# Usage Based Tag Enhancement of Images

Balaji Vasan Srinivasan[1], Noman Ahmed Sheikh[2], Roshan Kumar[3], Saurabh Verma[4], and Niloy Ganguly[5]

[1] BigData Experience Lab, Adobe Research, Bangalore, India
balsrini@adobe.com
[2] Indian Institute of Technology, Delhi
nomanahmedsheikh11@gmail.com
[3] Indian Institute of Technology, Kanpur
roshankr1995@gmail.com
[4] Indian Institute of Technology, Rourkee
saurv4u@gmail.com
[5] Indian Institute of Technology, Kharagpur
ganguly.niloy@gmail.com

**Abstract.** Appropriate tagging of images is at the heart of efficient recommendation and retrieval and is used for indexing image content. Existing technologies in image tagging either focus on what the image contains based on a visual analysis or utilize the tags from the textual content accompanying the images as the image tags. While the former is insufficient to get a complete understanding of how the image is perceived and used in various context, the latter results in a lot of irrelevant tags particularly when the accompanying text is large. To address this issue, we propose an algorithm based on graph-based random walk that extracts only image-relevant tags from the accompanying text. We perform detailed evaluation of our scheme by checking its viability using human annotators as well as by comparing with state-of-the art algorithms. Experimental results show that the proposed algorithm outperforms base-line algorithms with respect to different metrics.

## 1 Introduction

A popular English idiom says "An image is worth a thousand words". Content writers always look out for good visual supplements to enrich their content and make it more appealing to the target audience. Fortunately, a huge repertoire of such content (images, video, etc.) is available in the Internet - however proper annotation with appropriate tags is necessary for their efficient retrieval. The size of online visual data clearly calls for an automatic approach to tag them.

**Table 1.** Example: An image of Apple co-founder Steve Jobs along with the text from an article using a similar image in InShorts[6], an on-line news aggregator.

**Apple sells its 1 billionth iPhone**
Apple on Wednesday announced that it sold its one billionth iPhone last week. The news comes about two years after the company sold the 500 millionth unit of its handheld device. The iPhone was first introduced in 2007 by late Co-founder Steve Jobs and had registered its one millionth sale after 74 days of the launch.

Existing tagging systems work towards capturing the denotational aspects of the image, viz. what the image denotes/contains. This includes tags capturing the various aspects present in the image. These details are either captured via the visual features of the images or via human added tags. However, the former tags are often generic and do not capture the entire information that is contained in the image. Let us consider an example in Table 1 which shows an image of the Apple co-founder, Steve Jobs from the web. Fig 1(a) shows the set of tags for the image based on the visual tagging system in [18]. It can be seen that the tags thus obtained are generic in nature e.g. 'person', 'business' and do not capture any deeper information about the image e.g. Steve Jobs, Apple Inc., etc. While an author uploading these images can be expected to add some of these tags, it is not possible to cover all aspects of the image.


(a) Visual Tags (b) Text based Tags (c) UBTer Tags

**Fig. 1.** Tags for the image in Table. 1 based on a visual tagger [18], textual parsing and our system - UBTer.

Often such images are used in different illustrations which contain valuable information about the image. To address the shortcomings of the visual tags, the accompanying content of the images can be analyzed to extract the tags. Such information can enhance both the denotational and connotational (how the image is perceived) understanding of the image. To test this hypothesis, we conducted a survey among 30 participants to rate the relevance of the text around an image in several articles on the web and its usefulness to enhance the understanding of the image. It was observed that in 91.23% of cases, the participants found the text relevant to the image. Survey respondents further opined that while the original image tags were very appropriate, the image had a different connotation when appeared along with the text, thus calling for a need to incorporate these into the image tags.

We identified an article (text included in Table. 1) from InShorts[6], an on-line news aggregator using an image similar to the one in Table. 1. A simple text based tagging can add a lot of noise to the tags as seen in Fig. 1(b), where the text in Table. 1 was parsed to extract the textual tags e.g. "days", "week", "billionth". These noise occur primarily because of the extract textual tags that are prominent in the text but irrelevant to the image's context. The level of noise will increase with the size of the accompanying content. This calls for an automated tagging system that optimally combines the tags from accompanying text with the image tags capturing the right denotational and connotational

[6] https://inshorts.com/news/apple-sells-its-1-billionth-iphone-1469693675991

information around the images while discarding the unrelated tags from the accompanying text. resulting tags.

