Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

ANNOTATING IMAGES BY SEMANTIC REPRESENTATION USING THE OPEN KNOWLEDGE BASE

J. M. Ghazali[1,*], S. N. Ahmad[1], K. A. Elias[1], S. M. N. Khan[1], S. A. M. Noah[2], L. Q. Zakaria[2]

[1]Kolej Universiti Islam Antarabangsa Selangor (KUIS), Malaysia

[2]Universiti Kebangsaan Malaysia (UKM), Malaysia

Published online: 24 February 2018

ABSTRACT

A good image annotation scheme is highly desired especially when images with crude description provide inadequate information and images with no description are not accessible by text based search. In the area of image annotation, this study aims to propose a new approach by combining image low level features and semantics available in open knowledge base. Image classification is one of the steps in image annotation. The best classifier was determined by conducting a comprehensive experiment where various machine learning algorithms performances were compared. Using feature extraction, initial tag population were generated by retrieving tags from the most similar images identified. Experiments were carried out to determine the best parameters that yield the best performance. Finally, tags related to domain of interest were given semantic meaning by optimizing ontologies and the open knowledge base. Comparing image annotation performance before and after linking to the open knowledge base is the main evaluation of this study. Evaluation is based on the standard performance metrics; precision, recall, and F-Measure. This study demonstrates that representing the identified concept of image annotation semantically is most useful in increasing image annotation performance.

Keywords: image annotation, image classification, semantic, open knowledge base.

Author Correspondence, e-mail: author@gmail.com

doi: http://dx.doi.org/10.4314/jfas.v10i3s.59

1. INTRODUCTION

The advancement of Web applications has led to a significant growth of multimedia content sharing and accessibility. One of the most used media is images. Images on the Web are mainly reached by text-based searching where text query is mapped to text description of images. Textual description that is assigned to images is called image annotation.

There are a variety of approaches that have been adopted by researchers in the area of image annotation. The researchers adopted a combination of techniques and approaches. These include image analysis which takes into account the visual feature of an image like in Wang et al. (2008a) and Khan (2006). Other than that, is the text analysis approach which includes natural language processing like in Zhou et al. (2007), Wang et al. (2008b), Deschacht and Moens (2007), Xia et al. (2008), and Sigurbjornsson and van Zwol (2008), tags exploitation like in Wang et al. (2008a) and Sigurbjornsson and van Zwol (2008), and using ontologies like in Sahrani and Houlari (2015), Khan (2006) and Shi et al. (2007).

Adopting a combination of various approaches seems very promising; optimizing the potential of each approach. This study is somewhat similar yet a distinctive set of approaches were adopted in the quest to improve image annotation results. The study has undertaken image and text analysis approaches to formulate a unique scheme for image annotation which includes optimizing the open knowledge base.

2. PROPOSED APPROACH

Fig. 1 illustrates the proposed approach to image annotation. Social media sharing websites like Flickr offers rich annotations that can be taken advantage of. Features database is made available by careful selection and extraction of tagged Flickr images. These features are used for two purposes; the first is to classify images accordingly and the second, to identify the degree of similarity of images. Concepts of image classification are chosen manually based on popular and distinctive visual characteristics.

Tags (or annotation) of images that are very similar with the target image are used as seeds or initial annotation for the latter. This is done by first computing the feature of both target image and annotated images, and then searching is done to find visually similar images via a similarity measure technique. Annotations associated to these similar images are appointed as initial annotation to the target image. As there can be more than one similar image, a suitable strategy for merging such annotations is necessary. Furthermore, as Flickr social based annotation is sometimes prone to 'junks' further refinement is non-trivial. It is at the refinement stage whereby the annotations are mapped to the open knowledge base in order to further enrich its semantic meanings. The automatic mapping of these annotations will be a challenging task as the open knowledge base contains huge information and its content is frequently unknown to annotators.

