Semantic Similarity Between Images: A Novel Approach Based on a Complex Network of Free Word Associations

Enrico Palumbo[1(✉)] and Walter Allasia[2(✉)]

[1] Physics University of Torino, via Giuria, 1, 10025 Torino, Italy
enrico.palumbo@edu.unito.it
[2] EURIX, via Carcano, 26, 10153 Torino, Italy
allasia@eurix.it

Abstract. Several measures exist to describe similarities between digital contents, especially for what concerns images. Nevertheless, distances based on low-level visual features embedded in a multidimensional linear space are hardly suitable for capturing semantic similarities and recently novel techniques have been introduced making use of hierarchical knowledge bases. While being successfully exploited in specific contexts, the human perception of similarity cannot be easily encoded in such rigid structures. In this paper we propose to represent a knowledge base of semantic concepts as a *complex network* whose topology arises from free conceptual associations and is markedly different from a hierarchical structure. Images are anchored to relevant semantic concepts through an annotation process and similarity is computed following the related paths in the complex network. We finally show how this definition of semantic similarity is not necessarily restricted to images, but can be extended to compute distances between different types of sensorial information such as pictures and sounds, modeling the human ability to realize synaesthesias.[1]

Keywords: Semantic similarity · Complex networks · Free word associations · Image analysis

1 Introduction

Content-based image retrieval is a well established research branch, working on low-level visual features, such as color or texture, that can be automatically extracted from digital contents [7,19]. Unfortunately, it is often the case that purely visual features do not encode similarities regarding high-level concepts. Smeulders et al. define the *semantic gap* as "the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation" [19]. Image annotation

[1] This work was partially funded by the European Commission in the context of the FP7 ICT project ForgetIT (under grant no: 600826)

G. Amato et al. (Eds.): SISAP 2015, LNCS 9371, pp. 170–175, 2015.
DOI: 10.1007/978-3-319-25087-8_16

attempts to fill the semantic gap by mapping low-level visual features into high-level concepts, either manually or through machine learning algorithms such as Support Vector Machines (as done in [13], possibly combined with more structured hierarchical knowledge bases [7,21]). After the annotation, a picture is represented as a Bag of Words, namely a vector whose elements indicate the presence (or the absence) of the concepts utilized in the annotation process and distances are evaluated in multidimensional L^p Lebesgue spaces or more generalized topological spaces [17].

2 Graph-Based Similarity

The representations of images as vectors in a metric space all rely on the assumption of independence between the words used in the annotation process [8]. As also argued in [12], this is rarely true. Let us suppose to make use of three concepts for the annotation, tree, leaf and window, and to have three pictures, one containing only a tree, one only a leaf and one only a window. The vector representation would be: $tree = (1, 0, 0)$, $leaf = (0, 1, 0)$, $window = (0, 0, 1)$ and the distance between the images would be the same, even if intuitively the concept of tree should be closer to the concept of leaf than to that of window. Indeed, the natural semantic correlations among the concepts used in the annotation make inadequate the euclidean representation. To comply with the necessity of a structure which well expresses relations, it is common to make use of *semantic graphs*. A semantic graph is a pair of sets $G = (C, A)$ where C is the set of nodes and A is the set of edges, i.e. links between nodes. In semantic graphs the nodes are concepts (or words) and the edges represent either logical relations between them, e.g. 'is-a', 'has-a', or simple conceptual associations. Notable examples of semantic graphs are the models of semantic memory developed by Collins and Quillian [5] and Collins and Loftus [4] or networks of words such as WordNet [14], Roget's Thesaurus [18], Word Association networks or network of tags such as those of [3] and the like. Ontologies as well may be seen as semantic graphs whose structure must be logically consistent and is often hierarchical and in which formal rigor is added by means of logical axioms and inference rules [11]. In the last years, many have tried to overcome the limitations of the euclidean representation utilizing semantic graphs [8,9,12]. In [8], for instance, the authors use ImageNet, a logically organized database of images, analogous to WordNet, to evaluate semantic similarities between images. These methods, however, only account for logical similarities, namely for shared taxonomical categories. Oppositely, humans can analyze images at different semantic levels [21] and can establish more complex relations between them, which cannot be easily encoded in a hierarchical structure (Fig. 1). A pair of images can be considered to be related because the objects represented often occur together, because they evoke similar feelings or belong to the same context. Statistical evidence of this fact is presented in [10]. The authors compare the associations of the Word Association Network of the Human Brain Cloud [22], a web-based "massively interplayer word association game", which they have validated for scientific purposes, with the logical relations of WordNet. They map the word association

