USER STUDY OF THE ASSIGNMENT OF OBJECTIVE AND SUBJECTIVE TYPE TAGS TO IMAGES IN INTERNET

Evaluation for Native and non Native English Language Taggers

David Nettleton
Department of Information Technology and Communications, Pompeu Fabra University
Tanger, 122-140, 08018 Barcelona, Spain

Mari-Carmen Marcos
Department of Journalism and Audiovisual Communication, Pompeu Fabra University
Roc Boronat,138, 08018 Barcelona, Spain

Bartolomé Mesa-Lao
Department of Translation and Interpreting, Autonomous University of Barcelona
Edifici K – Campus UAB, 08193 Barcelona, Spain

Keywords: Image tagging, Tag recommendation, User support, Statistical analysis, Data modeling.

Abstract: Image tagging in Internet is becoming a crucial aspect in the search activity of many users all over the world, as online content evolves from being mainly text based, to being multi-media based (text, images, sound, …). In this paper we present a study carried out for native and non native English language taggers, with the objective of providing user support depending on the detected language skills and characteristics of the user. In order to do this, we analyze the differences between how users tag objectively (using what we call 'see' type tags) and subjectively (by what we call 'evoke' type tags). We study the data using bivariate correlation, visual inspection and rule induction. We find that the objective/subjective factors are discriminative for native/non native users and can be used to create a data model. This information can be utilized to help and support the user during the tagging process.

1 INTRODUCTION

The ability to share multimedia information on the Social Web has created the need to describe all of this information. Nowadays, users uploading information to the web have the possibility to tag (ie, describe) content using keywords.

Among the possible limitations of tags created by users, one could mention inconsistency among users and typos, but there are also other factors limiting the quality of these tags: the level of linguistic competence in the language used by the tagger. It seems reasonable that native language users will tag in a more accurate and diverse way than non native users.

Currently, English is the most common language used on the Internet and many users describe images, video and music in English even though it is not their native language. For those users tagging in a non-native language, it could be very useful to have a system which can suggest tags already used by other users, so they can have access to similar content descriptions.

The main aim of this study is to discover how a group of English native and non-native users tag images on the Internet. To do so we have shown them ten pictures and we asked them to describe them both in an objective way (what do you actually see in this picture?) and in a subjective way (which feelings are aroused by this picture?).

Our hypothesis assumes that: **(i)** Tags created by

Nettleton D., Marcos M. and Mesa-Lao B. (2009).
USER STUDY OF THE ASSIGNMENT OF OBJECTIVE AND SUBJECTIVE TYPE TAGS TO IMAGES IN INTERNET - Evaluation for Native and non Native English Language Taggers.
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, pages 99-104
DOI: 10.5220/0002273900990104

native speakers will be of higher quality (quality is defined here in terms of quantity and variety of tags used, once errors have been eliminated); **(ii)** The quality of those tags created by native speakers will become more apparent when they have to describe feelings evoked by the picture rather than when they objectively describe what is seen in the picture.

If this assumption is valid, we will have objective data to design a recommendation system in which tags would be automatically proposed to users based on previous tagging sessions. These previous sessions would only be selected from users providing high quality tags (i.e. good tags in terms of quantity and variety). This recommendation system would help non-native taggers to work with tags used by native taggers.

Goals and main contributions: to the best of our knowledge there are non or few investigators working on support for non-native taggers of images, and making the distinction and support for subjective versus objective tagging, which are two of the main lines of our work presented in this paper.

2 STATE OF THE ART AND RELATED WORK

We ask up to what point users with different language skill levels vary in their way of indexing contents which are similar or the same. Specifically, we will look at the description of images, and the difference between tags which represent feelings, emotions or sensations compared with tags which represent objective descriptions of the images (Boehner, DePaula, Dourish, Sengers, 2007)(Isbister, Hook, 2007).

In recent years tag recommendation has become a popular area of applied research, and of commercial interest for the major search engine and content providers (Yahoo, Google, Microsoft, AOL…). Different approaches have been made to tag recommendation, such as that based on collective knowledge (Sigurbjörnsson, van Zwol, 2008), approaches based on analysis of the images themselves (when the tags refer to images) (Anderson, Raghunathan, Vogel, 2008), collaborative approaches (Lee, 2007), a classic IR approach by analyzing folksonomies (Lipczak, Angelova, Milios, 2008), and systems based on personalization (Garg, Weber, 2008). With respect to considerations of non-native users, we can cite works such as (Sood, Hammond, Owsley, Birnbaum, 2007). Finally we can cite approaches based on complex statistical models, such as (Song, 2008).

