Relevance of Tags to Images: A Brief Survey of Major Parameters and Evaluation Measures

Fouzia Jabeen, Shah Khusro, Saeed Mehfooz, Azhar Rauf, Jehangir Khan

Department of Computer Science, University of Peshawar, Peshawar 25000, Pakistan
Fouzia.jabeenphd@upesh.edu.pk

Abstract: An exciting way of finding or searching for images is by using tags. Tags are user given words to an image. These tags serve the purpose of metadata. The freedom of giving tags in any form raises questions that whether tags are relevant to image or not. The correct tag assignment will facilitate in search and navigation not only for the users that have contributed the image but also for other users interested in same. In contrast irrelevant tags contribute to semantic noise. In this paper we give brief survey of the major parameters in the literature that are used for estimation of relevance of tags with images along with evaluation measures.
[Jabeen F, Khusro S, Mehfooz S, Rauf A. **Relevance of Tags to Images: A Brief Survey of Major Parameters and Evaluation Measures**. *Life Sci J* 2013;10(12s):401-404] (ISSN:1097-8135). http://www.lifesciencesite.com. 68

Keywords: Folksonomy tags; tag refinement; relevant tags; evaluating parameters

1. Introduction

Folksonomy applications like Flickr gives freedom to its users to freely assign tags to images. These assigned tags are used for retrieval of these images later on. In addition these tags can also be used to find out the related relevant images. As users are free to assign tags, many erroneous tags can also be assigned. These erroneous tags are called spam tags or noisy tags. The noisy tags would result in problems in retrieving images accurately. There are many motivations behind giving noisy or misleading tags. However these tags are irrelevant. Figure 1 shows the relevant and irrelevant tags associated with an image.

More precisely the appropriateness of tags can be defined by three points. Which are (1) the tags must appropriately describe the contents of the image (2) Other members of the community easily understand these tags and (3) Make it easy to find the images.

This is a survey paper that is briefly focused on two objectives. Firstly, to identify the major parameters used by the researchers to find relevant set of tags for images. Secondly, c the major evaluation metrics used in the liter test the algorithms developed for the solutior problem.

The paper is organized in four s Section 2 identifies and explains the relevant parameters used in the literature. Section 3 is about the overview of the evaluation metrics used to measure the accuracy of the algorithms developed for finding out the relevant and appropriate tags for images. Lastly we end up with discussion in Section 4.

2.0 Relevance Estimation Parameters

In this section we have identified and discussed the major parameters in the literature to find appropriate and relevant tags for images. We have classified these parameters as following.

a) Image Dependent: Parameters that consider image for finding appropriate Tags.
b) Image Independent: Parameters that do not consider image for finding the appropriate tags.

2.1 Image Dependent Parameters

2.1.1 Visual Similarity/Dissimilarity of Images

This parameter considers similar images or dissimilar images as a basis for selection of appropriate tags.

Based on logic that if different individuals annotate visually similar images using the same tags, these tags are likely to describe same objective aspects of the visual content. Li et al (Li, G.M.Snoek, and Worring, 2009) neighbor voting algorithm accurately and efficiently learns tag relevance by gathering votes from visual neighbors.

However Lee et al (Lee, Neve, and Ro, 2012) argued that neighbor voting algorithm only make use of similar images. So the author proposed the technique based on both visually similar and dissimilar images.

2.1.2 Semantic Similarity between Tags

The strong relation between tags is considered to decide the appropriateness.

Liu et al (Liu, Wang, Yang, Hua, and Zhang, 2009) based their scheme of improvement of tag quality assigned to image by removing noisy tags and suggestion of additional correct tags. The proposed method improves tag quality by exploiting both semantic and visual similarity. Semantic

similarity is estimated by calculating the similarity between their associated tag sets.

For same objective Lee et al (Lee, Neve, and Ro, 2010) based their scheme of detection of noisy tag on image-tag and tag-tag relation measurement. For calculating the image- tag relation visual similarity of images is calculated and for tag-tag relation co-occurrence statistics is considered. The decision whether tag is noisy or not is based on the probability of both that is image-tag and tag-tag similarity results. If the probability is less than a given threshold, the corresponding tag assignment is assumed to be in category of noisy tag. Otherwise, the tag assignment in question is found to be correct. Almost 36% of noisy tags has been reduced at the cost of 10% correct tag assignment.

