

Classification of Photos based on Good Feelings

ACM MM 2012 Multimedia Grand Challenge Submission

Mathias Lux
Klagenfurt University
Universitätsstrasse 65-67
Klagenfurt, Austria
mlux@itec.uni-klu.ac.at

Mario Taschwer
Klagenfurt University
Universitätsstrasse 65-67
Klagenfurt, Austria
mt@itec.aau.at

Oge Marques
Florida Atlantic University
777 Glades Road
Boca Raton, FL, USA
omarques@fau.edu

ABSTRACT

Many photo-taking scenarios have emotional triggers such as happiness, joy, or in general a *good feeling*. In this paper we postulate that if a photo was taken to preserve a good feeling, it will most likely trigger a good feeling when re-visiting the photo later on. We present an approach to classify photos and decide automatically whether they have been taken to preserve a good feeling or not. We employ an annotated data set to create a model and apply this model to enhance classical image search. The result is a novel search approach, demonstrated by a proof-of-concept application capable of performing keyword-based search and filtering out the results that are considered unlikely to convey a good feeling to the viewer.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

General Terms

Experimentation, Human Factors

Keywords

User intentions, affection, image search, image classification

1. INTRODUCTION

The process of taking a photograph is typically triggered by an intention¹. For example, a picture of a car's engine could be used to document a repair process or a visible engine damage. The same photo, on the other hand, might have been taken because the photographer just bought the car and is really happy about her purchase. In the latter case, the photo has been taken to preserve a *good feeling* and to recall it at a later time. Examples of situations where the majority of associated photos are shot to preserve good feelings include: pictures of kids, pictures taken on a holiday, and pictures taken at social events such as concerts or birthday parties, among many others.

Photographs *taken* to preserve good feelings are often revisited to *recall* such feelings, not only in the person who

took the photo but in other viewers as well. For instance, when a mother captures the first steps of her baby in a photograph, the feeling of happiness will most likely be triggered even years after the actual event, when re-visiting the photos. Moreover, similar feelings may be experienced by people who don't even know the baby or the mother, because the photo reminds them of a related situation in their lives. Our work is based on the assumption that photos taken with the specific intention to preserve a *good feeling* (e.g., love, joy or happiness) are more likely to trigger a similar response in the viewer of such photographs.

In this paper we contribute a classifier that decides automatically whether photographs have been taken with the intention of preserving a good feeling. We also present an application that makes use of this classification model, which serves as a proof of concept to demonstrate that the method is applicable in a production scenario. The application performs image search by keywords and returns a ranked result list filtered by the classifier, which allows for showing (only) photos that preserve good feelings.

2. APPROACH

We built a data set of images annotated with different intentions of the photographers. We acquired a data set of 1,309 photos, selected randomly from Flickr, employing an RSS feed featuring "recent additions to Flickr", between June and September 2011. For each photo the original intention of the photographer was surveyed and the survey results were validated with Amazon Mechanical Turk (mturk)². We extracted a subset with cross-validated annotations indicating whether the image was taken to preserve good feelings or not. The resulting subset was used to train a classifier that is capable of making the same decision (i.e. determine if the image was taken to preserve good feelings or not) on new (test) images. After experimenting with several image and metadata features and classifier algorithms, a purely text-based classifier based on Bayesian networks was chosen. The remainder of this section explains the pertinent technical details.

Our main assumption is that if a photo has been taken to preserve a good feeling, it will likely trigger a similar feeling while viewing the photo. Therefore we employ the novel approach of focusing on the *intentions of photographers while taking the photos* to train our classifier. The first step was to reduce the number of 1,309 instances to those that are consistent with mturk validation. We focused on the re-

¹a *thing intended; an aim or plan* as defined by the Oxford Dictionary.

²<http://www.mturk.com/>

sult of the survey statement “*I took the photo to preserve a good feeling*”. Using the mturk input we employed a voting scheme that includes only those instances where the answer $a \in \{1, 2, 3, 4, 5\}$ of the photographer was supported by at least 3 out of 5 mturk HITS h_i with $a - 1 \leq h_i \leq a + 1$. This resulted in 859 instances. Our class attribute c was then:

$$c = \begin{cases} 0, & \text{if } a \in \{1, 2, 3\} \\ 1, & \text{otherwise.} \end{cases}$$

We assume that the photo has been taken to preserve a good feeling only if $c = 1$. This labeling splits the training data set into two sets with 296 (34%, $c = 0$) and 563 (66%, $c = 1$) instances, respectively.