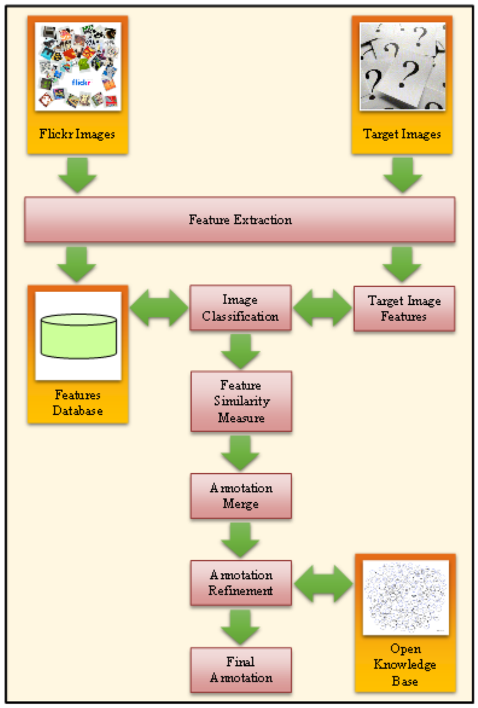

Fig.1. Proposed approach to semantic image annotation

2.1. Semantic Representation Using Open Knowledge Base

Specific knowledge bases are required to identify concepts that are related to the domain. Three knowledge bases were used, namely, GeoNames, DBpedia and Malaysia Tourism Ontology.

GeoNames is a geographical database that is accessible by a number of webservices. We use a webservice to search the location in GeoNames base on the latitude and longitude value retrieved from the image metadata. At the same time, nearby locations are also retrieved by linking to DBpedia.

Malaysia Tourism Ontology (Lailatul, 2011) stores information related to tourism in Malaysia consisting of two main roots, that is, Attraction and Event. Both roots have information such as names, descriptions and locations. This study is particularly interested in locations.

The following are the steps and rules of action in optimizing the three knowledge bases in giving semantic representation as illustrated in Fig. 2:

i. Given latitude and longitude value, the location name is retrieved from GeoNames and DBpedia

ii. MTO categorizes it being in what state and country.

iii. MTO is used to identify named location of tags and compare it with the location names found in (i) and (ii). If there is no name location identified in the tags, the generated location names are added as tags. If there is a match, the name location is ignored. Whereas if there is conflict, the initial location tag is replaced by the generated location names.

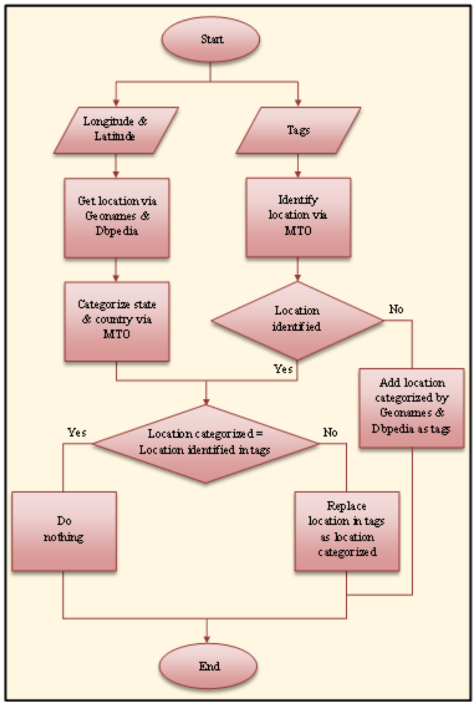

Fig.2. The flow of how the three knowledge bases are used to improve tags.

3. RESULTS AND DISCUSSION

3.1. Image Classification

Comparative results of all classifiers are shown on corrected classified instances which are measured by percentage, time taken to produce results, kappa statistics, false positive, precision, recall, F-Measure, ROC Area and accuracy.

Tables 1 and 2 show the classification performance for various learning algorithms; Support Vector Machine (SVM), Multilayer Perceptron (MP), Bagging (BG), DECORATE (DEC), C4.5 Decision Tree (C4.5), and Random Forest (RF).

Table 1 shows the scores of the classifiers with 64 bins. Bagging showed better performance with an accuracy of 72.8%, slightly outperforming DECORATE by 0.7%. A specific experiment was carried out by (Pal, 2007) to evaluate the effect of noise on the classification performance using four different ensemble approaches; bagging, DECORATE, random subspace and boosting. The fact that bagging is able to handle dataset with noise the best, as reported by (Pal, 2007), may be a contributing factor to the results.

The scores for classification with 216 bins is as shown in Table 2. DECORATE showed the best performance with 87.6% accuracy. This is followed by Bagging and Random Forest with both achieving 86.8% accuracy.

Table 1. Results for classification with 64 bins

Algo	Corrected Classified Instances	Time taken	Kappa Statistic	False Positive	Precision	Recall	F-Measure	ROC Area	Accuracy
SVM	46.0	0.09	0.270	0.194	0.457	0.460	0.449	0.633	63.3
MP	46.0	5.63	0.281	0.177	0.469	0.460	0.456	0.701	70.1
BG	50.6	0.19	0.342	0.163	0.509	0.506	0.499	0.728	72.8
DEC	48.9	1.51	0.316	0.172	0.492	0.489	0.482	0.721	72.1
C4.5	43.5	0.08	0.243	0.196	0.420	0.435	0.415	0.631	63.1
RF	47.2	0.06	0.294	0.179	0.478	0.473	0.468	0.707	70.7

Table 2. Results for classification with 216 bins.

