About Emotion Identification in Visual Sentiment Analysis

Olga Kanishcheva
Institute of Information and
Communication Technologies,
Bulgarian Academy of Sciences
kanichshevaolga@gmail.com

Galia Angelova
Institute of Information and
Communication Technologies,
Bulgarian Academy of Sciences
galia@lml.bas.bg

Abstract

In this paper we present an approach for analysis of sentiments and emotions in image tagging using SentiWordNet as an external linguistic resource of emotional words. Our aim is to design and implement algorithms that assess the emotions and polarity given a set of image tags. The approach is not limited to object analysis only (considering informational keywords) but deals with the involvement tags and employs some techniques used for sentiment analysis in social networks. We consider the issue of tag sense disambiguation when image keywords are mapped to SentiWordNet. The Lesk algorithm helps to identify correctly the meaning of about 50% of the ambiguous single keywords of 200 images. The total number of tags we process is about 10,000. Calculating a "sentiment score" for each image, the system classifies images into three classes (positive, negative, neutral). These classes are compared to emotional assessments done *(i)* by humans and *(ii)* by training of a SVM classifier that provides the baseline of 69.7% precision, 29.9% recall and 41.8% F-measure. Our approach works with 63.53% precision, 58.7% recall and 61.02% F-measure. The experiments are performed using the annotations of the industrial auto-tagging platform Imagga that identifies automatically image objects with high precision.

1 Introduction

Folksonomies are recognized as a recent type of internet classification system where non-professional users add their own keywords (tags) to information objects. These tags could then be used by anyone to sort and share items. "*Folksonomy*" became the word most commonly used to refer to this annotation approach that is also known as ethnoclassification, social classification/tagging, collaborative tagging, social indexing and distributed classification.

Peters and Weller (2008) wrote that annotation development via crowdsourcing and the resulting folksonomies provide many advantages such as diverse opinions, independent decision-making, decentralization of power, and a way of aggregating opinions. Most current systems that facilitate tagging do not require any sort of text verification or controlled vocabulary. In this way the diversity of opinions allowed in tagging is limitless and the annotators independently select the tags they want to use. Finally, folksonomies provide the aggregation of opinions in the form of systems such as Flickr (https://www.flickr.com), Instagram (https://instagram.com/), Picasaweb (picasaweb.com/), Photobucket (http://photobucket.com/) and others.

Images in these large collections are retrieved using keywords specified by users. For example, searching with tag "*London*" returns the list of links to all photos annotated with this keyword. Thus the semantic information, which is saved up in the metadata, enables the development of various searching strategies that rely significantly on automatic text processing, lexical hierarchies and information search techniques.

Emotional words take special place among folksonomy tags. For instance Beaudoin (2007) suggests that the emotional elements and other parts of speech that express sentiments, such as adjectives, are classified in various categories. To define "tag sentiment", it is necessary to use tags from various categories. It is well known that user-defined tags in folksonomies contain a lot of emotional markers. Sentiments are most often expressed by adjectives (*attractive, cool, funny, pretty, beautiful, happy* etc.), verbs (*hate, admire*) and especially interjections – words that

Proceedings of Recent Advances in Natural Language Processing, pages 258–265,
Hissar, Bulgaria, Sep 7–9 2015.

bear no meaning by themselves but are well loaded with emotionality, such as *ah, wow, oops, hey,* etc. Besides users can enter emoticons and expressive lengthening of words for identification or strengthening the emotional relation to the image (*:D, ☺, coooool* etc.). All these emotional markers are considered as one category – the so-called "*subjective tags*" because they express users' opinion and emotion, e.g., *funny* or *cool*. They can help evaluating qualities and recommendations. Subjective tags are assigned to digital objects primarily with a motivation of self-expression.

In this paper we propose an integrated approach to sentiment analysis of image tags – coming from user-defined folksonomies and auto-tagging systems. The paper is organized as follows: Section 2 summarizes some related work; Section 3 overviews the sentiment analysis achievements and linguistic resources that provide information about emotional words. In Section 4 we present the Imagga Auto-Tagging Program (ATP) – the source of our test corpus of automatically annotated images. Section 5 describes the suggested approach in more detail; Section 6 presents current results. Section 7 contains the conclusion and plans for future work.

