Set Similarity Measures for Images Based on Collective Knowledge

Valentina Franzoni[1,2(✉)], Clement H.C. Leung[3], Yuanxi Li[3], Paolo Mengoni[1], and Alfredo Milani[1,3]

[1] Department of Mathematics and Computer Science, University of Perugia, Perugia, Italy
valentina.franzoni@dmi.unipg.it, paolo@netdirect.it
[2] Department of Computer, Control and Mgmt. Engineering, University of Rome La Sapienza, Rome, Italy
[3] Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong
{clement,csyxli}@comp.hkbu.edu.hk, milani@dmi.unipg.it

Abstract. This work introduces a new class of group similarity where different measures are parameterized with respect to a basic similarity defined on the elements of the sets. Group similarity measures are of great interest for many application domains, since they can be used to evaluate similarity of objects in term of the similarity of the associated sets, for example in multimedia collaborative repositories where images, videos and other multimedia are annotated with meaningful tags whose semantics reflects the collective knowledge of a community of users. The group similarity classes are formally defined and their properties are described and discussed. Experimental results, obtained in the domain of images semantic similarity by using search engine based tag similarity, show the adequacy of the proposed approach in order to reflect the collective notion of semantic similarity.

Keywords: Group similarity · Semantic distance · Image retrieval · Data mining · Collective knowledge · Knowledge discovery

1 Introduction

Search engines, continually exploring the Web using its semantic meaning, are a natural source of information on which to base a modern approach to semantic annotation. The similarity measurement between documents and text has been extensively studied in information retrieval. [15][17]

Content Based Image retrieval (CBIR) [3][18] enables satisfactory similarity measurements of low level features. However, the semantic similarity of deep relationships among objects is not explored by CBIR or other state-of-the-art techniques in Concept Based Image Retrieval [9][10]. The example in Fig. 1 shows the different similarity recognition of humans and computers.

A promising idea is that it is possible to generalize the semantic similarity, under the assumption that semantically similar terms behave similarly [12][18-20]. In this

O. Gervasi et al. (Eds.): ICCSA 2015, Part I, LNCS 9155, pp. 408–417, 2015.
DOI: 10.1007/978-3-319-21404-7_30

work the features of the main semantic proximity measures used in literature are analyzed and used in a new group similarity measure, as a basis to extract semantic content, reflecting the collaborative change made on the web resources. We propose a Context-based Group Similarity to measure deep semantic similarity of images. The proposed algorithm measures the similarity between a pair of images with clouds including the tags provided by the user i.e., image author/owner.

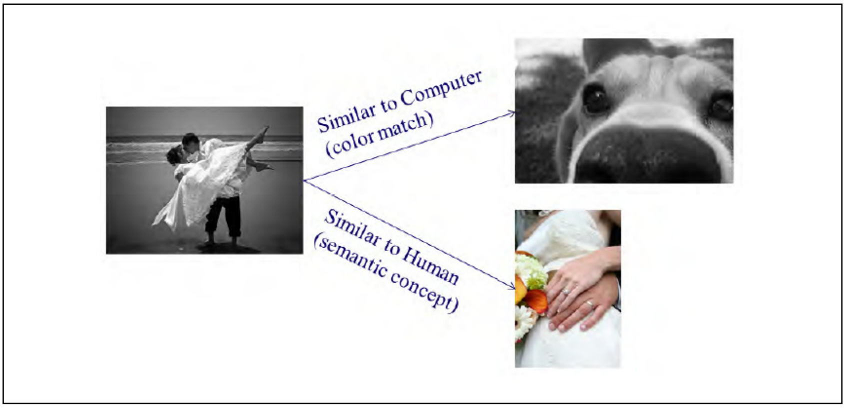


Fig. 1. Image similarity discovery comparison between computer and human

2 Related Work

2.1 WordNet Similarity

WordNet [1], is one of the applications of semantic lexicon propose for the English language and is a general knowledge base and common sense reasoning engine [3]. Recent researches [2] on the topic in computational linguistics has emphasized the perspective of semantic relatedness of two lexemes in a lexical resource, or its inverse, semantic distance. The work in [3] brings together ontology and corpus, defining the similarity between two concepts *c1* and *c2* lexicalized in WordNet, named *WordNet Distance (WD)*, by the information content of the concepts that subsume them in the taxonomy. Then, [4] proposes a similarity measure in WordNet between arbitrary objects:

$$d(c_1, c_2) = \frac{2 \times \log p(lso(c_1, c_2))}{\log p(c_1) + \log p(c_2)} \tag{1}$$

The advantage of a WordNet similarity is that it is a very mature and comprehensive lexical database, which provides measures of similarity and relatedness: it reflects universal knowledge because it is built by human experts, however WordNet Distance is only for nouns and verbs in WordNet, which is not dynamically updated.

