About Tag Sense Disambiguation in Image Annotation

Olga Kanishcheva, Galia Angelova

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 25A Acad. G. Bonchev Str., 1113 Sofia, Bulgaria

kanichshevaolga@gmail.com, galia@lml.bas.bg

(Extended abstract)

Abstract. In this paper we present an approach for word sense disambiguation (WSD) of image tags from different sources, using knowledge-based methods and semantic measures. We focus on the resolution of lexical ambiguity that arises when a given keyword has several different meanings. The results may be used for machine translation of tags or measuring similarity between images. Our approach combines some knowledge-based methods (Lesk algorithm and Hyponym Heuristics) and semantic measures. Mainly, the paper concentrates on how to combine these methods in order to achieve successful disambiguation. Finally, we present experimental results related to performance evaluation when the technologies are combined to process image tags in practical settings.

Keywords: Word sense disambiguation, WSD, semantic similarity measure, social tagging, lexical semantics, lexical ambiguity, tag sense disambiguation

1 Introduction

Word sense ambiguity (WSA) is a central problem for many established Human Language Technology applications (e.g., machine translation, information extraction, question answering, information retrieval, text classification, and text summarization). This is also the case for the associated subtasks (e.g., reference resolution, acquisition of subcategorization patterns, parsing, and, obviously, semantic interpretation) [1, 2].

WSA concerns also tagging: in blogs, videos, images and digital objects in general; site authors attach keyword descriptions (also called tags) as a category or topic to identify sub-units within their sites. Digital resources with identical tags can be linked together allowing users to search for similar or related content. Many users are interested in multimedia platforms supporting photo archives, due to the prevalence of digital cameras and the growing practice of photo sharing in community-contributed image websites like Flickr and Zooomr. All such services are based on tagging for images and video. Tags also support image retrieval, classification, clusterization etc.

Tags ambiguity leads to inaccurate results. For example (Fig. 1) one cannot distinguish the images about the crane (birds) or the crane (machine), since they share the same tag “crane”. Resolving the ambiguity of tags can help improving the accuracy of machine translation of keyword annotation and image classification.

adfa, p. 1, 2011.

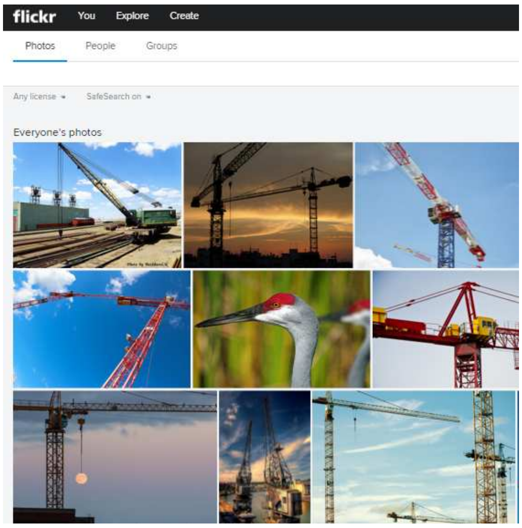

Fig. 1. Results from Flickr for query "crane"

The paper will be organized as follows: Section 2 will discuss related work and summaze different methods for WSD and similarity measures which are applied in our experiments. In Section 3 we'll describe our approach for tag sense disambiguation of annotated images. The results and evaluation using different component configurations will be reported in Section 4. Finally, in Section 5 we conclude the paper.

2 Our Approach to TSD

The task we deal with can be formulated as follows: Given an annotated image that is tagged with ambiguous English keywords, can we identify automatically the correct sense of the ambiguous keywords in the particular annotation tagset? An example of such an image is given in Figure 1; Table 1 contains the list of senses for "*orange*" in WordNet – there are six senses, S_1-S_5 correspond to nouns *(n)* and S_6 is an adjective *(adj)*; the tag "*fruit*" has 5 senses and "*food*" has 3 senses.

The image on the left has tags {"***orange***", "***fruit***", "***food***"}. Our task is to find which sense is referred to in the annotation of this particular image. Or, more practically, we want to design algorithms that suggest with certain accuracy the most probable sense that is meant in the image annotation. This is uneasy task as no textual context is available in image annotation.

Fig. 1: Annotation with 3 ambiguous tags

adfa, p. 2, 2011.