Algo.	Corrected Classified Instances	Time taken	Kappa Statistic	False Positive	Precision	Recall	F-Measure	ROC Area	Accuracy
SVM	64.1	0.17	0.525	0.115	0.676	0.641	0.639	0.763	76.3
MP	47.3	72.29	0.314	0.149	0.623	0.473	0.488	0.775	77.5
BG	66.8	0.56	0.552	0.119	0.662	0.667	0.656	0.868	86.8
DEC	68.8	3.72	0.580	0.114	0.685	0.688	0.683	0.876	87.6
C4.5	61.2	0.20	0.474	0.145	0.594	0.612	0.597	0.730	73.0
RF	66.7	0.13	0.549	0.126	0.662	0.667	0.656	0.868	86.8

Overall, DECORATE has shown the best performance in this experiment. In terms of time taken, Random Forest seems to be the quickest and Multilayer Perceptron being the slowest. Different number of bins can reveal different features of data. In this experiment, representing image features as 216 bins performed better than the 64 bins. The results show that all learning algorithms performed best with 216 bins with an increased accuracy between 7.4% - 16.1% as compared to the 64 bins.

The result presented indicates the potential performance of DECORATE in classifying images especially involving small training set. The best overall performance shown by DECORATE is consistent to Melville & Mooney (2004) where they reported that DECORATE is consistently more accurate than the base classifier of Bagging, AdaBoost and Random Forests. But given large training sets, DECORATE is still competitive with AdaBoost and outperform Bagging and Random Forests.

Another experiment carried out by (Caruana et al., 2006) reported that Random Forest perform about 0.6% better than the next best method, ANN. This is followed by boosted decision trees and SVMs. These results are in line with ours where Random Forest has outperformed decision trees and SVM. The aforesaid experiment however did not include DECORATE.

Due to the high effectiveness and reliability of using multi-class SVM in image classification as reported by previous researchers, SVM was chosen without any specific experiments done (Molitorisová, 2012; Lindstaedt et al., 2009; Khan, 2007; and Cusano et al., 2003). Nonetheless, this study concluded that other classification algorithms perform better than SVM.

Table 3 shows the accuracy category for the various classifiers based on the area under the ROC curve. It clearly shows that the DECORATE, Random Forest and Bagging model achieved ‘good’ discriminating ability as compared to other classifiers. However, DECORATE has slightly higher ROC area as compared to the Random Forest and Bagging classifier.

Table 3. Overall classification performance of six machine learning algorithms

Algo.	ROC Area (64 bins)	Accuracy Category	ROC Area (216 bins)	Accuracy Category
SVM	0.633	Poor	0.763	Fair
MP	0.701	Fair	0.775	Fair
BG	0.728	Fair	0.868	Good
DEC	0.721	Fair	0.876	Good
C4.5	0.631	Poor	0.730	Fair
RF	0.707	Fair	0.868	Good

3.2. Tag Population

The performance of the four classes is measured by calculating the average F-Measure using ten images with varying parameters. These are shown in Tables 4 and 5. The horizontal header in table 4 denotes the number of similar images taken and considering with and without word co-occurrence. For example, Top10 represent 10 similar images selected. The vertical header in Table 5, for example HF7, denotes the top seven words with the highest frequency chosen.

Table 4. Performance of four classes on average base on the number of similar images selected and impact of word co occurrence

	Top10		**Top15**		**Top20**		**Top25**		**Top30**	
	With-out WC	with WC	With-out WC	with WC	With-out WC	with WC	With-out WC	with WC	With-out WC	with WC
Beach	**0.590**	**0.591**	**0.601**	**0.602**	**0.619**	**0.620**	**0.640**	**0.638**	**0.622**	**0.614**
Building	**0.562**	**0.555**	**0.574**	**0.538**	**0.589**	**0.552**	**0.558**	**0.553**	**0.569**	**0.563**
Festival	**0.645**	**0.615**	**0.636**	**0.611**	**0.624**	**0.608**	**0.613**	**0.599**	**0.625**	**0.627**
Mountain	**0.574**	**0.570**	**0.617**	**0.617**	**0.603**	**0.603**	**0.586**	**0.586**	**0.587**	**0.587**
Average	0.593	0.583	0.607	0.592	0.608	0.596	0.599	0.594	0.601	0.598