2 Related Work

A large number of research works appeared recently that address sentiment analysis and its relation to image annotation, including: visual aspects of sentiment analysis (Borth et al., 2013; Jia et al., 2012; Machajdik and Hunbury, 2010; Hailin et al., 2015) and hybrid approaches which analyze emotions using additional resources (Yang et al., 2013; Yang et al., 2014). The basic idea is to build a sophisticated feature space that can effectively represent the sentiment status of texts and/or images.

Jia et al. (2012) present a prediction of sentiment reflected in visual content. The authors propose a systematic, data-driven methodology to construct a large-scale sentiment ontology built upon psychology and web crawled folksonomies using SentiBank. The authors also used the psychological theory Plutchik's Wheel of Emotions as the guiding principle to construct a large-scale visual sentiment ontology that consists of more than 3,000 semantic concepts.

Chen et al. (2014) created a hierarchical system to model object-based visual sentiment concepts. The system handles sentiment concept classification in object-specific manner. It tackles the challenges of concept localization and resolving sentiment attribute ambiguity.

The systems presented in (Borth et al., 2013; Chen et al., 2014) are based only on analysis of Adjective Noun Pairs (ANP) such as "*beautiful flower*" or "*disgusting food*". The advantage of using ANPs, compared to nouns or adjectives only, is the potential to turn a neutral noun like "*dog*" into an ANP with strong sentiment like "*cute dog*" by adding an adjective with a strong sentiment. Authors claim that such phrases also make the concepts more detectable than single adjectives (e.g. "*beautiful*") which are typically abstract and difficult to detect.

Yang et al. (2014) applied a lexicon-based sentiment method from (Esuli and Sebastiani, 2006) to analyze the corresponding textual sentiment that is further used to cluster and rank the related images. Then, to link images in social networks that have similar emotions but different visual contents, the authors combine the social links with visual similarity between images, constructing a "*visual-social similarity matrix*" that quantifies image similarities from both visual and social perspectives. They propose the ViSoRank algorithm to identify representative images on the inferred visual-social similarity graph and the VSTRank algorithm to combine them together to discover the emotionally representative images for social events. Only two sentiment categories are used (positive and negative).

Ignacio Fernández-Tobías et al. (2013) present a model which is built upon an automatically generated lexicon that describes emotions by means of synonym and antonym terms, and that is linked to multiple domain-specific emotional folksonomies extracted from entertainment social tagging systems. Using these cross-domain folksonomies, the authors develop a number of techniques that automatically transform tag-based item profiles into emotion-oriented item profiles. This approach is applied for folksonomies in the movie and music domains.

Siersdorfer et al. (2010) consider the "bag-of-visual words" representation as well as the color distribution of images, and make use of the SentiWordNet thesaurus to extract numerical values for image sentiment from associated textual metadata. Then they perform a discriminative feature analysis based on information theoretic methods and apply machine learning techniques to predict the sentiment of images.

3 Modelling Sentiment

Sentiment analysis aims to determine the attitude of speaker or writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication. The overall scheme of attitudes is often called "*tonality*".

In general approaches to tonality classification are based: *(i)* on dictionary matching, *(ii)* on (supervised or unsupervised) Machine Learning, and *(iii)* on hybrid methods. Some methods require dictionaries; others need annotated corpora. More sophisticated methods try to identify the mood and the object to which feelings are expressed. Tonality is measured by a predefined rank of emotional intensity of the feelings expressed by words or phrases. Often no text context is available in image tagging to help evaluating the emotional content as we deal mostly with isolated keywords.

In our work we use a dictionary-based approach for determining the tone of tags. Affective lexicons contain lists of words with tonality value for each word. One of the most popular linguistic resources for sentiment analysis is SentiWordNet, see Esuli and Sebastiani (2006).

SentiWordNet (http://sentiwordnet.isti.cnr.it/) is a lexical resource for opinion mining that consists of more 117,000 words. It appeared after automatic annotation of each WordNet synset with scores according to its degree of positivity, negativity, and objectivity. In this way three numerical values are assigned to each WordNet synset to define explicitly the objective, positive or negative component of the synset. Each value ranges in the interval [0,1] and their sum is 1. Words have various senses and therefore, can be assigned various respective values for objective, positive or negative components. SentiWordNet in used in our experiments because it is large enough to cover many tags we consider.