Table 1. The set of WordNet senses for tag "orange"	
S_1	S: *(n) orange (round yellow to orange fruit of any of several citrus trees)*
S_2	S: *(n) orange,* orangeness *(orange color or pigment; any of a range of colors between red and yellow)*
S_3	S: *(n) orange,* orange tree *(any citrus tree bearing oranges)*
S_4	S: *(n) orange (any pigment producing the orange color)*
S_5	S: *(n) Orange,* Orange River *(a river in South Africa that flows generally westward to the Atlantic Ocean)*
S_6	S: *(adj) orange,* orangish *(of the color between red and yellow; similar to the color of a ripe orange)*

The material. In our experiment we used 4,221 "professional" images from Professional image marketplace (http://www.stockpodium.com/) and 5,118 "social" images from Flickr (https://www.flickr.com/). Each image has also tags provided by the Imagga's auto-tagging platform (http://imagga.com/) that are appended to the original annotation. Table 2 presents features of the test datasets by mapping tags to WordNet.

Table 2. Some statistical information about the test datasets

	Number of files	Number of tags	Tags with 1 sense in WordNet	Tags with many senses in WordNet	Tags not included in WordNet
Professional images	4,221	292,418	35,167	250,127	7,127
Social images	5,118	277,065	31,649	237,135	8,281

The test sets contain about 23% more social images however the comparative analysis of the figures in Table 2 shows that 1) the total numbers of tags for these two groups do not really differ, e.g. for tags with a single sense and with ambiguous ones we even find more tags in the "professional" data set; 2) the number of tags which are not included in WordNet is significant for both groups (7,127 for the professional images; 8,281 for social images).

Finding 1 can be explained as follows: social images are often low-quality photos without clearly focused objects which is somewhat problematic for the auto-tagging platform. To understand more deeply finding 2, we analyzed the "unknown" tags. These are such causes as: a) phrases (the auto-tagging program assigns phrasal annotations and many of them are outside WordNet, e.g. "*one person*"); b) some tags are not included in WordNet because they are rather new words (for example, "*clipart*").

Our approach. We combine several WSD methods and similarity measures, using their data structures and ideas:

- *WUP measure*, a WordNet-based similarity score between senses [4];
- *Gloss overlap* (Lesk algorithm [5]): all the glosses from the synsets of an ambiguous word are checked looking for those that contain words from the context; each match increases the synset count by one;

- *Using hypernym/hyponym heuristics* – i.e. the glosses of the hypernym/hyponym synsets of the ambiguous word; the overlap scoring mechanism is also parametrized and can be adjusted to take into account gloss length (i.e. normalization) or to include function words [6].

Given an image with ambiguous tags $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, .., $\boldsymbol{t}_k$, for all senses of t_1 - $S(t_1)$ - we build a matrix M_1, where the rows are juxtaposed to the $\boldsymbol{t}_1$ senses and the columns are juxtaposed to all senses of the neighboring tags $\boldsymbol{t}_2$, .., $\boldsymbol{t}_k$. Each cell of M_1 contains the WUP similarity score between the corresponding senses. For each row, we sum up the similarity scores inserted in the row's cells. These sums are called "sense weight". The paper contains an algorithm how to calculate the matrix M_i. This is done for each ambiguous tag separately. After building the matrices M, we calculate similarity among vectors using two classical WSD algorithms: Lesk algorithm and Hyponym Heuristics. Quantitative evaluation showed that our method can effectively disambiguate tags. After manual evaluation, *"Correct"* are: 96.80% of the selected senses for professional images and 91.72% for social images. In the category *"Not correct, but related"* the evaluators include 1.4% for professional images and 5.59% for social images. There is relatively small amount of *"Neither correct, no related"*: 2.17% for professional images and 3.66% for social images.

In future we plane to use our approach for human tags for images from different sources. Also we want to use not only the WordNet ontology, but and modern and complexity resources such as BabelNet (BabelNet is a multilingual lexicalized semantic network and ontology) and pyWSD (it is python implementations of WSD technologies). We hope that these resources could be help with a problem of modern tags.

This research work is partially supported by the FP7 grant 316087 AComIn "Advanced Computing for Innovation", funded by the European Commission. It is also related to the COST Action IC1307 "Integrating Vision and Language (iV&L Net): Combining Computer Vision and Language Processing for Advanced Search, Retrieval, Annotation and Description of Visual Data". The authors are thankful to Imagga company for their comments, recommendations and experimental datasets.

References

1 Montoyo A., A. Su´arez, G. Rigau and M. Palomar. 2005. *Combining Knowledge- and Corpus-based Word-Sense-Disambiguation Methods*. Journal of Artificial Intelligence Research 23, 299-330.

2 Navigli R. 2009. *Word Sense Disambiguation: A Survey*. ACM Computing Surveys, Vol. 41, No. 2, Article 10.

3 WordNet, http://wordnetweb.princeton.edu

4 Wu Z. and M. Palmer. 1994. *Verb semantics and lexical selection*. Las Cruces, Mexico, 1994, 133-138.

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

6 Navigli R. 2009. *Word Sense Disambiguation: A Survey*. ACM Computing Surveys, Vol. 41, No. 2, Article 10.

adfa, p. 4, 2011.