky is expected to help detect Bird. In the literature, inter-concepts graph can be built by modeling inter-concept relationships through annotation provided by either manual labeling or machine tagging [2].

The above type of relation is a locally statistical correlation dependent on a limited set (manual annotated training set, automatic annotated test set). To get a more robust analysis, we seek other entrances to enrich the representation of concepts relatedness. Social media sharing web sites as Flickr are considered as the largest public available multimedia corpus.

In this section, we define the contextual information inter-concepts based on their co-occurrence by exploring Flickr resources. Analogous to the principle of Google distance [3], we first estimate the distance between two concepts C_i and C_j as follows:

$$NGD(Ci,Cj) = \frac{\max\{\log h(C_i), \log h(C_j)\} - \log h(C_i, C_j)}{\log X - \min\{\log h(C_i), \log h(C_j)\}} \quad (1)$$

where $h(C_i)$ denotes the number of images contained tag C_i , $h(C_j)$ denotes the number of images contained tag C_j and $h(C_i, C_j)$ denotes the number of images contained both tags C_i and C_j; X is the total number of images on Flickr, which is roughly estimated as 5 billion. The NGD is then converted to Flickr context similarity FCS [4] and defined as:

$$FCS(Ci,Cj) = e^{-NGD(Ci,Cj)} \quad (2)$$

Using FCS similarity, concepts are organized to form the context graph G in off-line process. Edge weights are set equal to corresponding similarities for modeling the contextual closeness among concepts. Since the FCS measure is symmetric, G is an undirected graph. Fig.1 shows a partial view of the inter-concepts graph G constructed from ImageCLEF2011 concepts. The thick lines illustrate a strong contextual relationship such as that between “Sky” and “Clouds”. In contrast, the thin lines represent a negligible contextual relationship such as that between “Indoor” and “Sea”.

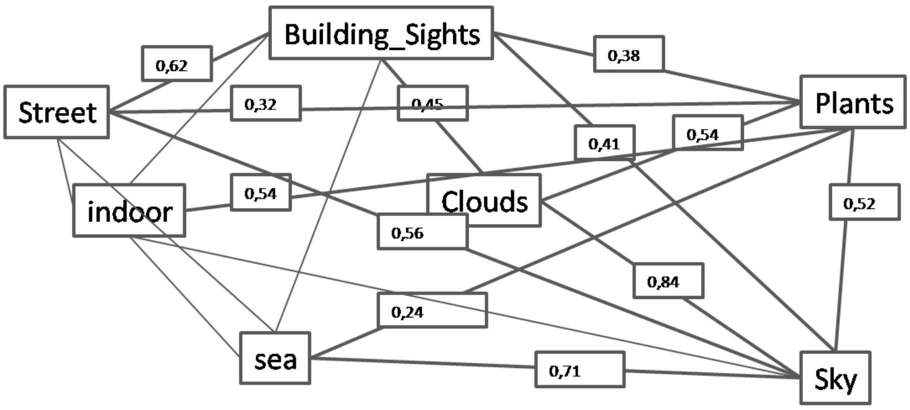

Fig.1. Partial view of the inter-concepts graph G constructed from concepts proposed in ImageCLEF2011

2.2 Modeling concept-tags graph

We propose to model each concept by a graph of Flickr's related tags. Flickr's Related Tags are “a list of tags ’related’ to the given tag (in our case, the concept), based on clustered usage analysis” [5]. For example, the top related tags for the concept “Party life” are “birthday, friends, fun, music, dance, girl, wedding, people, cake, family, night, club”. It can be seen that these words have high co-occurrence with this concept. However, these related tags generally are in a random order without any importance or relevance information, which limits the effectiveness of these tags. So, we propose to apply the same measure applied in inter-concepts graph, FCS measure [5], for ranking these related tags to the concept by weighting concept-tag correlation. As a consequence, each tag has its Flickr Context Similarity score FCS which defines the relatedness of a tag to the concept.