4 Auto-Tagging of Images

The company IMAGGA (http://imagga.com) has developed an original technology for image auto-tagging by English keywords. The technology is based on machine learning and assigns to each image a set of keywords depending on shapes that are recognized in the image. For each learned item the system "sees" in an image, appropriate tags are suggested. In addition the system proposes more tags based on multiple models that it has learned. They relate the visual characteristics of each image with associated tags of “similar” images in ImageNet or big external manually created data sets (e.g. Flickr). The intuition and motivation is that more tags serve better in searching because users may express their requests by different wordforms. The platform developers believe they have found the right practical way to offer best possible image annotation solution for a lot of use-cases.

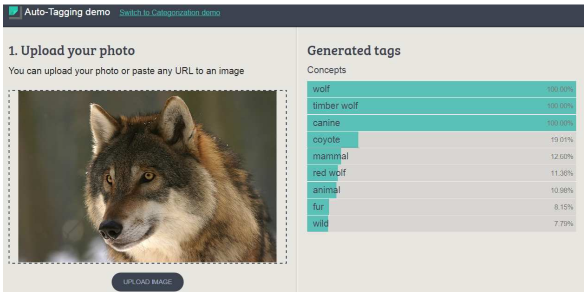

Figure1: Imagga's auto-tagging platform with automatically generated tags and their relevance scores: *wolf* 100%, *timber wolf* 100%, *canine* 100%, *coyote* 19%, *mammal* 12,6%, *red wolf* 11,36%, *animal* 10,98%, *fur* 8,15%, *wild* 7,79%

Quite often, when the image contains a close up object, Imagga's platform assigns correctly the most relevant tags to the central object (Fig. 1). In the right part of Fig. 1 keywords are ordered according to their relevance score. Associating external tags imports emotional keywords in the annotation of Imagga's images. We note that complex tags are not limited only to ANPs like in (Borth et al., 2013; Chen et al., 2014) – see e.g. "*timber wolf*" in Fig. 1.

5 Emotional Classification of Images

Our approach is sketched in Fig. 2: an image with folksonomy (i.e. manual) tags arrives to the system, then it is analyzed by the auto-tagging program and the central object is defined together with the corresponding tags. We designate the user-defined tags as t_{f_i}, $i=\overline{1,n}$, where n is the number of tags from the folksonomy. The tags assigned by the auto-tagging program are denoted by t_{p_j}, $j=\overline{1,m}$. Thus we receive a set of $n+m$ tags which characterize the image.

Keywords assigned by users have higher priority than tags calculated by the program. To distinguish the contribution of these tags’ emotional content to the unified tagset sentiment, we define two addition coefficients α and β: the coefficient α shows the degree of priority of folksonomy (authos’) tags and coefficient β denotes

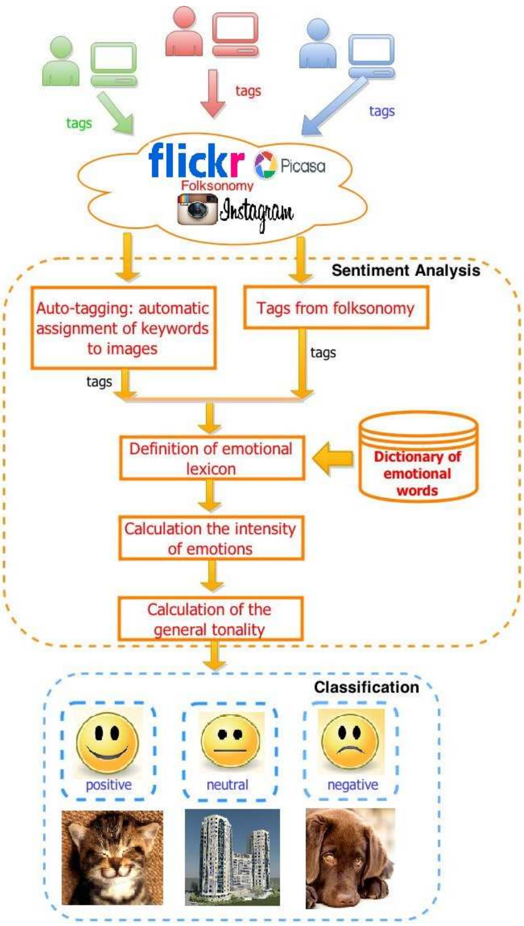

Figure 2: General scheme of our approach

the degree of priority of the ATP-keywords. SentiWordNet contains the predefined emotional polarity *p(w)* of a sense of the word *w* with values for its positive and negative components – *PosScore(w)*, *NegScore(w)* as shown in Table 1.