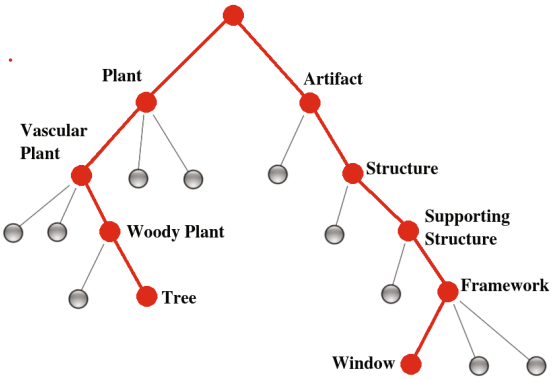

Fig. 1. *Sketch of the structure of ImageNet. The only way to reach Window from Tree is to go up and down the hierarchy: 'Tree-WoodyPlant-VascularPlant-Plant-Root-Artifact-Structure-SupportingStructure-Framework-Window'. On the Word Association Network built from the data of [16] the path is 'Tree-Shade-Window'.*

network, which completely lacks of semantics, onto WordNet and what they observe is that "human beings often construct associations with probabilities that could strongly deviate from what would be the pure statistical structure of WN", entailing that conceptual associations are often based on other criteria than pure logic.

3 Complex Networks

A further limitation of the hierarchical models is that they rarely exhibit *complex* structures. In fact, in the last years, starting from the article of Steyvers and Tenenbaum [20], many have pointed out the strong analogies between semantic graphs and *complex networks*. The study of *complex networks* is a new and emerging field, born in the late 90s as a consequence of the discovery that many real networks (WWW, Internet, science collaboration graph, the web of human social contacts,...) are *small world*, i.e. the distance between two nodes scales logarithmically with the number of nodes N, highly *clustered* and *heterogeneous*, i.e. the degree distribution is considerably different from binomial, poissonian or gaussian distributions, since it is markedly right-skewed and fat-tailed, often well approximated by a power-law distribution [1,2].

In [20] the authors have shown that Wordnet, the Roget's thesaurus and the Word Association network built from the experiment done by the University of South Florida [16] exhibit a small average path length, a high clustering coefficient and a power-law distribution of degree. Word associations are obtained through a simple experiment: subjects are asked to write down the first word that comes to their mind which is meaningfully related to a cue word, provided by the experimenters. A network can be built by identifying the words as nodes and the edges as associations, which can be weighted by the frequency of that

particular association. In [10] this analysis is extended to another network of word association, the Human Brain Cloud [22], an online multiplayer word association game. In [15] similar results are obtained, without aggregating data from different individuals. In [3] the topological properties of the semantic networks spontaneously emerging from co-occurring tags of digital resources on website such as del.icio.us also exhibit the typical properties of complex networks. These independent studies, obtained from semantic graphs of diverse nature and origin, yielding similar conclusions suggest that *complexity* is a fundamental property of the structure of semantic networks. This fact has remarkable consequences on the shortest paths, therefore on similarities. Scale-free networks are more than *small world*, with average shortest path $< l > \simeq \frac{\log N}{\log \log N}$ [2] with N vertices. This is due to the hubs of the networks which act as bridges between "distant" nodes, providing shortcuts across the web.