Given a “concept i”, we consider two graphs: the first is constructed without tags ranking and the second shows the scores related to each tag.

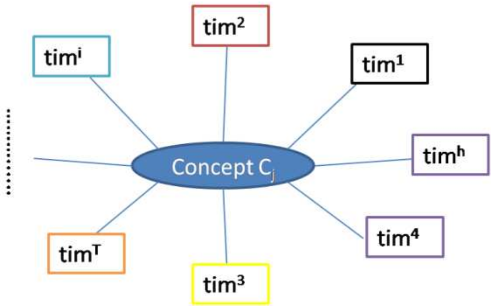

Fig.2a. Concept-tags graph without tags ranking

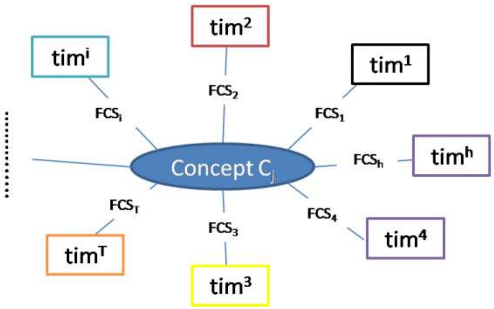

Fig.2b. Concept-tags graph with tags ranking

3 Experiments within photo annotation and concept-based retrieval tasks

In this section, we describe the proposed processes of the two tasks.

3.1 Photo annotation

The visual concept detection and annotation task is a multi-label classification challenge. It aims at the automatic annotation of a large number of consumer photos with multiple annotations from a predefined set of keywords (99 visual concepts).

A variety of techniques have been proposed for photo annotation in recent years [6], [7], [8], [14]. Most of these methods try to model the probabilistic relationship between tags and images. Although, these methods have achieved varying levels of success, the absence of context information in these methods limits the accuracy of automatic photo annotation. Our strategy is to add the notion of context to extract accurate and reliable relations inter-concepts and concept-tags to improve annotation results.

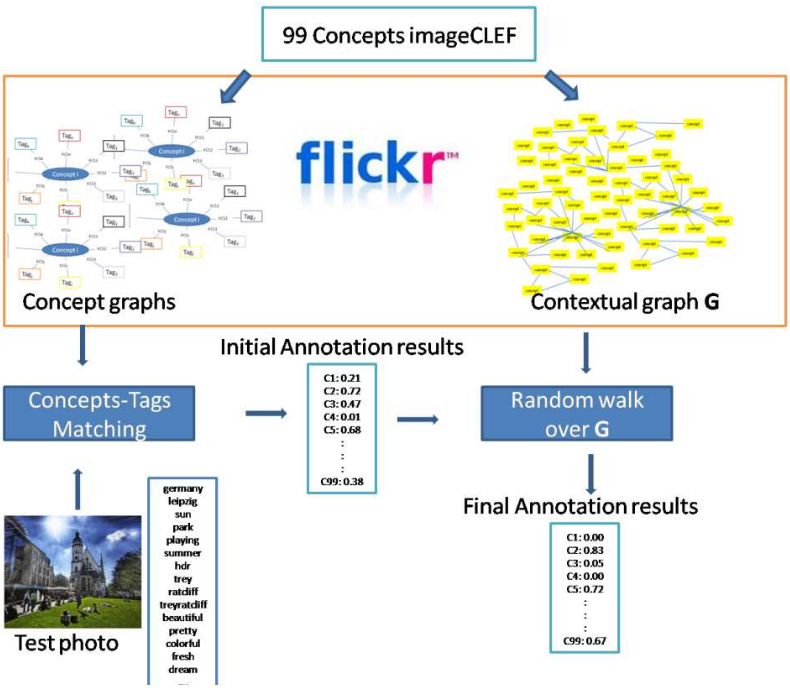

Fig.3. Proposed photo annotation process

The proposed process for automatic annotation is illustrated in Fig.3. After the construction of the contextual graph and modeling each concept by its Concept-tags graph, we estimate for each image the relevance scores of concepts $(Score_j)_{j=1..N}$ based on context similarity estimation as the following:

Given an image, for each conceptC_j, we define the $Score_j$ by:

$$Score_j = \frac{\sum_{i=1}^{M} \sum_{h=1}^{T} f_h(i,j)}{card\left(\left(\sum_{h=1}^{T} f_h(i,j)!=0\right)_{i=1..M}\right)} \quad \textbf{(3)}$$

$$\text{Where} \quad f_h(i,j) = \begin{cases} FCS(C_j, \mathrm{tc}_j^i), & \mathrm{tc}_j^i = tim^h \\ 0, & else \end{cases} \quad \textbf{(4)}$$

To investigate the relationship between concepts, we perform, then, a random walk over the inter-concepts graph by propagating these relevance scores among concepts. Random walk methods have been widely applied in machine learning and information retrieval fields [11], [12], [13].

Given a contextual graph G with N nodes (concepts), the random walk process is modeled as $W = \left(G, P, x_{(k)}\right)$ Where $P = \left[p_{ij}\right]_{N*N}$ is the transition matrix, and $x_{(k)}$ is a column vector encapsulating the stationary probabilities of the concepts at iteration k. p_{ij} indicates the probability of the transition from node i to node j and it is computed as

$$p_{ij} = \frac{FCS(C_i, C_j)}{max_{l=0..N}(FCS(C_i, C_l) - min_{l=0..N}(FCS(C_i, C_l)} \quad \textbf{(5)}$$

The stationary probabilities in $x_{(0)}$ are initialized based on the set of concepts selected by a given visual query. For concepts not selected, their weights are initialized to zero. The random walk process is thus formulated as:

$$x_{(k)}(j) = \propto \sum_{i \neq j} x_{(k-1)}(i) p_{ij} + (1-\propto) x_{(0)}(j) \quad \textbf{(6)}$$

Where $\propto \in [0,1]$ is a parameter to control the speed of convergence, and $x_{(0)}$ is the initial stationary probability. The stationary probability of a concept C_j is iteratively updated at each time instance k, until meeting the convergence condition of $\left|x_{(k+1)} - x_{(k)}\right| \to 0$.
Using the random walk process, some scores of concepts which have strong correlations with the selected concepts will be increased. This process will promote the concepts that have many close neighbors and weaken isolated concepts.

3.2 Concept based retrieval

The second challenge is a concept-based retrieval task and is performed on MIR Flickr dataset. 40 proposed topics consist of a logical connection of concepts from the annotation task. To give an example, we could ask to "Find all images that depict a small group of persons in landscape scenery showing trees and a river on a sunny day."

Intuitively, if queries can be automatically mapped to related semantic concepts, search performance will benefit significantly. For example, a query as “scenes with snow” will surely benefit from the concept “Snow”, or even “Sky” since a snowy scene is often with sky present. However, it remains unclear. An important research problem rises here as how to map the query to the concepts automatically and take into consideration the logical connection inter-concepts?

A few works positively support the usefulness of query-to-concept-mapping in the query-by-concept paradigm [9], [10] by projecting the query in semantic spaces constructed from either local manual annotation or fixed ontology as WordNet[1] or LSCOM[2]. As a consequence, these methods are disabled to update their knowledge. In order to tackle this limitation, we attempt to explore the social web to extract contextual relations between concepts to perform the query to concept mapping in the proposed contextual space.

Fig. 4 shows the proposed process for concept based retrieval by context reasoning. Given a visual query (photo(s)) which is binary labeled with the 99 concepts, a query vector is constructed by normalizing the number

1 http://wordnet.princeton.edu/
2 http://lscom.org/

of annotation. After that, this vector is projected to the inter-concepts graph to expand the query by taking into consideration the relationships inter-concepts. Then, the query matching is accomplished by calculating the Euclidian distance between the expanded vector and annotation vectors of the test photos. Finally, the results ranking is done in decreasing way according to the scores of query matching.