For each image annotated with tagset S_{tags} we define the value of the positive tag component *P* for this image: $P = \sum_{w \in S_{tags}} PosScore(w)$.

Pos-Score	Neg-Score	Synset Terms	Gloss
0.875	0	attrac-tive#1	pleasing to the eye or mind especially through beauty or charm; "a remarkably attractive young man"; "an attractive personal-ity"; "attractive clothes"; "a book with attractive illustrations"
0.125	0.375	long#9	having or being more than normal or neces-sary: "long on brains"; "in long supply"

Table 1: Structure of the SentiWordNet dictionary

The value of the negative tag component *N* is calculated similarly, by summing up all *NegScore*-s of the keywords. Note that *P* and *N* can be zero. Calculating positive and negative scores of certain image helps to measure the emotional intensity of the keywords as a whole, no matter whether it is positive or negative.

We introduce the notion of image sentiment *T*:

$$T = \begin{cases} \dfrac{\sum_{w \in t_f} \alpha \cdot PosScore(w) + \sum_{w \in t_p} \beta \cdot PosScore(w)}{\sum_{w \in t_f} \alpha \cdot NegScore(w) + \sum_{w \in t_p} \beta \cdot NegScore(w)} + \gamma \cdot N_{EE}, & N \neq 0, \\ 1 & N = 0, \quad P = 0, \\ \sum_{w \in t_f} \alpha \cdot PosScore(w) + \sum_{w \in t_p} \beta \cdot PosScore(w) + \gamma \cdot N_{EE}, & N = 0, \quad P \neq 0, \end{cases} \qquad (1)$$

where γ is a coefficient defining the intensity of emotional elements and N_{EE} is the number of emotional elements in the tagset. The formula (1) takes into account the number of emotional elements such as emoticons (*:D*, ☺), lengthenings (*coool*), interjections (*bravo*, *oh*) etc. In our experiments, all emotional elements have equal weight given by the coefficient γ that ranges in the interval [0,1]. Note that image sentiment can be calculated for authors' tag and the auto-tags separately, for instance formula (1') defines image sentiment using folksonomy tags only:

$$T' = \begin{cases} \dfrac{\sum_{w \in t_f} \alpha \cdot PosScore(w)}{\sum_{w \in t_f} \alpha \cdot NegScore(w)} + \gamma \cdot N_{EE}, & N \neq 0, \\ 1 & N = 0, \quad P = 0, \\ \sum_{w \in t_f} \alpha \cdot PosScore(w) + \gamma \cdot N_{EE}, & N = 0, \quad P \neq 0, \end{cases} \qquad (1')$$

After calculating the tonalities T_1, T_2, T_3, … of all images, we classify the latter into three categories *positive*, *negative*, and *neutral* and rank them within each group. Images with value $1 \le T \le 1.5$ are considered *neutral*; with $T > 1.5$ – *positive*; and with $T < 1$ – *negative*.

6 Experiments and Discussion

We deal with 200 images from 7 Flickr categories (people, animals, cars, houses, flowers, nature and miscellaneous) that have original author's annotation, in average 19 tags per image. To ensure independent opinion about their sentiment, all images were classified manually by two independent humans into three categories: *ExPos* (92 positive images), *ExNeg* (89 negative images) and *ExNeur* (19 neutral ones). No tags were shown to these annotators so they gave individual assessment looking at the image only. Images with controversial judgment are rejected. The resulting 200 pictures are the dataset we use.

In addition these 200 images were annotated by the Imagga ATP. Tables 2 and 3 present numbers of tags and their intersection with the SentiWordNet items. Note that "tags" come from the dataset but when mapping them to SentiWordNet we split them to tokens, e.g. "*Tokina 11-16mm f/2.8*" will be split into 3 sub-strings.