In this work, we propose a novel framework **UBTer** - **U**sage **B**ased image **T**agg**er** that combines the tags derived from accompanying (usage) content with the image tags based on the visual features [18], thus integrating the information from content and usage cues. We thus achieve a balance between connotational and denotational aspects of an image. The resultant tags are shown in Fig. 1(c). We show that such a combination beats the state-of-the-art (visual and textual) tagging engines in our subjective and objective evaluations.

The paper is organized as follows. In Section 2, we describe the existing state of image tagging and position our framework with respect to existing systems. Section 3 introduces UBTer, - the proposed usage based tagger along with its key components. In Section 4 we compare the performance of UBTer against existing works via subjective and objective evaluations. We also evaluate the different parameters of UBTer to arrive at the right system configuration. Section 5 concludes the paper.

## 2 Related Work

Tagging and understanding textual content has been widely studied. The first step in textual tagging is extracting and detecting named entities; the popular one here is the Stanford NLP parser [11]. Once the named entities are identified, they are disambiguated and resolved into various categories [9]. Finally, the inter relationships in the content or hierarchies are identified by a semantic understanding of the text. In these works, the entities in the textual content are typically processed into a rich semantic representation (e.g. [1]) which is utilized to gain a deeper understanding of their inter-relationships.

Yang et al. [23] extract the textual tags based on a nearest-neighbor based approach and utilize the neighbors to extract the relationships between entities. Nallapatti et al. [13] use "event threading" to join different pieces of text and identify the undercurrent events in the textual topics. Shahaf et al. [16] estimate the importance and "jitteriness" of the entities in the text and use it to infer the connections between different parts of the textual content.

With the advent of knowledge bases like YAGO [19], relationships from these sources are used to further enhance the understanding of the textual content. Kuzey et al. [6] resolve temponym based on a YAGO based entity resolution to understand textual content with temporal scopes. They develop an Integer Linear Program that jointly optimizes the mappings to knowledge base for a rounded document representation. Tandon et al. [20] mine activity knowledge from Hollywood narratives to answer questions around these activities. They capture the spatio-temporal context of the topics by constructing multiple graphs to capture relationships among activity frames which is leveraged for effective understanding. However, none of these works aim at understanding images based on a combination of visual tags and usage context which is the key challenge in our problem, where we have to combine the content and usage cues in tagging.

There also exists a large body of literature in the space of image tagging. Li et al.[8] propose methods for assignment of tags from visual aspects and use

them for effective retrieval of images. Once an image is tagged, its relationships with other images have been used for further enhancing the tag set [14] or alternatively, using these tags to enhance tags of similar images [3]. The visual tags can also be enhanced and disambiguated with knowledge bases and conceptnets [22]. With the successful emergence of deep learning for image understanding, convolution neural networks have been used to find an intermediary representation *Visual Word2Vec* [5] in order to generate the image tags from this latent space. However, all these works focus on tagging the image from their visual cues/content. In our problem, we capture the usage of the images along with the visual content in the image tags to have a rounded understanding of the image.

One work that is close to the proposed solution framework is by Leong et al. [7], which relies exclusively on accompanying content for mining information relevant to the image. They construct relationships among entities based on multiple factors to arrive at the final set of tags. However, they do not use the visual tags of the images to align the accompanying content to the image and therefore have the same pitfall that we illustrated in our example in Table. 1.

## 3 UBTer - Usage Based Tagger

We propose a novel framework, **UBTer**, which enriches the tags around an image which may not be initially contained in the set of image based tags based on the visual features. UBTer takes as input the image tags (author given and the auto tags) along with the "usage" content which uses the image for illustration. The content is processed to extract key tag candidates. Many of these tags may not be directly related with the image and hence needs to be pruned. The pruning is initiated by establishing the context of a tag. This is done by a). scoring the importance of the tag by measuring its usage pattern in the local textual context and b). capturing the inter-tag relationship based on certain global knowledge base. Thus we obtain a graph with weighted nodes (local importance) and weighted edges. The final tags are selected by performing a biased (based on node weight and edge weight) random walk starting from the image tags. Those nodes reached by random walk are selected in the final set. They are found not only rich and appropriate but also diverse bringing out various connotational aspects of the same image.