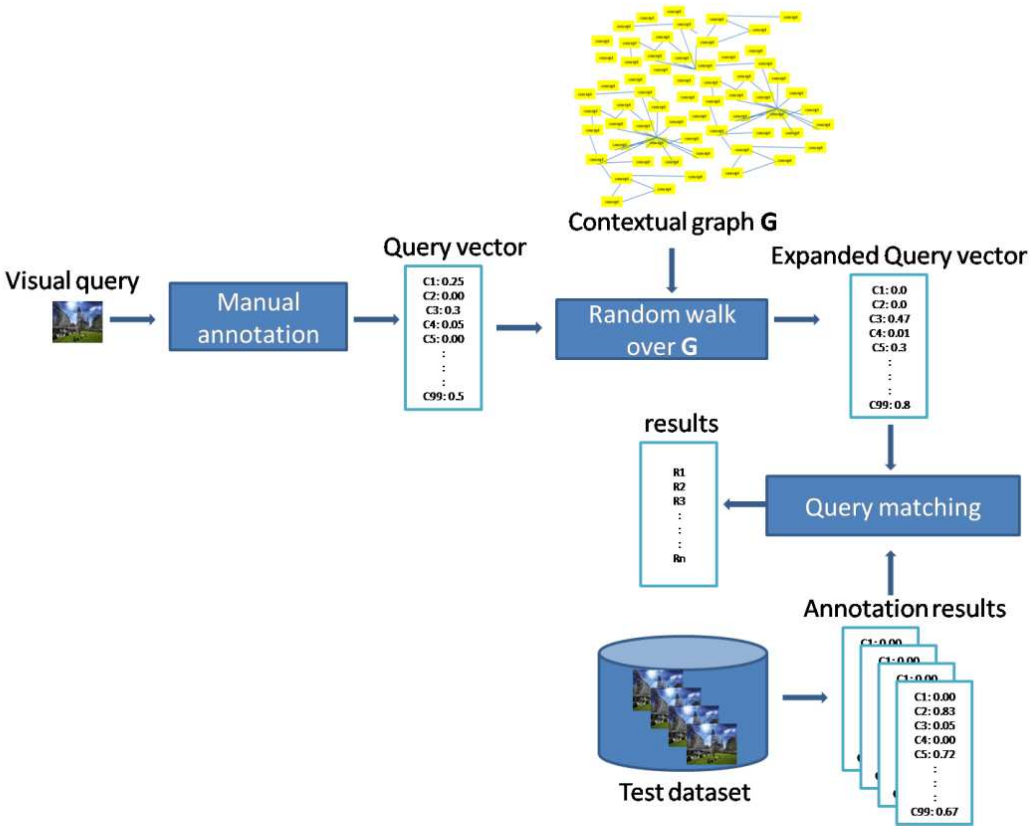

Fig.4. Concept based retrieval by context reasoning

4 Submitted run and results

We submitted two runs based on the processes described above. The first run has the following identifier: REGIMVID_1308507656911__REGIMVID.txt_confidence.txt and the second has the following identifier REGIMVID_1308548603241__REGIMVIDT1FRT.txt The official measure used in photo annotation and concept based retrieval tasks in ImageCLEF2011 is the Mean Average Precision (MAP).

For the first task, we obtained the rank 74 on 79 submissions, with a MAP value of 0.204274. This value is 0.1 lower than the median value for these runs, 0.3159. For the example-based F-measure, we obtained the rank 78 on 79. The value obtained is 0.140913, and the median value is 0.49; we achieve then poor results according to this measure. For the concept based retrieval task in ImageCLEF2011, we obtained the rank 29 on 31 submissions, with value of 0.0042. This value is lower than the median value for theses runs, 0.066.