2.2 Wikipedia Similarity

WikiRelate [5] was the first research to compute measures of semantic relatedness using Wikipedia. This approach took familiar techniques that had previously been applied to WordNet and modified them to suit Wikipedia. The implementation of WikiRelate follows the hierarchical category structure of Wikipedia.

The *Wikipedia Link Vector Model (WLVM)* [6] uses Wikipedia to provide structured world knowledge about terms of interest.

These approaches use the hyperlink structure of Wikipedia rather than its category hierarchy or textual content [7][11]. The probability of WLVM is defined by the total number of links to the target article over the total number of articles. Thus if t is the total number of articles within Wikipedia, the weighted value w for the link $a \rightarrow b$ is:

$$w(a \rightarrow b) = |a \rightarrow b| \times \log\left(\sum_{x=1}^{t} \frac{t}{|x \rightarrow b|}\right) \tag{2}$$

where a and b denote the search terms.

Wikipedia similarity reflects relationships as seen by the user community [18][19], which is dynamically changing, as links and nodes are changed by the users' collaborative effort. However, it only can apply to knowledge base organized as networks of concepts.

2.3 Flickr Similarity

Flickr distance (FD) [8] is another model for measuring the relationship between semantic concepts in visual domain. For each concept, a collection of images is obtained from Flickr, based on which the improved latent topic-based visual language model is built to capture the visual characteristics of the concept. The Flickr distance between concepts *c1* and *c2* can be measured by the *square root of Jensen-Shannon divergence* [10, 11] between the corresponding visual language models as follows:

$$\mathrm{D}(\mathrm{C\_1}, \mathrm{C\_2}) = \sqrt{\frac{\sum_{i=1}^{K}\sum_{j=1}^{K} \mathrm{D_{JS}}(\mathrm{P_{Zi}C_1}|\mathrm{P_{Zj}C_2})}{\mathrm{K}^2}} \tag{3}$$

where

$$\mathrm{D_{JS}} = \left(\mathrm{P_{Zi}C_1}|\mathrm{P_{Zj}C_2}\right) = \frac{1}{2}\mathrm{D_{KL}}(\mathrm{P_{Zi}C_1}|\mathrm{M}) + \frac{1}{2}\mathrm{D_{KL}}\left(\mathrm{P_{Zj}C_2}|\mathrm{M}\right) \tag{4}$$

K is the total number of latent topics, which is determined by experiment. $P_{Zi}\ C_1$ and $P_{Zj}\ C_2$ are the trigram distributions under latent topic *zic1* and *zjc2* respectively, with M representing the mean of $P_{Zi}\ C_1$ and $P_{Zj}\ C_2$. The FD is based on visual language models (VLM), which is a different concept relationship respect to WordNet Similarity and Wikipedia Similarity.

3 Context-Based Semantic Group Distance

3.1 Web-Based Proximity Measures

We use Web-based proximity Measurements [18] as compositor of Group Distance because it can be applied to any retrieval engine. It makes good use of the occurrence/co-occurrence of words, terms, tags, users, objects etc. It reflects the user community current state of believes and dynamically changes as results change [16][11].

3.1.1 Normalized Google Distance (NGD)

Normalized Google distance (NGD) [9] quantifies the extent of the relationship between two concepts by their correlation in the search results from Google search engine when querying both concepts, with

$$\mathrm{NGD}(x,y) = \frac{\max\{\log f(x), \log (y)\} - \log [f(x,y)]}{\log N - \min\{\log f(x), \log [f(y)]\}} \tag{5}$$

Where *f(x)* and *f(y)* are the numbers of the web pages and documents returned by Google search engine when typing *x* and *y* as the search term respectively, with *f(x,y)* denoting the number of pages containing both *x* and *y*. *N* denotes the total number of documents indexed by Google search, which is deductible or approximable by any number greater than every page count of the results.

3.1.2 Pointwise Mutual Information (PMI)

Mutual Information (MI) is a measure of the information overlap between two random variables.

Pointwise Mutual Information (PMI) [13][14] is a point-to-point measure of association, which represents how much the actual probability of a particular co-occurrence of events differs from what we would expect it to be on the basis of the probabilities of the individual events and the assumption of their independence. Even though PMI may be negative or positive, its expected outcome over all joint events (i.e., PMI) is positive.

PMI is used both in statistics and in information theory.