### 3.1 Tag Shortlisting

The input to UBTer is the image along with its visual tags and the accompanying text(s). The accompanying text might contain several entities that could be ambiguous e.g. Apple, Jobs in Table 1. The algorithm therefore starts with disambiguating the accompanying content for such ambiguous entities via Ambiverse [4]. Ambiverse provide a technology to automatically analyze a textual data and disambiguate named entities. It relies on the knowledge base YAGO for an accurate characterization of all the entities in the text. These entity characteristics are used along with the context of the entity in the text to disambiguate them into formal YAGO entries. We replace each occurrence of the entity with their disambiguated version. The disambiguated content is extensively parsed

to identify all named entities and noun phrases using the Stanford NLP Parser [11]. Note that the image may/may not be relevant to the entirety of the entities in the accompanying text and we address this in Section 3.3. At the end of this step, we have a set of all candidate tags for consideration in the final tags.

### 3.2 Tag Importance

For each tag candidate, a score is assigned based on their importance in the local context. We calculate the total frequency of the candidate tag occurrence in the usage content accounting for the co-reference of the candidates via proper nouns by co-reference parsing. Thus, not just the direct mentions, the indirect mentions of the entities are also accounted in their local importance. We normalize the frquency counts by the counts of all entities in the text to keep the measure between 0 and 1.

For every tag candidate we also compute the average distance of the entity from the root of the corresponding dependency tree (obtained by passing the accompanying content through a dependency parser[2]). A candidate tag at the root (distance = 1) is the central topic of discussion in a sentence and hence is more important indicating the local relevance of the entity in the discussed subject. The inverse distance is considered as the tag importance (tags at the root gets a value of one).

The average of the two measures yields the final tag importance ($n_i$) whereby the tags that are in the center of discussion in the accompanying content getting higher value. We assume that a picture is added to further emphasize the central point of discussion.

### 3.3 Inter-tag Relationship

We build the relationships between each tag candidates leveraging two independent global knowledge base. (A). We used the Word2Vec [12] model trained on a corpus of Google News dataset with 100 billion words resulting in a final corpus of about 3 million word representations. Word2Vec yields a 300 dimensional vector for every tag candidate that represents the word in the space of the trained deep neural network. We compute the cosine-similarity between the vectors in this space which captures the semantic closeness between the tags. (B). We calculate the point-wise mutual information [21] between two entities based on their co-occurrences in the Wikipedia articles. This yields a similarity score based on how coherent the two tags are with respect to the entire Wikipedia corpus (English).

The Word2Vec based measure captures the semantic similarity between the tags because the Word2Vec space groups similarly meaning entities together. Therefore, entities closer in this space can often be interchangeably used in several context. On the other hand, the Wikipedia based measure captures the topical closeness - since entities that occur together in the several articles are closer in this space. Our final edge weight ($e_{ij}$) is the average of the two measures.

The edge weights along with the node importance yield a graphical representation of the candidate tags with the edge weights capturing the global relationship between the tags and the node weights indicating their local importance in the usage content.

**Infusing Image Tags:** To extract the usage-specific tags from the accompanying content, it is important to understand how these tag candidates relate to the visual tags. However, there may be duplicates or near duplicates to the visual tags already present within the tag set. Therefore, we first calculate the edge weight between the visual tags and every tag candidate in the graph based on the combined measure above. The tag pairs with similarity greater than a threshold (0.95 in our experiment) are merged into a single node, thus avoiding duplicity in tags. We then propagate the importance of the merged node to the adjacent nodes (at a distance of 2 edges) using an exponential decay. This ensures the propagation of the strength of the merged nodes to its neighbors and thus emphasizing the relevant pieces of the tag graphs with respect to the visual tags.

For tag pairs less than the matching threshold, an edge is added between every tag candidate whose similarity with the visual tag is significant ($> 0.1$ in our experiments). This ensures that the visual tags are connected to the relevant parts of the tag-graph. The series of steps is summarized in Algorithm 1.

**Algorithm 1** Tag Unifier

```
1: procedure UNIFY(tagsFromImage, TagGraph)
2:     for tag ∈ tagsFromImage do
3:         tag ← normalize(tag)
4:         for node ∈ TagGraph do
5:             val ← similarity(tag,node) (from Sec.3.3)
6:             if val > σ1 then
7:                 MergeNodes(tag,node)
8:                 node.weight ← MergedWeight()
9:                 PropagateWeight(node)
10:            else if val > σ2 then
11:                edge ← createNewEdge(tag,node)
12:                edge.weight ← val
13:            else
14:                continue
15:            end if
16:        end for
17:    end for
18: end procedure
```

### 3.4 Tag Extraction

With the graphical representation of the tags, the problem of extracting the tags that capture the context around the image boils down to identifying the top nodes in the tag graph that are closely connected to the image tags. For this we use a random walk based algorithm [15], starting the random walk from the visual tags, thus ensuring the node ranking relevant to the tag images and avoiding irrelevant tags from the accompanying text.