5 Conclusion

In our first participation in ImageCLEF, we proposed to integrate the notion of context in the two submitted runs about photo annotation and concept based retrieval. So, we constructed two contextual graphs that define the co-occurrence relationships inter-tags obtained from the social Web resources of Flickr. In the concept-tags graph, we explored the Flickr related tags to model concepts by measuring the co-occurrence similarity between related tags and concepts. In the inter-concepts graph, we propagated the relevance scores of concepts deduced from the annotation task, to refine the annotation results and to expand the query in concept based retrieval.

Unfortunately, the proposed processes did not have as good result as we expected. The problem is most probably due to ignoring the semantic correlation inter-concepts and between tags and concepts. Also, it could be due to using only the textual features. So, as a future perspective, it will be necessary to introduce the visual features to model concepts. Moreover, we intend to merge the semantic relationships and the contextual relationships in our models.

References

1. Nowak, S., Nagel, K., Liebetrau, J.: The CLEF 2011 Photo Annotation and Concept-based Retrieval Tasks, In: CLEF 2011, Amsterdam, The Netherlands (2011).
2. Wei, X.Y., Jiang, Y.G., Ngo, C.W.: Exploring interconcept relationship with context space for semantic video indexing, In: ACM International Conference on Image and Video Retrieval, Santorini, Greece (2009).
3. Cilibrasi, R., Vitanyi, P. M. B.: The Google Similarity Distance. In: IEEE Transactions on Knowledge and Data Engineering, Vol. 19, pp. 370 (2007).
4. Jiang, Y.G., Ngo, C.W., Chang, S.F.: Semantic context transfer across heterogeneous sources for domain adaptive video search. In: ACM International Conference on Multimedia, pp. 155--164, Beijing (2009).
5. Xu, H., Zhou, X., Wang, M., Xiang, Y., Shi, B.: Exploring Flickr's related tags for semantic annotation of web images. In Proceedings of CIVR. Santorini Island (2009).
6. Ksibi, A., Elleuch, N., Benammar, A., Alimi, M.A.: Semi-Automatic Soft Collaborative Annotation for Semantic Video Indexing. In: IEEE EUROCON Conference, pp. 1--6, Lisbon (2011).
7. Liu, X., Ji, R., Yao, H., Xu, P., Sun, X., Liu, T.: Cross-media manifold learning for image retrieval & annotation. In Multimedia Information Retrieval, pp. 141--148, Canada (2008).
8. Sigurbj¨ornsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proceeding of the 17th international conference on World Wide Web, pp. 327--336, (2008).
9. Wei, X.Y., Jiang, Y.G., Ngo, C.W: Concept-Driven Multi-Modality Fusion for Video Search. In: IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 1, pp. 62--73 (2011)
10. Feng, B., Cao, J., Bao, X., Bao, L., Zhang, Y., Lin, S., Yun, X.: Graph-based multi-space semantic correlation propagation for video retrieval. In: The Visual Computer, pp. 21--34, (2011)
11. Hsu, W.H., Kennedy, L.S., Chang, S. F.: VideoSearch Reranking through Random Walk over Document-Level Context Graph. In: Proceeding of 14th ACM International Conference on Multimedia, Santa Barbara(2006).
12. Jing, Y.S., Baluja, S.: VisualRank: Applying PageRank to Large-Scale Image Search. In: Transactions on Pattern Analysis and Machine Intelligence, (2008).
13. Liu, D. Hua, X.S., Yang, L., Wang, M., Zhang, H.J.: Tag ranking. In: Proceedings of the 18th international conference on World wide web, pp. 20--24, Madrid (2009)
14. Wu, L., Hoi, S.C., Zhu, J., Jin, R., Yu, N.: Distance metric learning from uncertain side information with application to automated photo tagging. In: Proceedings of ACM International Conference on Multimedia, pp. 19--24, Beijing (2009).