3 METHODOLOGY – DESIGN OF EXPERIMENTS FOR USER EVALUATION

For this study we have selected 10 photographs from Flickr. The photographs we have used have been chosen for their contrasting images and for their potential to require different tags for ‘see’ and ‘evoke’. Image 1 is of a person with his hands to his face; Image 2 is of a man and a woman caressing; Image 3 is of a small spider in the middle of a web; Image 4 is of a group of people dancing in a circle with a sunset in the background; Image 5 is of a lady holding a baby in her arms; Image 6 is of a boy holding a gun ; Image 7 is of an old tree in the desert, bent over by the wind; Image 8 is of a hand holding a knife; Image 9 is a photo taken from above of a large cage with a person lying on its floor; finally, Image 10 is of a small bench on a horizon.

We have created a web site with a questionnaire in which the user introduces his/her demographic data, their tags for the photographs (tag session) and some questions which the user answers after completing the session. The capture of tag sessions has been carried out for native and non-native English, and our website reference is:

http://www.tradumatica.net/bmesa/interact2007/index_en.htm .

Tag Session Capture. During a tag session the users must assign between 4 and 10 tags which are related to the objects which they can see in the image and a similar number of tags related to what each image evokes for them, in terms of sensations or emotions. With reference to Figure 1, in the first column the user writes the tags which express what they see in the image, while in the second column the user writes the tags which describe what the image evokes. We have currently accumulated a total of 162 user tag sessions from 2 different countries, involving the tasks of description of the photographs in English. For approximately half of the users, English is their native language and for the other half it is a second language.

Raw Data and Derived Factors. From the tags collected and the information which the users have provided, we can compare results in the English language used by native and non natives in that

Words that describe what you actually see in this picture:	Words that describe what is suggested by this picture:
1. hands	1. peace
2. army	2. agreement
3. uniform	3. encounter
4. sand	4. treaty
5. greeting	5. truce
6.	6. union
7.	7. concord
8.	8. accordance
9.	9.
10.	10.

Figure 1: Example of how the user enters the tags for a given image.

language. Our data is captured from taggers in the United States (native) and from Spain (non native). For each tag session, we collect the following information: language in which the tag session is conducted; easiest image to tag (user is asked); most difficult image to tag (user is asked); the tags themselves assigned for each image, for "See" and "Evoke" separately, and the order in which the tag is assigned. We also record the type of language (if the current tagging language is native or not for the user).

The following factors were derived from the tagging session data (statistically averaged and grouped by user and image):

-Easiness: average number of tags used for "see" and "evoke". This value is compared with the question which refers to the ease or difficulty which a user had to tag the image for "see" and in "evoke". One assumption is that the images evaluated as easier to tag should have more tags. Also, users who possess a greater descriptive vocabulary in the tagging language should define a greater number of tags.

-Similarity: frequency of the tags used for "see" and for "evoke". The tags which present a greater frequency in each image will be compared to detect similarities or differences between native and non-native taggers.

-Spontaneity: tags used as first option for "see" and for "evoke". The tags which appear as first option in each image will be compared to detect similarities or differences between native and non-native taggers.

4 DATA PROCESSING

In this section we explain the factors we derived from the raw data and some statistics about the tags themselves.

4.1 Derived Factors

The following factors were derived from the tag session data:

"Easiness" is represented by the following six factors: "anumTagsSee", "anumTagsEvoke", "asnumTermsSee", "asnumTermsEvoke", "aanumTermsSee" and "aanumTermsEvoke". These factors represent, respectively, the average number (for all images) of tags used for "See", the average number (for all images) of tags used for "Evoke", the average of the sum (for each image) of the number of terms used in each tag for "See", the average of the sum (for each image) of the number of terms used in each tag for "Evoke", the average number of terms (for each tag) used for "See" tags and the average number of terms (for each tag) used for "Evoke" tags. We recall that all these values are summarized by image and user, and that a tag consists of one or more terms (individual words).

"**Similarity"** is represented by the following four factors: "asimSee", "asimEvoke", "atotSimSee" and "atotSimEvoke". The factor "aSimSee" represents the average similarity of a given tagging of an image by a given user for "See", in comparison with all other taggings of the same image by all other users. This is essentially a frequency count of tag coincidences. The factor "aSimEvoke" represents the same statistic as "aSimSee", but calculated for the "Evoke" type tags. The factor "atotSimSee" is equal to "asimSee' divided by the number of users, which gives a sort of 'normalized' value. The factor "atotSimEvoke" represents the same statistic as "atotSimSee", but calculated for the "Evoke" type tags.