		Numbers
	Total tagsets	200
Tags assigned by authors	Total	3761
	of them unique	1715
Tags assigned by the ATP	Total	6103
	of them unique	597
Avg #tags per image, given	by authors	19
	by the ATP	30

Table 2: Assignment of 9864 tags in the test dataset

	Human tags	ATP tags	Total
Only pos-score	260	86	346
Only neg-score	233	75	308
Neutr(no score)	1505	746	2251
Pos.& neg. scores	107	38	145
Single sense	358	157	515
Many senses	1747	788	2535
Interjections	6		6
Lengthening	17		17

Table 3: Mapping test dataset' tags to SentiWordNet

Among the 9,864 tags in the test dataset, some 3,050 were found in SentiWordNet: 2,105 are assigned by authors and 945 by the ATP. Table 3 shows that 2,535 of these tags are polysemous so we used the Lesk WSD algorithm (Lesk, 1986) to distinguish which tag sense is mentioned in a particular image annotation. For each polysemous tag, we mapped the whole tagset of the respective image to a SentiWordNet gloss. The sense that overlaps maximally with the "annotation context" was considered to be the correct one. Some examples follow below:

***Example 1 for "homeless"*: Author's Tags** – Nikon D80 homeless man lisbon portugal obdachlos street life *poor* man; **SentiWordNet** homeless#2 – *poor* people who unfortunately do not have a home to live in ..... Here "*poor*" is a tag that appears in the gloss so "homeless#2" is chosen;

***Example 2 for "ancient"*: Program Tags** – architecture *old* ancient; **SentiWordNet** ancient#2 – very *old*; "an ancient mariner".... Here the sense ancient#2 is selected as the correct one due to the fact that the tag "*old*" appears in the SentiWordNet gloss.

The evaluation shows that the WSD precision in this case is about 50%. For empty overlaps the first sense in the SentiWordNet list is chosen.

Our experiment aims to study whether formula (1) provides a reasonable sentiment score for images. The tests support the rationality of including the keywords, assigned by the ATP, in the calculation of image sentiment. It happens often that the manually-annotated images have small amount of tags. But the ATP delivers further tags and then numerous keywords are associated from external collections with similar images, so the accumulated polarity increases.

We made a number of experiments to assess the behavior of coefficients in formula (1). To give an idea about these tests we present at Fig. 3 the changes of precision for positive and negative classes when α=1 and $\beta \in [0.1, 1]$. The best results *Precision(positive)*=63.53% and *Precision(negative)*=58.93% were received for values α=1 and β=0.4. Similar test were performed for β=1 and $\alpha \in [0.1, 1]$.The optimal coefficient values are α=1 and β=0.4. We assumed that γ = 0.1.

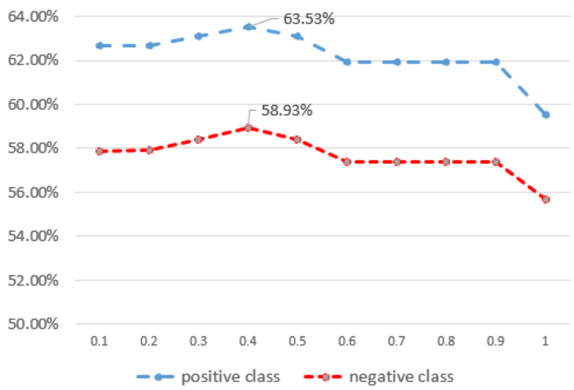

Figure 3: Tests with changes of coefficient β: the dash blue line corresponds to the positive class, the dot red line corresponds to the negative class.

For all pictures in the test collection, we compared the human-defined classes *ExPos*, *ExNeg* and *ExNeut* to image sentiments calculated using the ATP tags in formula (1). Regarding the 89 images in *ExNeg*, the histogram at Fig. 4 shows that the ATP assigned (correctly) keywords with negative tonality only to 57. However the ATP assigns also relevance scores to the keywords so we checked the tonality of auto-tags with relevance score higher than 20%. The success rate improves – 73% (65 out of 89 images) are annotated with negative sentiment by the ATP.