PMI between two particular events w_1 and w_2, in this case the occurrence of particular words in Web-based text pages, is defined as follows:

$$PMI(w_1, w_2) = log_2 \frac{P(w_1,w_2)}{P(w_1)P(w_2)} \tag{6}$$

This quantity is zero if w_1 and w_2 are independent, positive if they are positively correlated, and negative if they are negatively correlated.

On particularly low frequency data, PMI does not provide reliable results [19]. Since PMI is a ratio of the probability of w_1, w_2 together and w_1, w_2 separately, in the case of perfect dependence, PMI will be *0*.

PMI is a bad measure of dependence, since the dependency score is related to the frequency of individual words. PMI could not always be suitable when the aim is to compare information on different pairs of words.

3.2 Confidence and Average Confidence (CM)

Given a rule *X→Y*, *Confidence* [4] is a statistical measure that, given the number of transactions that contain *X*, indicates the percentage of transactions that contain also *Y*.

$$confidence(x \to y) = \frac{P(x,y)}{P(x)} = \frac{f(x,y)}{f(x)} \tag{7}$$

From a probabilistic point of view, confidence approximates the conditional probability:

$$confidence(x \to y) = \frac{\mathrm{P}(x \wedge y)}{\mathrm{P}(x)} = \mathrm{P}(y|x) \tag{8}$$

Since Confidence is not a symmetric measure, the Average Confidence (CM) [19] can be defined a

$$CM = \frac{confidence(\mathrm{x \to y}) + confidence(\mathrm{y \to x})}{2} \tag{9}$$

3.3 Context-Based Group Similarity

We propose a Context-based Group Similarity to measure deep semantic similarity of images. The proposed algorithm measures the image similarities with semantic proximity based one user provided concept clouds. Image semantic concept clouds include any semantic concept associated to or extracted from images. Typical sources for semantic concepts are tags, comments, descriptors, categories, or text surrounding the image. As shown in Fig. 2, Image *Ii* and Image *Ij* are a pair of images to be compared. *Ti1*, *Ti2,…, Tim* are original user provided tags of image *Ii*, while *Tj1; Tj2,.., Tjn* are original user provided tags of image *IJ* .

Given *DIij* as the distance (or equivalently, the similarity) of image *Ii* and image *Ij*, we define the ***Group Distance (GD)***:

$$DI_{ij} = AVG2\left\{AVG1[SEL(dT_{im \to jn})], AVG1[SEL(dT_{jn \to im})]\right\} \tag{10}$$

where *SEL* could be the maximum *MAX,* the average *AVG* or the minimum *MIN* of *d*, the similarity calculated by algorithm (*Confidence* or *NGD* or *PMI*), as in equations (11-14).

$$dT_{im} \to dT_{jn} = \begin{pmatrix} dT_{i1\to j1}, & dT_{i1\to j2}, & dT_{i1\to j3}, & \dots & dT_{i1\to jn} \\ dT_{i2\to j1}, & dT_{i2\to j2}, & dT_{i2\to j3}, & \dots & dT_{i2\to jn} \\ \dots, & \dots, & \dots, & \dots & \dots \\ dT_{in\to j1}, & dT_{in\to j2}, & dT_{in\to j3}, & \dots & dT_{in\to jn} \end{pmatrix}$$

$$dT_{im} \to dT_{jn} = \begin{pmatrix} dT_{j1\to i1}, & dT_{j1\to i2}, & dT_{j1\to i3}, & \dots & dT_{j1\to im} \\ dT_{j2\to i1}, & dT_{j2\to i2}, & dT_{j2\to i3}, & \dots & dT_{j2\to im} \\ \dots, & \dots, & \dots, & \dots & \dots \\ dT_{jn\to i1}, & dT_{jn\to i2}, & dT_{jn\to i3}, & \dots & dT_{jn\to im} \end{pmatrix} \tag{11}$$

$$AVG1[SEL(dT_{im\to jn})] = avg[SEL(dT_{i1\to jn}), SEL(dT_{i2\to jn}), , SEL(dT_{im\to jn})] \tag{12}$$

$$AVG1[SEL(dT_{jn\to im})] = avg[SEL(dT_{j1\to im}), SEL(dT_{j2\to im}), , SEL(dT_{jn\to im)}] \tag{13}$$