"Spontaneity" is represented by the following two factors: "aespSee" and "aespEvoke". The factor "aespSee" represents the spontaneity of a given tagging of an image in a given tag session for "See", by comparing it with the most frequent tags chosen as first option for the same Image.

Table 1: Derived dataset for data analysis and modeling.

Factor Name	Factor Type*
easiest	
mostDifficult	
anumTagsSee	E
anumTagsEvoke	E
asnumTermsSee	E
asnumTermsEvoke	E
aanumTermsSee	E
aanumTermsEvoke	E
asimSee	SI
asimEvoke	SI
atotSimSee	SI
atotSimEvoke	SI
aespSee	SP
aespEvoke	SP
typeLanguage	

*E=easiness, SI=similarity, SP=spontaneity

The factor "aespEvoke" represents the same statistic as "aespSee", but calculated for the "Evoke" type tags. With reference to Table 1, the derived 'See' and 'Evoke' factor session data is held in a table with this structure. All the data has been aggregated by user and image. The attribute "typeLanguage" is the "point of reference" for the data analysis and modeling. If the users' native language is not English, then 'typeLanguage'=2, whereas if the users native language is English, then 'typeLanguage' = 1. This indicator is used as the output, or labeling class.

4.2 Basic Statistics of Users and Tags

In this section we present the basic statistics and frequencies for the tags assigned by the users, which represents the data contained in the structure of Table 1.

With respect to the user "demographic" attributes, there was a similarity of characteristics in terms of the proportions of each type of category, between the native and the non-native groups. The average user age is quite young, 28 and 19 years respectively for native and non native taggers. This is because the majority of the 'volunteers' were university students.

4.2.1 Tag Statistics

With reference to the tags, in Table 2 we summarize some of the most frequent tags for image 10 (small bench on a horizon). Image 10 was considered one of the most difficult images, only by the non natives.

Table 2: Most popular tags for see and evoke (Image 10).

	See (tag, %*)		Evoke (tag, %*)	
Native	bench	91.9	peace	29.7
	sky	89.1	nature	16.2
	grass	83.7	open	13.5
Non native	sky	83.9	peace	34.7
	grass	82.2	loneliness	29.7
	bench	77.3	freedom	24.6

*percentage of the total of users who chose the tag

With reference to Table 2, we observe a clear tendency for see and evoke type tags: the most popular tags for 'see' have a much higher percentage of users who chose them than the most popular tags for 'evoke'. This implies that for the evoke tags, users chose tags which were more different with respect to those of other users, and with a greater distribution over a more diverse set of tags.

This is consistent with the hypothesis that a see tag is assigned in a more stereotypic and spontaneous manner, and that the evoke tag requires more thought and is assigned as a more individual/personal response to the image. If we compare the tags of natives to those of non natives, we see a general coincidence for the see type tags (first three) and for the first evoke type tag.

In Figure 2, we see a plot of the log of the frequency of occurrence of the tags (on the y axis) against the tag id/index (on the x-axis). In general there is a 'zipf' type distribution with a small number of high frequency tags and a larger number of unique tags. A clear trend is evident between native and non native taggers: the natives have a significantly shorter tag distribution (the range for native evoke tags {(d) in Figure 2} is from 1 to 83, whereas the range for non-native evoke tags (b) is from 1 to 244. This is due to the significantly higher error rate in tag definition for non-natives which gives rise to a significantly higher incidence of 'unique' tags. A second trend is evident if we compare see and evoke type tags: the see tags have a shorter distribution. For example, in Figure 2 native see tags (c) range from 1 to 32, whereas the native evoke tags (d) range from 1 to 83. A similar subtrend is shown for non native taggers. This confirms the hypothesis that evoke tags are more diverse than see tags.

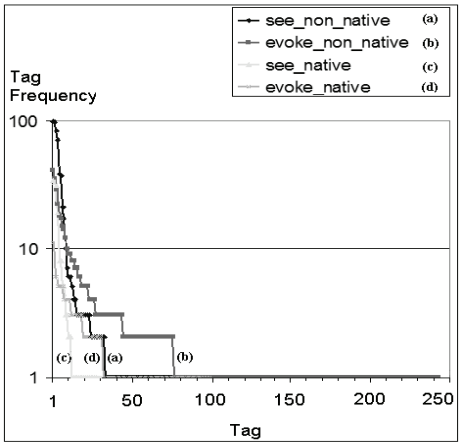

Figure 2: Distributions of frequencies (log scale of y axis) of tags for Image 10.