4 The Model

To account for the role of complexity and to encompass the possibility of free conceptual association, we propose a model for evaluating semantic similarities between images based on a Complex Network of Free Word Associations. Both the Word Association Network built from the experiment of the University of South Florida [16] and the one built from the web-based experiment of HumanBrainCloud, have been proven to be *complex networks* and to share a similar structure [10,20]. Therefore, in the following, we shall generically refer to a Word Association Network (WAN). The model works as follows (Fig. 2):

1) Build a Word Association Network
2) Annotate the images I_i and I_j with words of the WAN: f_i and f_j are the vectorial representations of I_i and I_j, whose component $f_i^k \in [0,1]$ is a confidence score of the word k in I_i
3)Turn the most relevant words into weighted links and anchor the images to the WAN
4)The distance is the shortest weighted path length [6], namely $d(I_i, I_j) = min_{\gamma_{i,j}} \sum_{l \in \gamma_{i,j}} l$, where $l = \frac{1}{w}$ is the length of a link (the stronger the association, the closer the nodes) and $\gamma_{i,j}$ is a generic weighted path connecting I_i and I_j

Note that "most relevant" is vague and needs to be further specified. Different criteria may be applied to determine what number of words should be turned into links, but a robust method is to normalize f_i, sort the components by magnitude and select the first k concepts containing a fixed percentage α of the total norm. In the demonstration of Fig. 2, we have used $\alpha = 0.9$, but different values may be selected. We suggest that this free parameter could be set optimally through a learning process onto a training set of images whose similarity have been already evaluated by well established methods.

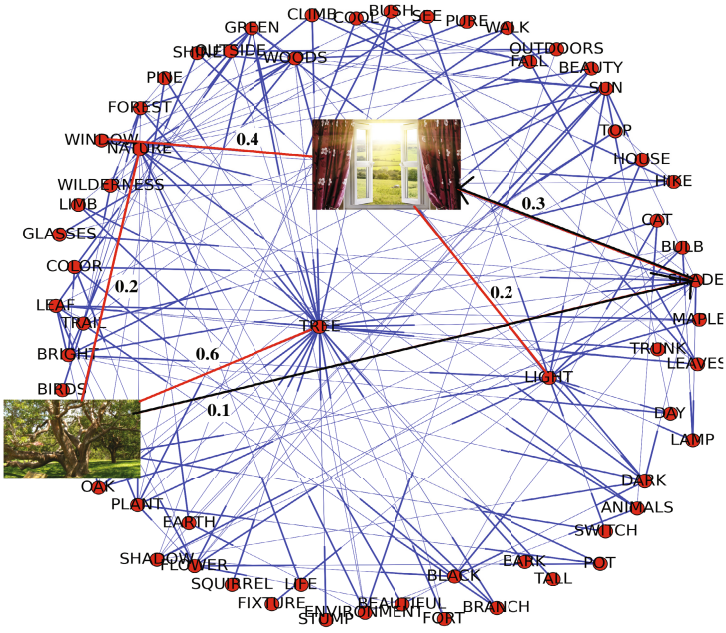

Fig. 2. *Suppose that the annotation yields* $f_1 = (window = 0.4, shade = 0.3, light = 0.2, nature = 0.05, sheep = 0.05)$ *and* $f_2 = (tree = 0.6, nature = 0.2, shade = 0.1, branch = 0.1)$. *Setting* $\alpha = 0.9$ *we select the first three attributes and turn them into weighted links (represented in red and black). The shortest path (in black) between the two images is Img1* $\rightarrow$ *Shade* $\rightarrow$ *Img2, hence the similarity is* $d = 1/0.1 + 1/0.3 \approx 13.3$ *in the WAN built from [16].*