Lee *et al* (Lee et al., 2010) proposed removal of semantic noise means by utilizing two semantic measurements. First is number of times tag occur (tag frequency) and second is tag co-occurrence. Almost 40% of noisy tags are discovered.

Figure 1. Relevant and Irrelevant Tags[1]

2.1.3 Representing Interesting Portions of the Image

This parameter considers all those tags assigned to image appropriate and relevant if they represent interesting or important portions of an image. Levy *et a*l (Levy et al., 2009) tested this parameter on music folksonomy.

2.1.4 Representing certain event

It might be the case that tags assigned to an image do not represent its contents but still are relevant. For example if there is an image of a match played in Olympics and tag "Olympics" is assigned. Then this tag does not represent contents that image has but represents the event in which this image was taken. In other words we can use the term semantic context for such tags. We can say these tags represents the semantic context of an image.

[1] http://www.flickr.com

2.2 Image Independent Parameters

2.2.1 Wisdom of Crowd

Agreement of users on tags that are assigned to an image, in other words wisdom of crowd is one of the important parameter considered for appropriateness of tag assigned to images. In other words popularity of tag(s) for an image. The drawback of this parameter is that to increase the view of certain image, spam users can assign tags that are popular at particular time irrespective of their appropriateness. So that any user that do search with that tag would find that image .

Based on this parameter Tang *et al* (Tang, Zuo, Xu, Zheng, and Wang, 2010) proposed to use this wisdom of crowd along with the semantic similarity between tags to access the appropriateness.

To end this section, Table 1 outlines the major parameters and Table 2 gives review of the parameters utilized by the researchers for finding tag appropriateness with images.

Table 1. Major Relevance Parameters

No	Parameter
1	Visual Similarity/Dissimilarity of Images
2	Semantic Similarity between Tags
3	Representing Interesting Portions of the Image
4	Representing certain event
5	Wisdom of Crowd

Table 2. Parameter used in the Literature

Paper Id	Parameter Used
Liu et al(Liu et al., 2009) Lee et al (Lee et al., 2010)	1,2
Lee et al (Lee et al., 2012)	1
Tang et al(Tang et al., 2010)	2,5
Levy et al(Levy et al., 2009)	3

3.0. Evaluation Criteria

In this section we give review of major evaluation metrics. These metrics are used for testing the results of algorithms developed for finding out the relevant tags and removal of noisy or irrelevant tags.

3.1 Confusion Matrix

Confusion matrix basically gives the counting of following four items.

1) Number of relevant tags correctly detected by the algorithm (w)
2) Number of relevant tags not correctly detected by algorithm (x)

3) Number of noisy tags detected as correct tags(y)
4) Number of noisy tags detected as noisy tags(z)

3.2 Precision, Recall and F measure

Based on the results of confusion matrix precision, recall and FMeasure is calculated using formulas given in Eq 1 ,Eq 2 and Eq3.

$$R = \frac{z}{y+z} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots Eq(1)$$

$$P = \frac{z}{x+z} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \text{Eq (2)}$$

$$FMeasure = \frac{2PR}{P+R} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots Eq(3)$$

3.3 Refinement Gain Measure

Basically this measure calculates ratio between removed corrected and removed noisy tags using Eq 4 and Eq 5 respectively**.** Where the values of variables w,x,y and z are taken from confusion matrix discussed above.

$$\text{RemovedNoisyTagRatio} = \frac{z}{y+z} \dots\dots \text{Eq(4)}$$

$$\text{RemovedCorrectTagRatio} = \frac{x}{w+x} \dots\dots \text{Eq(5)}$$

$$\text{Refinement Gain} = \text{RemovedNoisyTagRatio} - \text{RemovedCorrectTagRatio} \dots\dots \text{Eq(6)}$$

3.4 Normalized Discounted Cumulative Gain

This measure ranks the relevance of a tag in levels. To be precise most relevant tag ranked by assigning level 5, only relevant tag is at level 4, partially relevant tag is at level 3, weakly relevant tag is at level 2 and irrelevant tag is at level 1 (Lee et al., 2010)(Järvelin and Kekäläinen, 2002).

3.5 Noise Level Metric (NL)

This metric basically represents the ratio of irrelevant tags in the set of all user assigned tags. If there is high no of irrelevant tags then value of NL is close to 1,and if there are few number of irrelevant tags then NL is closed to Zero. NL is calculated using Eq7 (Chua et al., 2009).

$$\text{NL} = 1 - \text{FMeasure} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \text{Eq(7)}$$

3.6 Precision at rank m (P @ m)

Precision at rank m where m is a variable and it takes values 1,2…….m. P@ 5 for example means proportion of recommended tags that are relevant to the image content when recommending five tags (Lee et al., 2010)(Lee et al., 2012).