5 QUANTITATIVE EVALUATION

In this section we show results of the data analysis and data modeling using the IM4Data (IBM Intelligent Miner for Data V6.1.1) Data Mining tool (Im4Data, 2002).

Data Analysis – Statistical Methods and Visualization. Figures 3 and 4 are produced from the 'SessionD' dataset for native English taggers and non-native taggers, respectively. They are ordered by the Chi-squared statistic relative to the 'typeLanguage' label. We recall that this dataset contains attributes which represent the 'easiness', 'similarity' and 'spontaneity' factors for the user tag sessions. Refer to the definitions of these factors in Sections 3 and 4 of the paper. We observe that the first four ranked attributes in Figure 3 (native) and Figure 4 (non native) are 'atotSimEvoke', 'mostDifficult', 'asimEvoke' and 'aespSee', although the ordering is different for attributes 2 to 4. From this we observe that two of the attributes most related to the native/non native label (as indicated by Chi-Squared) are variables related to the similarity of the evoke type tags.

This is coherent with the hypothesis that non native users will find it more difficult to think of vocabulary to define emotions. If we look at the distributions of 'atotsimEvoke' and 'asimEvoke' in Figures 3 and 4, we see that the non-natives (Figure 4) have a greater frequency in the higher (rightmost) part of the distribution, which means that there is more coincidence between the non-native tags, and therefore less diversity. Rule Extraction. The IM4Data tree/rule induction algorithm was used for data modeling. For testing, we have manually created test datasets using a 5x2-fold cross-validation. We used 14 input attributes: easiest, mostDifficult, anumTagsSee, anumTagsEvoke, asnumTermsSee, asnumTermsEvoke, aanumTermsSee, aanumTermsEvoke, asimSee,

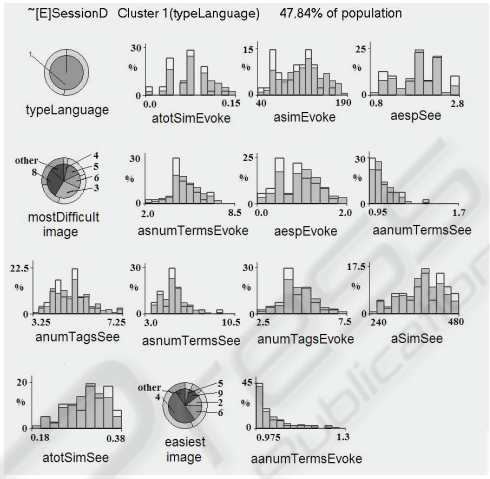

Figure 3: Distributions of variables of dataset 'SessionD', for native English taggers.

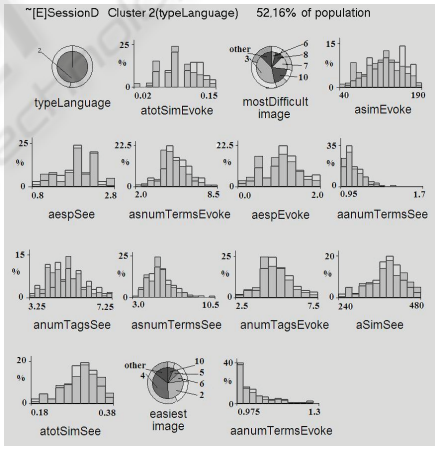

Figure 4: Distributions of variables of dataset 'SessionD', for non native taggers.

asimEvoke, atotSimSee, atotSimEvoke, aespSee, aespEvoke; and one output attribute (class): 'typeLanguage'.