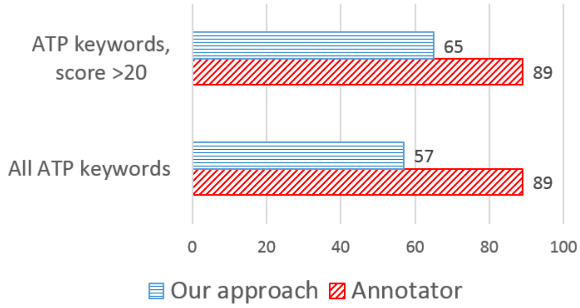

Fig. 4: Computing ATP tags' sentiment for ***ExNeg***

Fig. 5 shows that from all 92 images in *ExPos*, 67 (72.83%) are defined correctly when all ATP tags are considered. Filtering only the keywords with relevance score above 20% reduces also the images with positive sentiment to 54 (58.70%). Actually many ATP keywords with relevance scores lower than 20% are positive; therefore their removal influences significantly the calculations and the results are less successful for *ExPos* (but more successful for *ExNeg*).

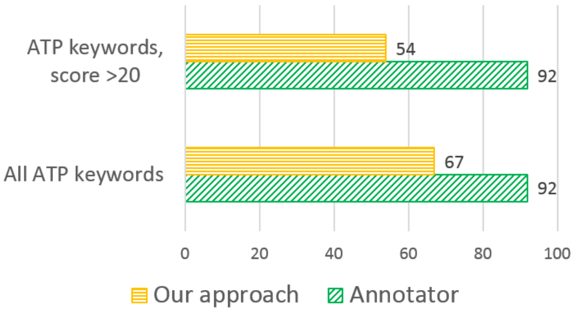

Fig. 5: Computing ATP tags' sentiment for ***ExPos***

The neutral class of 19 images turned to be the trickiest one. The default is – following the intuition behind formula (1) – that an image is "neutral" when it has no emotional tags at all, or when the sentiment of all the positive tags is equal or close to the sentiment of all negative tags. One of 19 images was classified incorrectly by the ATP. Another image with multiple correct tags was annotated with keywords that have strongly negative components in SentiWordNet:

NegScore(monkey) = 0.125,

NegScore(tropical) = 0.5

which lead to T<1 and assignment of negative sentiment. Apparently our approach significantly depends on the linguistic resources and the WSD success. In addition, SentiWordNet scores range in relatively small interval so one tag can change the image sentiment to either positive or negative. Due to this reason images are assigned different values (Fig. 6). To partially decrease these effects, *ExNeut* is defined for $T \in [1, 1.5]$.

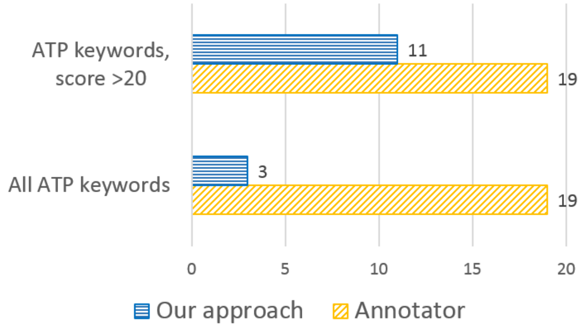

Fig. 6: Computing ATP tags' sentiment for ***ExNeut***

The test dataset contains ATP tags with relevance scores 7-100%. Fig. 7 shows how precision varies depending on the tags' relevance scores. The best precision is achieved for the class *ExPos* using only tags with relevance score>20%. This is related to the ATP features: all high relevance tags are not emotional.

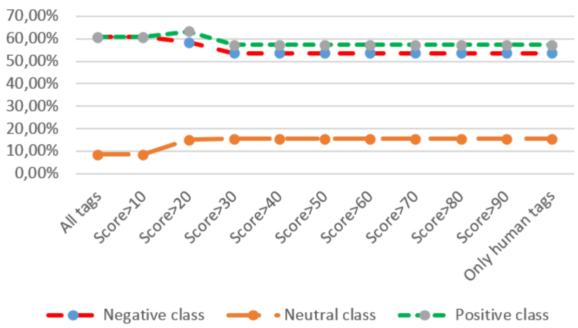

Figure 7: Precision in all classes (shown in Fig. 4, 5, and 6) depending on the tags' relevance scores

The emotional keywords in the test dataset have relevance scores from 20% to nearly 70%. But in general the majority of the positive tags, which are imported from external collection by Imagga's ATP, have relevance scores less than 20%. We remind that about 30% of all 9,864 tags are included in SentiWordNet. Table 3 shows that only 799 have a non-zero sentiment value.