$$AVG2 = AVGAVG1, AVG2 \tag{14}$$

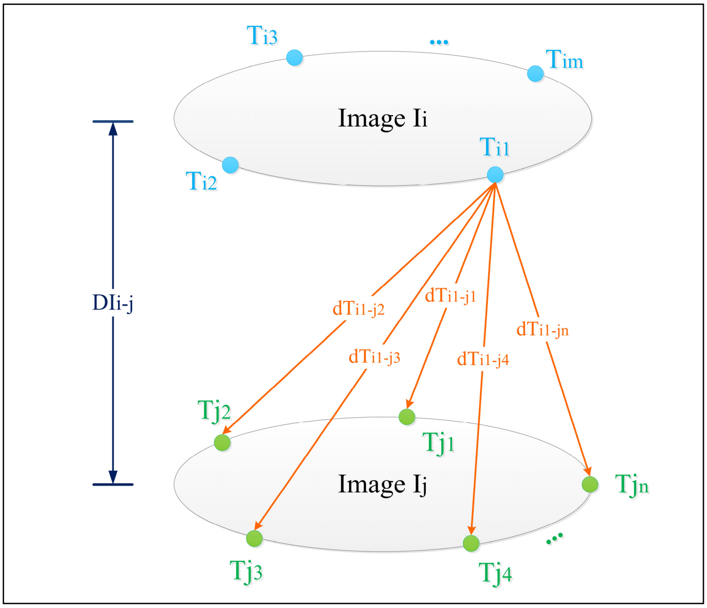


Fig. 2. Group Similarity Core Algorithm

We calculate the similarity of each pair of images, from the users' votes of how much similar of certain pair of images is. Users can score each pair of images with a number from 0 (=very different) to 5 (=very similar), based on their personal opinion.

The quantitative experimental results which compare the users' scores and proposed Context-based Group Distance will be shown in section 4.

The core algorithm of the proposed Context-based Group Distance calculations is shown in Figure 2.

4 Experimental Results

The system is evaluated quantitatively using more than 500 pairs of Web images on Flickr to compare the users' scores and the proposed Context-based Group Distance. We keep the tags as original ones provided by the owner of images in Flickr.

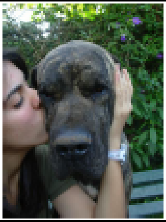
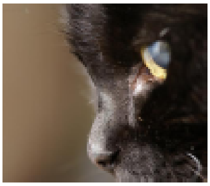
Fig. 3 shows a sample of image pairs similarity experiment results comparing user score and proposed Group Similarity.

Table 1. Experimental Results on Bing for avg-avg-avg GD

image1	image2	User Score	GD-Confidence	GD-NGD	GD-PMI
1	2	0.47826	0.39801	0.744305	5.99773
3	4	0.53043	0.33722	0.721759	6.24246
5	4	0.37391	0.29569	0.705348	6.0838
6	4	0.41739	0.28847	0.680995	5.78023
2	7	0.32174	0.27788	0.732283	5.94255
8	9	0.29565	0.25935	0.595794	4.23411
1	10	0.43478	0.24937	0.658751	5.13127
11	6	0.75652	0.23742	0.576982	5.40974
12	13	0.31304	0.22882	0.711011	6.05964
4	14	0.13043	0.2266	0.674191	5.52523
4	15	0.16522	0.21048	0.734764	6.24088
…	…	…	…	…	…

Table 1 shows a small subset of experimental results. *User Score* represents the average human evaluation score. *GD-Confidence, GD-NGD* and *GD-PMI* are respectively the Group Distance for Confidence, Normalized Google Distance and Pointwise Mutual Information.

Experimental results indicate that this approach can give more accurate and efficient similarity measurements than calculations based on image low level features, in terms of the deep semantic concept similarity.

Image Pair Number	Sample Image Pairs		User Voted Average Score	Similarit Calculate by Group Similarit
19-24			0.47826087	0.398011
Original Tag ID and Tags	19936 beagle 19936 dog 19936 black 19936 white 19936 nose	16054 dog 16054 people 16054 kiss		
19-20			0.434782609	0.249373
Original Tag ID and Tags	19936 beagle 19936 dog 19936 black 19936 white 19936 nose	17405 cat 17405 eye 17405 nose 17405 fur		
19-47			0.07826087	0.168459
Original Tag ID and Tags	19936 beagle 19936 dog 19936 black 19936 white 19936 nose	18239 airplane 18239 aviation 18239 blue 18239 california 18239 commercial		

Searching Engine: Bing
Similarity Type: Confidence
SEL Option Used: Average (AVG)

Fig. 3. Sample set of image pairs similarity experiment results – comparing user score and proposed Group Similarity

5 Conclusion

In this work we present our approach to measure the context-based group similarity among concepts and images. Comparing with Content Based Image Retrieval (CBIR), which measures the image content similarities by low level features, the proposed Context-based Image Similarity outperforms CBIR in effectiveness and accuracy of comparing the deep concept among images, since in order to recognize low level features in the CBIR higher computation capabilities are required. The systems are evaluated quantitatively involving the user judgments. Experimental results indicate that our approaches could enable advanced degree of semantic similarity measurements and deep meaning enrichment of images, also delivering highly competent performance, attaining excellent precision and efficiency.