For classification we ranked the terms and tags occurring at least twice in the training set according to their information gain. We then introduced a binary attribute for the first 500 to create an attribute vector for classification. Table 1 lists the top 16 ranked words, as well as their information gain (G) values with respect to the class attribute.

word	$G \times 10^3$	word	$G \times 10^3$
britain	7.19	slum	5.39
light	7.15	beauty	4.28
paris	6.68	flower	3.86
graffiti	5.39	bike	3.69
speed	5.39	thing	3.59
sony	5.39	around	3.59
canada	5.39	toronto	3.59
shanghai	5.39	nope	3.59

Table 1: Top 16 tag or title words of training data set ranked by information gain G with respect to class.

The training set with these selected attributes was used to train a special Bayesian network classifier called *averaged one-dependence estimator* [1]. The classification performance was evaluated using 10-fold cross-validation on the training data set, yielding a classification accuracy of 76.6% (658 instances correctly classified, 201 instances incorrectly). Other evaluation metrics are shown in Table 2.

Class	TP Rate	FP Rate	Precision	Recall
0	0.490	0.089	0.744	0.490
1	0.911	0.510	0.773	0.911

Table 2: Results of classifier cross-validation showing true-positive and false-positive rates, precision, and recall for negative (class 0) and positive (class 1) decisions on “taken to preserve a good feeling”.

Note that other classifiers and weighting schemes like TF-IDF commonly used for text mining gave significantly worse results. Also results from content based classification were significantly worse.

3. APPLICATION

The application presented in this paper triggers a search process on Flickr based on a set of keywords entered by the user and downloads and analyzes the first 36 results selected from the result list returned by the Flickr API³. Im-

³<http://www.flickr.com/services/api/>

plemented as a Java desktop application, it allows the user to search for arbitrary text (see Fig. 1). Additional options include searching for most interesting photos (ranked by the Flickr *interestingness* value) and restricting the search results to Creative Commons licensed ones. After downloading all 36 photos and processing them by our classifier, the results are shown. Clicking on the good feeling icon in the top right corner of the results list hides images that our classifier does not recognize as photos taken to preserve good feelings.

Fig. 1 shows four screenshots of a typical workflow of the application. Top-left: query form with the query “boat water sea”. Top-right: processing screen. Bottom-left: (un-filtered) results of the query. Bottom-right: result of the filtering process (photos not considered to convey a good feeling are greyed out).

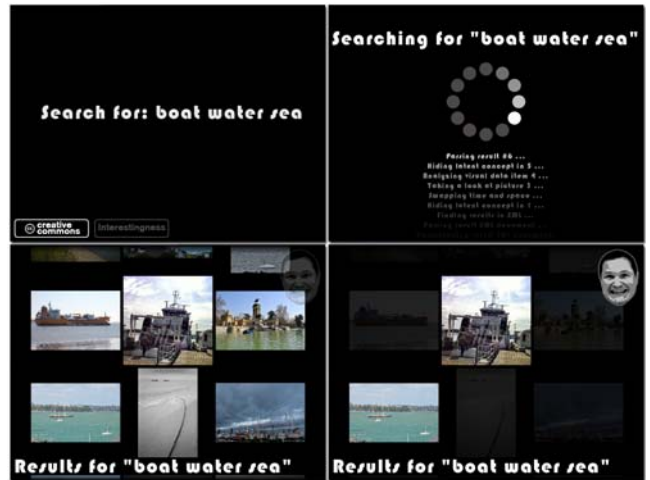


Figure 1: Screenshots of the application. Note the bottom-right picture, where the filter has been applied and only images conveying good feelings are shown.

4. CONCLUSIONS

In this submission we presented a novel approach to classifying photos based on the intentions of the photographers. Photos that have been taken to preserve a good feeling are recognized and highlighted by a demo application.

5. ACKNOWLEDGMENTS

We’d like to thank A. Kopper and D. Khura for data acquisition, cleaning and packaging of the employed data set.

Parts of this work were supported by the Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant 20214/17097/24774.

6. REFERENCES

- [1] G. I. Webb, J. R. Boughton, and Z. Wang. Not so naive Bayes: Aggregating one-dependence estimators. *Mach. Learn.*, 58(1):5–24, Jan. 2005.