To end this section, Table 3 represents the list of major evaluation metrics that are used to test the techniques developed for finding relevant tag set for images.

Table 3: Major Evaluation Metrics

No	Evaluation Metric	Paper ID
1	Confusion Matrix	(Castillo, Donato, and Murdock, 2007)
2	Precision , Recall and F Measure	(Castillo et al., 2007)(Chua et al., 2009)
3	Refinement Gain	(Lee et al., 2010)
4	Normalized discounted cumulative gain (nDCG)	(Järvelin and Kekäläinen, 2002)
5	Noise Level	(Chua et al., 2009)
6	P@1, P@ 2…….P@K	(Sigurbjörnsson and van Zwol, 2008)(Lee, Neve, and Ro, 2012)

4. Discussions

Finding appropriate and relevant set of tags for images is an important problem that need attention of the researchers. Because association of irrelevant tags with the images cause problem in search and retrieval and also in finding the relevant related images.

The major cause of this problem is the freedom that users have for free assignment of tags to images. In this survey paper we have briefly covered two aspects of this problem. Firstly, we have identified the major parameter that are considered by the researchers for finding the relevant and appropriate set of tags. Secondly we have identified the major evaluation metrics. These metrics test the worthiness of the results of algorithm developed for solving the problem of finding appropriate and relevant tag set for images.

Corresponding Author:
Fouzia Jabeen
Department of Computer Science
University of Peshawar, 25000, Pakistan
E-mail: Fouzia.jabeen@gmail.com

References

1. Castillo, C., Donato, D., & Murdock, V. (2007). Know your Neighbors : Web Spam Detection using the Web Topology, 423–430.
2. Chua, T., Tang, J., Hong, R., Li, H., Luo, Z., & Zheng, Y. (2009). NUS-WIDE : A Real-World Web Image Database from National University of Singapore, 0–8.
3. Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems. 20*(4), 422–446. Retrieved from http://dl.acm.org/citation.cfm?id=582418
4. Lee, S., Neve, W. De, & Ro, Y. M. (2010). Tag refinement in an image folksonomy using visual similarity and tag co-occurrence statistics. *Image Communication*, *25*(10), 761–773. Retrieved from http://dl.acm.org/citation.cfm?id=1891010
5. Lee, S., Neve, W. De, & Ro, Y. M. (2012a). Towards Data-driven Estimation of image Tag Relavence using Visually Similar and Dissimilar Folksonomy Images. In *Proceedings of the 2012 international workshop on Socially-aware multimedia* (pp. 3–8). Retrieved from http://dl.acm.org/citation.cfm?id=2390880
6. Lee, S., Neve, W. De, & Ro, Y. M. (2012b). Towards Data-driven Estimation of Image Tag Relevance using Visually Similar and Dissimilar Folksonomy Images. In *Proceedings of the 2012 Workshop on Socially-Aware Multimedia (SAM 2012; held in conjunction with ACM Multimedia 2012)* (pp. 3–8). Retrieved from http://www.citeulike.org/user/wmdeneve/article/11007815
7. Levy, M., Last.fm, London, & Sandler, M. (2009). Music Information Retrieval Using Social Tags and Audio. *IEEE Transactions on Multimedia*, *11*(3), 383–395.
8. Li, X., G.M.Snoek, C., & Worring, M. (2009). Learning Social Tag Relevance by Neighbor Voting. In *IEEE Transactions on multimedia*.
9. Liu, D., Wang, M., Yang, L., Hua, X., & Zhang, H. (2009). TAG QUALITY IMPROVEMENT FOR SOCIAL IMAGES. In *Proceedings of international Conference on Multimedia and Expo* (pp. 350–353).
10. Sigurbjörnsson, B., & van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge. In *Proceeding of the 17th international conference on World Wide Web - WWW '08* (p. 327). New York, New York, USA: ACM Press. doi:10.1145/1367497.1367542
11. Tang, R., Zuo, J., Xu, K., Zheng, J., & Wang, Y. (2010). An Intelligent Semantic-based Tag Cleaner for Folksonomies. In *IEEE* (pp. 773–776).

11/12/2013