Tables 4 and 5 summarize the comparison between the human-defined classes *ExPos*, *ExNeg* and *ExNeut* and the emotional image scores calculated using SentiWordNet. Table 4 shows calculations using only the author-defined tags and respectively, formula (1').

Low results for *ExNeut* are due to several reasons. First they illustrate the discrepancies of opinions of human-experts who defined *ExPos*, *ExNeg* and *ExNeut* (without seeing image tags)

Class	Recall	Precision	F_1-measure
ExPos	47%	57%	52%
ExNeut	74%	16%	26%
ExNeg	70%	53%	61%

Table 4: Mapping ***ExPos, ExNeg*** and ***ExNeut*** to calculations using formula (1’), for author-assigned tags

Class	Recall	Precision	F_1-measure
Positive	59%	63%	61%
Neutral	58%	15%	24%
Negative	73%	59%	65%

Table 5: Mapping ***ExPos, ExNeg*** and ***ExNeut*** to calculations using formula (1), for all ATP tags

and the picture authors. Table 6 shows further examples of various opinions and perspectives: authors’s tags and the ATP keywords differ substantially. Second, emotional tags are relatively scarce in principle. Finally the lack of adequate linguistics resources prohibits the development of standardized datasets and gold standards.

Author’s tags: *Derwentwater, Lake District, Weather, Wet, Very wet, Rain, Downpour, Torrential Rain, Cloud, Lake District Weather, Stairrods, Heavy rain*

Imagga’s tags: *landscape 41.91%, water 38.13%, lake 34.06%, river 29.05%, trees 27.65%, tree 26.29%, forest 26.22%, …*

Calculated sentiment using SentiWordNet and formula (1): *neutral*

Author’s tags: *garbage, dump*

Imagga’s tags: *food 21.17%, honeycomb 15.97%, spice 15.32%, apiary 14.29%, healthy 12.93%, …*

Calculated sentiment using SentiWordNet and formula (1): *neutral*

Table 6: Sample images belonging to ***ExNeg***

Given the human-defined classes *ExPos*, *ExNeg* and *ExNeut*, we trained a SVM classifier on a subset of 80 images (i.e. 40% of the original dataset). More precisely we used SVM classifiers, which are binary by nature, and combined them into *n*-ary classifiers using the Sequential Minimization Optimization (SMO, Platt 1988) implemented in Weka (Witten, 2011). The remaining 60% of the experimental dataset are used as a test corpus for classifying images as *positive*, *negative* and *neutral*. Thus we have a "SMO-baseline" how tags are related to the human judgment of image sentiment. Fig. 8 shows precision, recall and F_1-measure for the positive class *ExPos*, where our approach is compared to the SMO results. The highest F_1-measure 61.02% is achieved for the suggested formula (1) despite the fact that less than 10% of all tags (799 of 9,864) have non-zero emotional values in SentiWordNet. The proposed idea looks feasible assuming that the activities on development of linguistic resources with affective words will grow.

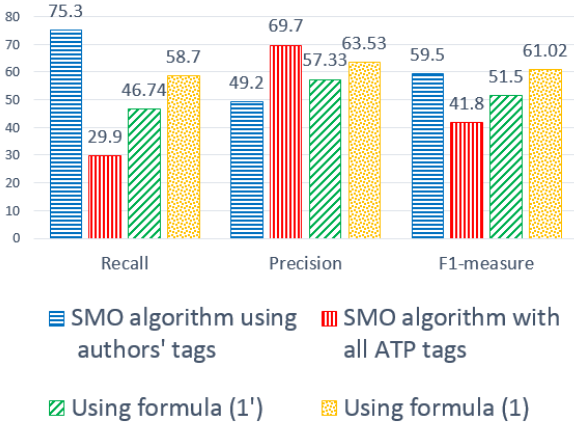

Figure 8: Precision, Recall and F_1-measure for ***ExPos*** using a SMO classification and our approach

Conclusion

The emotional classification of images depends on the individual opinion of each person, but we propose and investigate an idea how to compute image sentiment scores using external resources. Most keywords we use are meant for indexing the image content but the small percentage of positive/negative tags enables automatic calculations. The reported results are similar to those achieved in sentiment analysis and opinion mining where F-measures for evaluation of emotions in social networks are usually below 70%. As future work we plan at first to include colors in the emotional assessment of images.