We define the probability of the random walk moving from a node $i$ to another node $j$ as, $P(tr_{i \to j}) = e_{ij} \times n_j$ where, $e_{ij}$ is the weight of the edge (from

Section 3.3) between tags $i$ and $j$ and $n_j$ is the node importance of tag $j$ from Section 3.2. The probability of the node staying in the same node is defined as $P(tr_{i \to i}) = n_i$. The probabilities are normalized to conform to the requirements of a probability distribution. The final set of tags is then extracted by performing a random walk over several iteration starting from the visual/author tag nodes. This ensures that the tags selected are not just based on their importance from the accompanying text but also emphasizes on a strong relationship with the visual tags. The random walk is terminated after $k$ (20 in our experiments) iterations and the average number of visits to a node across all runs is used as the score of the tags. The top-$k$ tags is output as the final set of tags for the images.

## 4 Experimental Evaluation

We first evaluate the importance of usage tags from UBTer based on an annotator based evaluation. We then introduce 3 independent metrics that measure different aspects of the extracted tags and use them to extensively test the performance of UBTer against existing tagging baselines on the dataset from [7]. Finally, we evaluate the different parts of the UBTer to measure their significance in extracting the final tags.

### 4.1 Importance of Usage Tags

In order to assess the importance of usage tags over the visual tags, we conducted a survey among 45 participants to rate the overall relevance and diversity of the tags on a scale of $0-10$ for the outputs from UBTer as well as those provided by the visual tagger [18] on a subset of 20 images. On a scale of 10 for tag relevance to the image, usage tags were rated at $8.08 \pm 0.58$ on an average against the score of $5.73 \pm 1.19$ for the visual tags. For diversity, usage tags received a rating of $6.79 \pm 0.8$, whereas, the visual tags received $5.93 \pm 0.73$. This indicates that UBTer increases the overall relevance of the tags to the image and also performs better in terms of the diversity of the tags indicating the viability of UBTer.

### 4.2 Ground Truth Data Set

We utilized the dataset curated by Leong et al. [7] which contains 300 image-text pairs collected by issuing a query to Google Image API and processing one of the query results that has a significant amount of text around the images. Leong et al. [7] have also created a gold standard tag set based on manual annotations from 5 annotators via Amazon Mechanical Turk accepting annotations from annotators with approval rating $> 98\%$. The annotators have suggested the tags about the image based on their understanding of the accompanying text. We used the Clarifai API [18] to generate the visual tags for all our experiments.

### 4.3 Metrics for evaluation

Human annotations cannot be extended for a comprehensive evaluation of the tags. We therefore extend several existing metrics to measure different aspects of the tags which are described below.

The **term-significance [10]** is calculated as the significance of the tags to the textual content and is calculated by computing the Normalized Discounted

Cumulative Gain(NDCG) over the term frequency of the tags from the usage content normalized based on the tag's inverse document frequency in a global corpus. The intuition here is to compute how important a tag is to the given context (usage) and normalize it with its "commonness" across a bigger corpus (as computed by the idf). We use Wikipedia as the bigger corpus similar to Leong et al. [7].

The term-significance metric purely tests the relevance of the tags to the usage content. To further capture the **tag relevance** of the tags to the gold standard tags and its overall **diversity**, we propose two additional metrics. To determine how relevant our tags are to the gold standard tags, we compute a weighted cosine similarity between the Word2Vec [12] representation of the extracted tags and the gold tags as given by,

$$sim = \frac{1}{N} \sum_{i} \frac{\sum_{a_j \in TopK(G_i, I_i)} cos(a_j, I_i)\gamma^j}{\sum_j \gamma^j}, \tag{1}$$

where $N$ is the number of tags generated for the images, $I_i$ is the vector representation of the $i^{th}$ image tag and $G_i$ is the set of all vector representations of the gold standard tags. The inner sum above computes a weighted average of the similarity between the generated tag and the most similar gold-standard tags. An average of the similarity can lead to higher relevance only when the tag is relevant to all human annotated tags. Alternatively, a max over the similarities can lead to high scores for tags even if they are similar to a single human tag. The parameter $\gamma$, $(0 \leq \gamma \leq 1)$ addresses both these scenarios via a similarity-ranked-decayed-weighted-average.The outer summation averages this measure between all the generated image tags and the gold standard tags.