References

1. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM **38**(11), 39–41 (1995)
2. Budanitsky, A., Hirst, G.: Semantic distance in wordnet: an experimental, application-oriented evaluation of five measures. In: Proceedings of Workshop on WordNet and Other Lexical Resources, p. 641. North American Chapter of the Association for Computational Linguistics, Pittsburgh (2001)
3. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 448–453 (1995)
4. Lin, D.: An information-theoretic definition of similarity. In: Proceedings of the 15th International Conference on Machine Learning, pp. 296–304. Morgan Kaufmann (1998)
5. Strube, M., Ponzetto, S.P.: WikiRelate! computing semantic relatedness using wikipedia. In: Proceedings of the Twenty-First National Conference on Artificial Intelligence. AAAI Press, July 2006
6. Milne, D., Witten, I.H.: An effective, low-cost measure of semantic relatedness obtained from wikipedia links. In: Proceedings of first AAAI Workshop on Wikipedia and Artificial Intelligence, WIKIAI 2008, Chicago, IL, USA (2008)
7. Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., Studer, R.: Semantic wikipedia. In: Proceedings of the 15th International Conference on World Wide Web, WWW 2006, pp. 585–594. ACM, New York (2006)
8. Wu, L., Hua, X.-S., Yu, N., Ma, W.-Y., Li, S.: Flickr distance. In: Proceedings of the 16th ACM International Conference on Multimedia, MM 2008, New York, NY, USA, pp. 31–40 (2008)
9. Enser, P.G., Sandom, C.J., Lewis, P.H.: Surveying the reality of semantic image retrieval. In: Bres, S., Laurini, R. (eds.) VISUAL 2005. LNCS, vol. 3736, pp. 177–188. Springer, Heidelberg (2006)
10. Li, X., Chen, L., Zhang, L., Lin, F., Ma, W.: Image annotation by large-scale content-based image retrieval. In: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 607–610 (2006)
11. Franzoni, V., Milani, A.: PMING Distance: A collaborative semantic proximity measure. In: 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (IAT), vol. 2, pp. 442–449 (2012)

12. Leung, C.H., Li, Y., Milani, A., Franzoni, V.: Collective evolutionary concept distance based query expansion for effective web document retrieval. In: Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) ICCSA 2013, Part IV. LNCS, vol. 7974, pp. 657–672. Springer, Heidelberg (2013)
13. Manning, D., Schutze, H.: Foundations of statistical natural language processing. The MIT Press, London (2002)
14. Turney P.: Mining the web for synonyms: PMI versus LSA on TEOFL. In Proc. ECML (2001)
15. Chan, A.W.S., Liu, J., et al.: Intelligent Social Media Indexing and Sharing Using an Adaptive Indexing Search Engine. ACM TIST **3**(3), 47 (2012). doi:10.1145/2168752.2168761
16. Li, Y.X.: Semantic Image Similarity Based on Deep Knowledge for Effective Image Retrieval. Research Thesis (2014)
17. Cheng, V.C., Liu, J., et al.: Probabilistic Aspect Mining Model for Drug Reviews. IEEE Transactions on Knowledge and Data Engineering **99**, 1 (2014). doi:10.1109/TKDE.2013.175. vol. 99, no. PrePrints, p. 1
18. Franzoni, V., Milani, A.: Heuristic semantic walk for concept chaining in collaborative networks. International Journal of Web Information Systems **10**(1), 85–103 (2014). doi:10.1108/IJWIS-11-2013-0031
19. Franzoni, V., Mencacci, M., Mengoni, P., Milani, A.: Heuristics for semantic path search in Wikipedia. In: Murgante, B., et al. (eds.) ICCSA 2014, Part VI. LNCS, vol. 8584, pp. 327–340. Springer, Heidelberg (2014)
20. Franzoni, V., Milani, A.: Heuristic Semantic Walk. In: Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) ICCSA 2013, Part IV. LNCS, vol. 7974, pp. 643–656. Springer, Heidelberg (2013)
21. Franzoni V., Milani A.: Semantic Context Extraction from Collaborative Networks. In: IEEE International Conference on Computer Supported Cooperative Work in Design (CSCWD), Calabria, Italy (2015)