With reference to Figure 5, we see the pruned tree induced by IM4Data on the SessionD dataset, including the details of the decision nodes and classification nodes. We observe that attributes 'asimEvoke' and 'mostDifficult' have been used in the upper part of the tree (asimEvoke < 138.15, mostDifficult in [image9, image3, ,image10, image7]). Thus, they represent the most general and

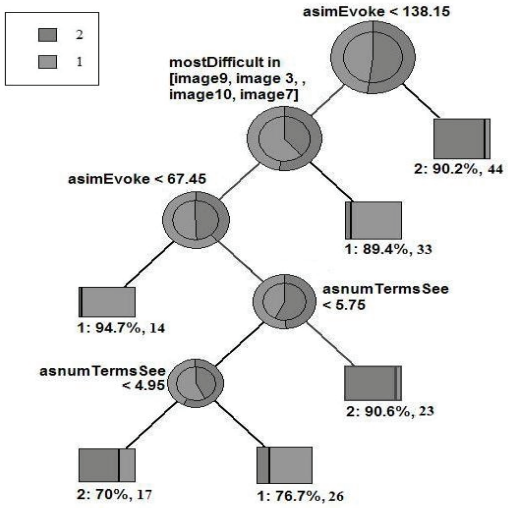

Figure 5: Pruned Classification Tree: dataset 'SessionD'.

discriminatory factors to classify 'typeLanguage', that is the native and non-native users. We note that lower down in the tree the attribute 'asnumTermsSee' has been used.

With reference to Table 3, we present the test results (test folds) for the tree induction model built from the SessionD factors. The overall precision of the model over 5 folds is 75.63%. The low percentage of false positives and false negatives over the five folds indicates that we have a 'robust' model. We conclude from the results that with the derived factors for 'Easiness', 'Similarity' and 'Spontaneity' we are able to produce an acceptably precise model (75.63%), using real data and 'typeLanguage' as the output class. This model distinguishes between English native and non-native taggers, based on the given input variables and derived factors.

Table 3: 'SessionD': test precision for 5x2 fold cross validation.

	native†	non-native††	MP*
fold1	65.5, 21.1	78.9, 34.5	71.08
fold2	88.3, 32.2	67.8, 11.7	77.07
fold3	85.2, 33.9	66.1, 14.3	76.17
fold4	70.6, 34.4	65.6, 29.4	77.60
fold5	89.6, 35.0	65.0, 10.4	76.42
Geometric mean for folds	79.2, 30.8	68.5, 17.7	75.63

*MP=Model Precision †{%Rate: True Positive, False Positive}, ††,{%Rate: True Negative, False Negative}

6 CONCLUSIONS

As conclusions from the present work and the available data and derived factors, we can reasonably infer that there is a significant difference between "see" and "evoke" type tags, which is related to if the user is native or not in the tagging language. We have successfully built a data model from the derived factors (Figure 5, Table 3). We have determined that non native taggers have distinctive characteristics especially for the similarity of subjective type tags. The initial hypothesis of greater quality and diversity of tags has been confirmed for native users. From a user support point of view, the findings can be used in online applications, for example, the recommendation of evoke type tags for non-native users by using the tags defined by the best native taggers.

REFERENCES

Anderson, A., Raghunathan. K., Vogel, A., 2008. TagEz: Flickr Tag Recommendation. Association for the Advancement of Artificial Intelligence (www.aaai.org). http://cs.stanford.edu/people/acvogel/tagez/

Boehner, K., DePaula R., Dourish, P., Sengers, P., 2007. How emotion is made and measured. International Journal of Human-Computer Studies. 65:4, 275-291.

Garg, N., Weber, I., 2008. Personalized, interactive tag recommendation for flickr. Proceedings of the 2008 ACM conference on Recommender Systems, Lausanne, Switzerland. pp. 67-74, ISBN:978-1-60558-093-7.

Im4Data, 2002. Using the Intelligent Miner for Data V8 Rel. 1. IBM Redbooks, SH12-6394-00.

Isbister, K., Hook, K., 2007. Evaluating affective interactions. International Journal of Human-Computer Studies, 65:4, 273-274.

Lipczak, M., Angelova, R., Milios, E., 2008. Tag Recommendation for Folksonomies Oriented towards Individual Users. ECML PKDD Discovery Challenge 2008, Proceedings of WWW 2008.

Sigurbjörnsson, B., van Zwol, R., 2008. Flickr Tag Recommendation based on Collective Knowledge. WWW 2008, Beijing, China, ACM 978-1-60558-085-2/08/04.

Song, Y., 2008. Real-time Automatic Tag Recommendation. SIGIR'08, July 20–24, 2008, Singapore. ACM 978-1-60558-164-4

Sood, S.C., Hammond, K., Owsley, S.H., Birnbaum, L., 2007. TagAssist: Automatic Tag Suggestion for Blog Posts. International Conf. on Weblogs and Social Media (ICWSM), 2007, Boulder, Colorado, USA.