Finally, for measuring the diversity in the tags, we use the **cophenet correlation coefficient** [17] (which is a measure of how faithfully a dendogram preserves the pairwise distances between the original un-modeled data points). We perform a hierarchical clustering on the tags based on their Word2Vec representation and compute the cophenet correlation coefficient as the diversity score. Cophenet correlation coefficient is then given by,

$$c = \frac{\sum_{i<j}(x(i,j) - \bar{x})(t(i,j) - \bar{t})}{\sqrt{[\sum_{i<j}(x(i,j) - \bar{x})^2][\sum_{i<j}(t(i,j) - \bar{t})^2]}} \tag{2}$$

where, $x(i,j)$ is the distance between the $i^{th}$ and $j^{th}$ tag. $t(i,j)$ is the height of the node at which the clusters corresponding to $i^{th}$ and $j^{th}$ clusters are first joined together. A higher value of the cophenet correlation coefficient indicates the presence of more significant clusters and hence more tag diversity.

### 4.4 Tagging Performance

To evaluate the proposed UBTer based tags, we compare it against the baseline algorithm in [7]. Leong et. al [7] propose 3 independent algorithms based on "Wikipedia Salience", "Flickr Picturability" and "Topic Modeling" to extract

tags for an image from its accompanying textual content. In their experiments, the Wikipedia Salience based tagger was best performing in terms of the precision and recall. We used this algorithm as the baseline for our evaluations. We also compare the performance of our UBTer against the visual tagger in [18]. Fig. 2 shows the Term Significance, Tag Relevance and Tag Diversity for the tags from [18], [7] and UBTer.

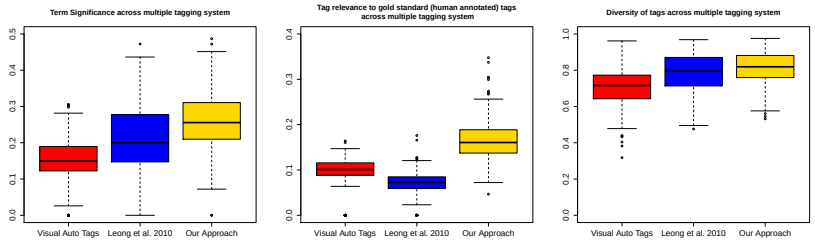



**Fig. 2.** Term Significance, Tag Relevance (Eq. 1) and Diversity (Eq. 2) for tags based on Clarifai [18], Wikipedia Salience [7] and UBTer

The term significance checks significance of the tags with respect to the accompanying text and hence text based taggers are expected to perform better in this measure. Along the expected lines, both the UBTer and the tagger by Leong et al. [7] perform better than the visual tagger. Between the text based taggers, the term significance is the best for UBTer indicating the superiority of the tags in capturing the local context.

The tags from UBTer are also more relevant/close to the human annotated tags based on the tag relevance (Eq. 1). A superior performance here indicate that UBTer captures the denotational aspects as well as the connotational aspects.

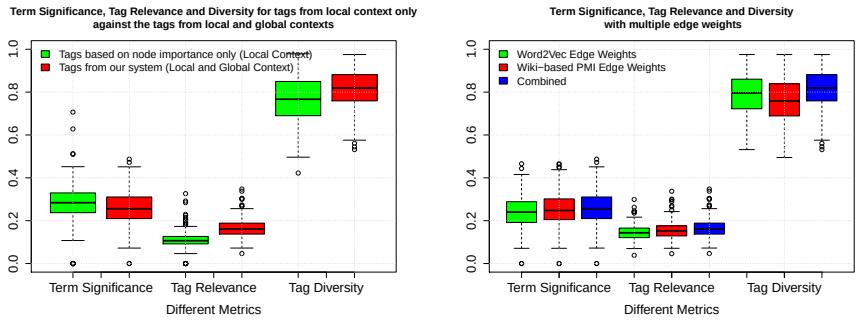
Capturing the connotational aspects of the images yields more diversity as indicated by the superior performances of both the text-based taggers on the scales of diversity. Here again, the tags from UBTer are marginally more diverse than the tags from Leong et al. [7].