Table 5. Performance of four classes on average base on the number of top highest frequency selected

	Beach	Building	Festival	Mountain	Average
HF5	N/A	**0.563**	N/A	**0.559**	0.561
HF6	N/A	**0.574**	**0.615**	**0.595**	0.595
HF7	**0.591**	**0.560**	**0.621**	**0.594**	0.592
HF8	**0.610**	**0.560**	**0.624**	**0.617**	0.603
HF9	**0.622**	**0.549**	**0.620**	**0.599**	0.597
HF10	**0.626**	N/A	**0.622**	N/A	0.624
HF11	**0.619**	N/A	N/A	N/A	0.619

Table 4 illustrates that taking the top 20 similar images and without considering word co-occurrence produces the best result on average. On the other hand, Table 5 shows average best performance which includes all four classes is taking the top eight highest frequency. Although the top ten highest word frequency yields the better performance, but it does not include all classes. Thus, top eight highest word frequency and not considering word co-occurrence are selected from top 20 similar images parameter.

3.3. Before linking to the Open Knowledge Base

Table 6 illustrates the performance of image annotation for all four classes before linking to the Open Knowledge Base. On average, with performance of 0.685, non-visual aspect performs better than the visual aspect with performance of 0.581. Building and Festival class perform best with a result of 0.660 whereas beach class performs least with 0.607 performance value.

With performance values 0.666 and 0.643 respectively, on average, precision has shown slightly better results compared to recall. This shows that the precision values are significantly more than the recall values for visual aspect throughout all classes except for class Mountain where precision equals recall. Recall seems to be significantly more than precision for non-visual aspect. The result is consistent throughout all classes except for class Beach where precision is more than recall.

Table 6. Performance of image annotation for all four classes, before linking to the open knowledge base

	Visual			**NonVisual**			**Average**		
	Preci-sion	Re-call	F-Mea-sure	Preci-sion	Re-call	F-Mea-sure	Preci-sion	Re-call	F-Mea-sure
Beach	0.719	0.447	0.541	0.810	0.602	0.683	0.750	0.513	0.608
Building	0.881	0.565	0.659	0.556	0.815	0.657	0.632	0.697	0.660
Festival	0.936	0.524	0.654	0.595	0.756	0.658	0.651	0.679	0.660
Mountain	0.492	0.492	0.471	0.715	0.784	0.743	0.630	0.683	0.654
Average	0.757	0.507	0.581	0.669	0.740	0.685	0.666	0.643	0.646

The number of words distributed in the benchmark annotation is related to low recall values in the visual aspect and high values in the non-visual aspect of image annotation. Visual aspect seems to have more words than non-visual aspect does in the benchmark annotation. This gives impact to recall as correctness is reflected on the benchmark annotation. Thus, because visual aspect of the benchmark annotation has more words then the recall value is low. However, because non-visual aspect of the benchmark annotation has less words, then the recall value is high.

As an exception, we observe that for class Beach, recall is lower than precision for non-visual aspect. This is probably due to the fact that class Beach has an average benchmark annotation much more than other classes. The average benchmark annotation for class Beach is 12 compared to the annotation scheme chosen, that is, eight. Thus, this gives impact to recall as word counts are fairly distributed.

3.4. Linking to the Open Knowledge Base

All four classes have shown results in an increase of image annotation performance after linking to the open knowledge base as shown in Fig. 3. The improvement of performance gives a T-test result of 98% confidence level significance.

Image annotation were categorized and evaluated in two main aspects; visual and non-visual. From Table 7, we can see that there was no improvement of performance for all four classes of visual aspect annotation after linking to the open knowledge base. Nonetheless, improvements were recorded entirely on the non-visual aspect. Class Beach recorded the highest average performance improvement from 0.601 to 0.824 which is a difference of

0.224. This is followed by class Mountain with an improvement of 0.129 and scored the highest non-visual annotation performance of 0.903.

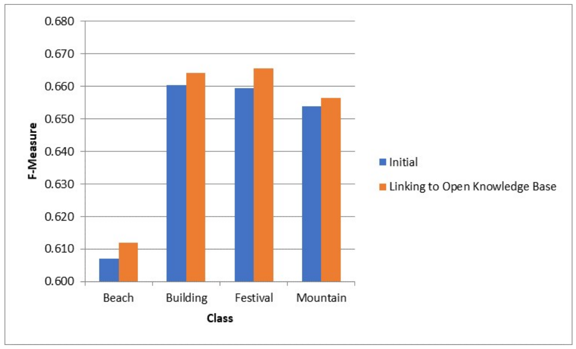

Fig.3. Image annotation performance for all four classes comparing before and after linking to the Open knowledge base.