Acknowledgements

This research is partially supported by the FP7 grant 316087 AComIn "Advanced Computing for Innovation", funded by the European Commission in 2012–2016. It is also related to the COST Action IC1307 iV&L Net. The authors are thankful to Imagga company for their comments, recommendations and experimental datasets.

References

Esuli A. and Sebastiani F. 2006. *Sentiwordnet: A publicly available lexical resource for opinion mining*. In Proc. of LREC, Vol. 6, pp. 417-422, 2006.

Beaudoin J. 2007. *Flickr Image Tagging: Patterns Made Visible*. In Bulletin of the American Society for Information Science and Technology (October/November, 2007), 26-29.

Borth D., R. Ji, Tao Chen, T. Breuel and Shih-Fu Chang. 2013. *Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs*. In Proc. of the 21st ACM Int. conference on Multimedia (Barcelona, Spain, October 21-25, 2013). MM'13 . ACM, New York, NY, 223-232. DOI= http://doi.acm.org/10.1145/2502081.2502282

Chen T., Felix X. Yu, Jiawei Chen, Yin Cui, Yan-Ying Chen and Shih-Fu Chang. 2014. *Object-Based Visual Sentiment Concept Analysis and Application*. In Proc. MM'14, 22st ACM Int. Conf. on Multimedia, ACM, NY, 367-376, 2014. DOI= http://doi.acm.org/ 10.1145/2647868.2654935

Hailin J., Jianchao Y., Quanzeng Y. and Jiebo L. 2015. *Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks*. In Proc. of the 29 AAAI Conference on AI (Austin Texas, USA), 381-388, 2015.

Peters I. andWeller K. 2008. *Tag Gardening for Folksonomy Enrichment and Maintenance*. Webology, Vol.5 (3).

Fernández-Tobías I., Cantador I. and Plaza L. 2013. *An Emotion Dimensional Model based on Social Tags: Crossing Folksonomies and Enhancing Recommendations E-Commerce and Web Technologies.*Lecture Notes in Business Information Processing, Vol. 152, 88-100, 2013.

Jia J., S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. 2012. *Can we understand van gogh's mood?: learning to infer affects from images in social networks*. In Proc. of the 20th ACM Int. conference on Multimedia (Nara, Japan). MM'12 . ACM, New York, NY, 857-860. DOI= http://doi.acm.org/10.1145/2393347.2396330

Lesk M. *Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone*. In SIGDOC '86: Proc. 5th Annual International Conference on Systems documentation, pp. 24-26, 1986 ACM, USA. DOI=10.1145/318723.318728

Machajdik J. and Hanbury A. 2010.*Affective image classification using features inspired by psychology and art theory*. In Proc. Int. Conf. on Multimedia (Firenze, Italy, October 25-29, 2010). MM'10 . ACM, New York, NY, 83-92. DOI= http://doi.acm.org/10.1145/1873951.1873965

Platt J. 1998. *Fast training of support vector machines using sequential minimal optimization*. In B. Schoelkopf, C. Burges, and A. Smola (eds.), Advances in Kernel Methods, Support Vector Learning. MIT Press, Cambridge, MA, USA.

Siersdorfer S., Minack E., Fan Deng and J. Hare. 2010. *Analyzing and Predicting Sentiment of Images on the Social Web*. In Proc. Int. Conf. on Multimedia (Firenze, Italy) ACM, NY, 715-718. DOI=http://doi.acm.org/10.1145/1873951.1874060

Witten I. H. 2011. *Data mining: practical machine learning tools and techniques*. – 3rd ed. / Ian H. Witten, Frank Eibe, Mark A. Hall. San Francisco: Morgan Kaufmann.

Yang, Y.,P. Cui, W. Zhu, and S. Yang. 2013. *User interest and social influence based emotion prediction for individuals*. In Proc. of the 21st ACM Int. Conf. on Multimedia (Barcelona, Spain), MM'13. ACM, NY, 785-788, 2013. DOI=http://doi.acm.org/10.1145/2647868.2654935

Yang Y., P. Cui, H. Vicky Zhao, Wenwu Zhu, Yuanyuan Shi and Shiqiang Yang. 2014. *Emotionally Representative Image Discovery for Social Events*. In Proc. Int. Conf. on Multimedia Retrieval (Glasgow, UK), ICMR'14. ACM, NY, 177-184. DOI=http://doi.acm.org/10.1145/2578726.2578749