### 4.5 Evaluation of Algorithmic Parameters

Finally, we independently evaluate the different parts of UBTer and their importance in extracting relevant and diverse tags capturing the image usage.

**Local vs Global Context:** In this experiment, we compare the local context captured by the node importance (Sec. 3.2) against the combined context captured in UBTer. We extract the top tags based on their node importance score and compare it against the UBTer tags.

Fig. 3(a) compares the Term Significance, Tag Relevance (Eq. 1) and Tag Diversity for the two cases. The term significance of the tags based on the local context with an average of 0.275 is marginally better than the term significance of UBTer (average at 0.26). Since the term significance captures the local importance of the tags in the accompanying text, hence the tags from local context is




(a) Local vs Global Context (b) Effect of Edge Weights

**Fig. 3.** Term Significance, Tag Relevance (Eq. 1) and Diversity (Eq. 2) for tags for different algorithmic parameters. Fig. 3(a) compares the tags extracted solely based on Node Importance against the tags from UBTer (where the local and global context of the tags are jointly accounted for). Fig. 3(b) compares the effects of different edge weighting mechanisms on the tagging performance

expected to be better here. However, the overall tag relevance (average of 0.1090 for local context against 0.1654 for the combined context) and tag diversity (average of 0.7554 for local context against 0.0.8155 for the combined context) is better with the combined approach since it accounts for the global relationship between the tags as well as similarity of connotational tags with visual tags. Hence better tags without compromising much on the term significance (since the difference between the two methods is not significant) is derived.

**Effect of edge weights:** In the next experiment, we compare the term significance, tag relevance and tag diversity among the edge weighting mechanisms based on Word2Vec, Wikipedia and the combined metric defined in Section 3.3.

From Fig. 3(b), it can be seen that while Word2Vec performs marginally better than the Wikipedia based relationship on the scales of term significance (average of 0.2516 for Word2Vec based metric against the 0.2398 average for the Wikipedia based metric) and tag relevance (average of 0.1567 for Word2Vec based metric against the 0.1451 average for the Wikipedia based metric). In terms of overall tag diversity, Wikipedia based metric is marginally better than Word2Vec (average of 0.7897 for Wikipedia based metric against the 0.7617 average for the Word2Vec based metric). This could perhaps be because Wikipedia includes more entities than the Google News Corpus on which the Word2Vec were trained, and hence aid in the extraction of diverse tags. Note that the combined approach yields the best tags across all metrics.

**Effect of visual tag quality:** We finally compare the correlation between the quality of the visual tags and the tags from UBTer.

Fig. 4 shows the correlation between the two sets of tags on the scales of Term Significance, Tag Relevance and Tag Diversity. It can be seen that there is a strong dependence of the term significance and relevance of UBTer tags with the visual tags as indicated by the slopes of 0.95 and 0.89 respectively of the corresponding line fits. This is expected since the algorithm starts the random walk from the visual tags and hence the output tag quality is directly dependent

on the quality of visual tags. However the tag diversity is less dependent on the visual tags, since the diversity of the output tags is obtained more from the accompanying text than from the visual tags indicated by a lower slope of the corresponding line (0.36).