All improvements observed in image annotation performance were wholly on the non-visual aspect of annotation. This is apparent since linking to the open knowledge base caters for location suggestion and/or rectification of image annotation and that location falls under non-visual aspect of image annotation.

Table 7. Visual and non-visual aspect of image annotation performance for all four classes, before and after linking to the open knowledge base

Class	Aspect	Before	After	Difference
Beach	Visual	0.676	0.676	0.000
	Non-Visual	0.601	0.824	0.224
Building	Visual	0.730	0.730	0.000
	Non-Visual	0.654	0.741	0.087
Festival	Visual	0.626	0.626	0.000
	Non-Visual	0.783	0.818	0.035
Mountain	Visual	0.484	0.484	0.000
	Non-Visual	0.774	0.903	0.129

4. CONCLUSION

All four classes showed an increase of image annotation performance after optimizing the open knowledge base. Improvements were observed totally on the non-visual aspect of image annotation. This is expected since linking to the open knowledge base accommodates for location suggestion and/or rectification of image annotation and that location falls under non-visual aspect of image annotation. This approach, however, is only applicable to images that have longitude and latitude values in their EXIF files.

From this research, it is proven that populating tags from similar tagged images via extracting low-level features and applying semantic meaning by linking to the open knowledge base has shown an improved image annotation performance.

REFERENCES

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning, June 25-29, 2006, Pittsburgh, PA., USA., pp: 161-168. doi:10.1145/1143844.1143865.

Cusano, C., Ciocca, G., Schettini, R., (2003). Image annotation using SVM. Proceedings Of Internet imaging IV, SPIE, January 20, 2003, Santa Clara, CA, USA.

Deschacht, K., and Moens, M.F. (2007). Text analysis for automatic image annotation. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. 1000-1007.

Khan, L. (2007). Standards for image annotation using Semantic Web. Computer Standards & Interfaces. 29, 196-204.

Lindstaedt S, Morzinger R, Sorschag R, Pammer V, Thallinger G. (2009). Automatic image annotation using visual content and folksonomies. Multimed Tools Appli., 42: 97-113.

Molitorisová, E. (2012). Automatic Semantic Image Annotation System, *International scientific conference of PhD. students and young scientists and pedagogues* September 19-20, 2012, Nitra, Slovak Republic, pp: 414-419.

Melville, P. and Mooney, R. (2004). Diverse ensembles for active learning. In Proceedings of the International Conference on Machine Learning (ICML), pages 584–591. Morgan Kaufmann, 2004.

Pal, M., (2007). Ensemble learning with decision tree for remote sensing classification. In: Proceedings of World Academy of Science, Engineering and Technology, vol. 26, pp. 735–737

Sahlani, H., & Hourali, M. (2015). A Novel Semantic Statistical Model for Automatic Image Annotation Using Ontology. Majlesi Journal of Multimedia Processing, 4(2).

Shi, R., Lee, C.H., and Chua, T.S. (2007). Enhancing Image Annotation by Integrating Concept Ontology and Text-based Bayesian Learning Model. In Proceeding of 14th ACM International Conference on Multimedia.

Sigurbornsson B and van Zwol R. (2008). Flickr tag recommendation based on collective knowledge. In WWW '08: Proceeding of the 17th international conference on World Wide Web, pages 327{336, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2.

Wang, C., Jing, F., Zhang, L., and Zhang, H.J. (2008a). Scalable search-based image annotation. Multimedia Systems, 14(4):205–220.

Wang, M., Zhou, X., and Chua, T.S. (2008b). Automatic image annotation via local multi-label classification. In ACM CIVR. pages 17-26, New York, NY, USA.

Xia, D., Wu, F., Zhuang, Y. (2008). Search-Based Automatic Web Image Annotation Using Latent Visual and Semantic Analysis. Pacific-Rim Conf. on Multimedia. p.842-845.

Zakaria, Lailatul Qadri. (2011). PhD Thesis. Analysing the content of Web 2.0 documents by using a hybrid approach.

Zhou, X., Wang, M., Zhang, Q., Zhang, J., and Shi, B. (2007). Automatic image annotation by an iterative approach: Incorporating keyword correlations and region matching. In CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval. pages 25-32, New York, NY, USA.

How to cite this article:
Ghazali J M, Ahmad S N, Elias K A, Khan S M N, Noah S A M, Zakaria L Q. Annotating images by semantic representation using the open knowledge base. J. Fundam. Appl. Sci., 2018, *10(3S), 691-702.*