5 Conclusions

In this position paper, we have highlighted two possible weak points of the state-of-the-art measures of semantic similarity between images: the excessive rigidity of purely *hierarchical structures* and the *absence of complexity*. Therefore, we have proposed a model which could possibly solve these issues. We want to underline the fact that this definition of similarity can be extended to digital contents of diverse nature, such as images, sounds and more generally media objects. Once the objects are semantically annotated, the proposed algorithm allows to measure distances between different sensorial information, modeling the natural human ability to associate sensations. The model still necessitates a thorough evaluation and its performances have to be compared with the available and consolidated IR techniques in order to confirm its proximity to human behaviors. However, if our intuition is right, it can provide a general method to evaluate relations between digital contents in a way more comprehensible to humans. This approach could have a vast array of applications in information management such as retrieval, clustering and the like.

References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Reviews of Modern Physics **74**(1), 47 (2002)
2. Barrat, A., Barthélemy, M., Vespignani, A.: Dynamical processes on complex networks, pp. 116–135. Cambridge University Press (2008)
3. Cattuto, C., Barrat, A., Baldassarri, A., Schehr, G., Loreto, V.: Collective dynamics of social annotation. Proceedings of the National Academy of Sciences **106**(26), 10511–10515 (2009)
4. Collins, A.M., Loftus, E.F.: A spreading-activation theory of semantic processing. Psychological Review **82**(6), 407–428 (1975)
5. Collins, A., Quillian, M.: Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior **8**, 240–248 (1969)
6. Dall'Asta, L., Barrat, A., Barthélemy, M., Vespignani, A.: Vulnerability of weighted networks, March 2006. arXiv:physics/0603163v1
7. Datta, R., Li, J., Wang, J.Z.: Content-based image retrieval: approaches and trends of the new age. In: Zhang, H., Smith, J., Tian, Q. (eds.) Multimedia Information Retrieval, pp. 253–262. ACM (2005)
8. Deselaers, T., Ferrari, V.: Visual and semantic similarity in imagenet. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, pp. 1777–1784. IEEE Computer Society, Washington, DC (2011)
9. Fang, C., Torresani, L.: Measuring image distances via embedding in a semantic manifold. In: European Conference on Computer Vision, pp. 402–415, October 2012
10. Gravino, P., Servedio, V.D.P., Barrat, A., Loreto, V.: Complex structures and semantics in free word association. Advances in Complex Systems **15**(3–4) (2012)
11. Guarino, N., Oberle, D., Staab, S.: What is an ontology? In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, 2nd edn. Springer (2009)
12. Kurtz, C., Beaulieu, C.F., Napel, S., Rubin, D.L.: A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations. J. of Biomedical Informatics **49**(C), 227–244 (2014)
13. Markatopoulou, F., Mezaris, V., Kompatsiaris, I.: A comparative study on the use of multi-label classification techniques for concept-based video indexing and annotation. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N. (eds.) MMM 2014, Part I. LNCS, vol. 8325, pp. 1–12. Springer, Heidelberg (2014)
14. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: Introduction to wordnet: an on-line lexical database. Int. J. Lexico. **3**, 235–244 (1990)
15. Morais, A.S., Olsson, H., Schooler, L.: Mapping the structure of semantic memory. Cognitive Science **37**, 125–145 (2012)
16. Nelson, D.L., McEvoy, C.L., Schreiber, T.A.: The university of south florida word association norms. http://w3.usf.edu/FreeAssociation
17. van Rijsbergen, K.: The Geometry of Information Retrieval. Cambridge University Press (2004–2007)
18. Roget, P.: Roget's thesaurus of English words and phrases. TY Crowell Co. (1911)
19. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
20. Steyvers, M., Tenenbaum, J.B.: The large scale structure of semantic networks: Statistical analyses and a model of semantic growth. Cognitive Science **29**, 41–78 (2005)
21. Tousch, A., Herbine, S., Audibert, J.: Semantic hierarchies for image annotation: a survey. Pattern Recognition **45**, 333–345 (2012)
22. Gabler, K.: The human brain cloud. http://www.humanbraincloud.com