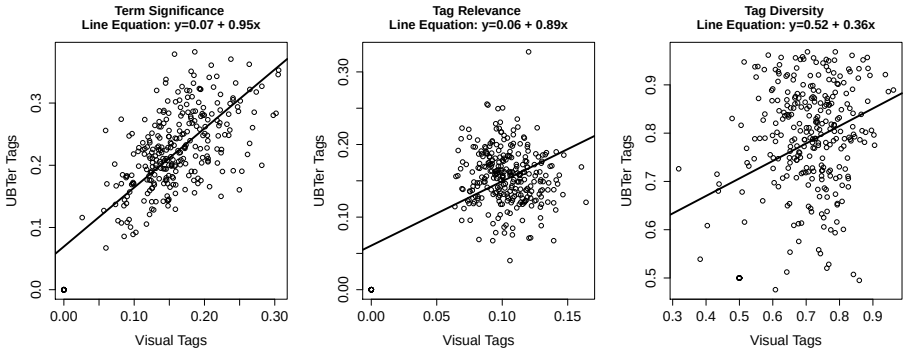



**Fig. 4.** Correlation between the quality of visual tags and the tags from UBTer

## 5 Conclusion

In this paper, we have proposed a novel system - UBTer to enhance the tags of an image by capturing its usage. Capturing usage through tags is not straightforward as majority of the tags describing the neighboring text of an image don't pertain to the image - our approach gleans out the relevant tags. This is done first through understanding the importance of the tag in local context (we conduct sophisticated dependency test to compute the importance) and then derive the inter-tag relationship (we use Word2Vec and Wikipedia-co-occurrence) and finally run a biased random walk to shortlist relevant tags. The tags thus obtained outperform the state-of-the art systems in the lights of several quality metrics capturing the relevance and diversity of the tags. Such a tagging system will serve well to improve the image retrieval and recommendation systems by effectively expressing the user's context.

## References


1. Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N.: Abstract meaning representation (amr) 1.0 specification. In: Conference on Empirical Methods in Natural Language Processing. ACL (2012)
2. Chen, D., Manning, C.D.: A fast and accurate dependency parser using neural networks. In: Conference on Empirical Methods in Natural Language Processing. ACL (2014)
3. Guillaumin, M., Mensink, T., Verbeek, J., Schmid, C.: Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In: IEEE International Conference on Computer Vision (2009)
4. Hoffart, J., Yosef, M.A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S., Weikum, G.: Robust disambiguation of named entities in text. In: Conference on Empirical Methods in Natural Language Processing. ACL (2011)

5. Kottur, S., Vedantam, R., Moura, J.M., Parikh, D.: Visual word2vec (vis-w2v): Learning visually grounded word embeddings using abstract scenes. arXiv preprint arXiv:1511.07067 (2015)
6. Kuzey, E., Setty, V., Strötgen, J., Weikum, G.: As time goes by: comprehensive tagging of textual phrases with temporal scopes. In: International Conference on World Wide Web. ACM (2016)
7. Leong, C.W., Mihalcea, R., Hassan, S.: Text mining for automatic image tagging. In: International Conference on Computational Linguistics. ACL (2010)
8. Li, X., Uricchio, T., Ballan, L., Bertini, M., Snoek, C.G., Del Bimbo, A.: Image tag assignment, refinement and retrieval. In: ACM International Conference on Multimedia (2015)
9. Lieberman, M.D., Samet, H.: Adaptive context features for toponym resolution in streaming news. In: ACM SIGIR conference on Research and Development in Information Retrieval. ACM (2012)
10. Lu, Y.T., Yu, S.I., Chang, T.C., Hsu, J.Y.j.: A content-based method to enhance tag recommendation. In: IJCAI (2009)
11. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The stanford corenlp natural language processing toolkit. In: ACL (System Demonstrations) (2014)
12. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems (2013)
13. Nallapati, R., Feng, A., Peng, F., Allan, J.: Event threading within news topics. In: ACM International Conference on Information and Knowledge Management. ACM (2004)
14. Ramanathan, V., Li, C., Deng, J., Han, W., Li, Z., Gu, K., Song, Y., Bengio, S., Rossenberg, C., Fei-Fei, L.: Learning semantic relationships for better action retrieval in images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
15. Sarkar, P., Moore, A.W.: Random walks in social networks and their applications: a survey. In: Social Network Data Analytics. Springer (2011)
16. Shahaf, D., Guestrin, C.: Connecting the dots between news articles. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (2010)
17. Sokal, R.R., Rohlf, F.J.: The comparison of dendrograms by objective methods. Taxon pp. 33–40 (1962)
18. Sood, G.: clarifai: R Client for the Clarifai API (2016), r package version 0.4.0
19. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: International Conference on World Wide Web. ACM (2007)
20. Tandon, N., de Melo, G., De, A., Weikum, G.: Knowlywood: Mining activity knowledge from hollywood narratives. In: International Conference on Information and Knowledge Management. ACM (2015)
21. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: 40th Annual Meeting on Association for Computational Linguistics (2002)
22. Xie, L., He, X.: Picture tags and world knowledge: learning tag relations from visual semantic sources. In: ACM International Conference on Multimedia (2013)
23. Yang, Y., Ault, T., Pierce, T., Lattimer, C.W.: Improving text categorization methods for event tracking. In: ACM SIGIR Conference on Research and Development in Information Retrieval. ACM (2000)