Engineering Applications of Artificial Intelligence 51 (2016) 150–162

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Online training of concept detectors for image retrieval using streaming clickthrough data

Ioannis Sarafis *, Christos Diou, Anastasios Delopoulos

Multimedia Understanding Group, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Greece

ARTICLE INFO

Available online 29 January 2016

Keywords:
Clickthrough data
Online learning
Image retrieval
Label noise
Fuzzy SVM
LASVM

ABSTRACT

Clickthrough data from image search engines provide a massive and continuously generated source of user feedback that can be used to model how the search engine users perceive the visual content. Image clickthrough data have been successfully used to build concept detectors without any manual annotation effort, although the generated annotations suffer from labeling errors. Previous research efforts therefore focused on modeling the sample uncertainty in order to improve concept detector effectiveness. In this paper, we study the problem in an online learning setting using streaming clickthrough data where each click is treated seperately when it becomes available; the concept detector model is therefore continuously updated without batch retraining. We argue that sample uncertainty can be incorporated in the online learning setting by exploiting the repetitions of incoming clicks at the classifier level, where these act as an implicit importance weighting mechanism. For online concept detector training we use the LASVM algorithm. The inferred weighting approximates the solution of batch trained concept detectors using weighted SVM variants that are known to achieve improved performance and high robustness to noise compared to the standard SVM. Furthermore, we evaluate methods for selecting negative samples using a small number of candidates sampled locally from the incoming stream of clicks. The selection criteria aim at drastically improving the performance and the convergence speed of the online concept detectors. To validate our arguments we conduct experiments for 30 concepts on the Clickture-Lite dataset. The experimental results demonstrate that: (a) the proposed online approach produces effective and noise resilient concept detectors that can take advantage of streaming clickthrough data and achieve performance that is equivalent to Fuzzy SVM concept detectors with sample weights and 78.6% improved compared to standard SVM concept detectors; and (b) the selection criteria speed up convergence and improve effectiveness compared to random negative sampling even for a small number of available clicks (up to 134% after 100 clicks).

1. Introduction

Every day millions of images are uploaded on the web; for example Flickr alone is reported to receive over 3 million image uploads daily. At such rates, the availability of new content implies that new visual elements constantly emerge, including new visual aspects of existing terms or concepts. Currently, modern image search engines rely on clickthrough data, in order to continuously update their text- or metadata-based retrieval models (Park et al., 2015; Yi et al., 2014; Liu et al., 2010). Clickthrough data, consisting of queries and the number of times each image was clicked for the query, act as a source of implicit feedback which is continuously provided by users and is effortlessly collected by the retrieval mechanism to identify and adapt to user information needs.

In the case of image retrieval, however, text-based retrieval models have significant limitations (Datta et al., 2008). For example, text metadata are not always available, and are almost always incomplete (i.e., they do not fully describe the visual content of the image). In addition, disambiguation of different visual aspects of the same term is hard to achieve using text alone. Concept-based retrieval approaches (Snoek and Worring, 2009; Yildizer et al., 2012; Wang et al., 2014), on the other hand, have demonstrated high effectiveness and show promise as a complementary approach for image retrieval based on the visual content. It therefore makes sense, to extend the use of the provided feedback (in the form of clicks) not only for the adaptation of retrieval models, but for the creation and continuous update of concept detector models as well. Clickthrough data can be used to

* Corresponding author.
E-mail addresses: sarafis@mug.ee.auth.gr (I. Sarafis), diou@mug.ee.auth.gr (C. Diou), adelo@eng.auth.gr (A. Delopoulos).

http://dx.doi.org/10.1016/j.engappai.2016.01.017
0952-1976/

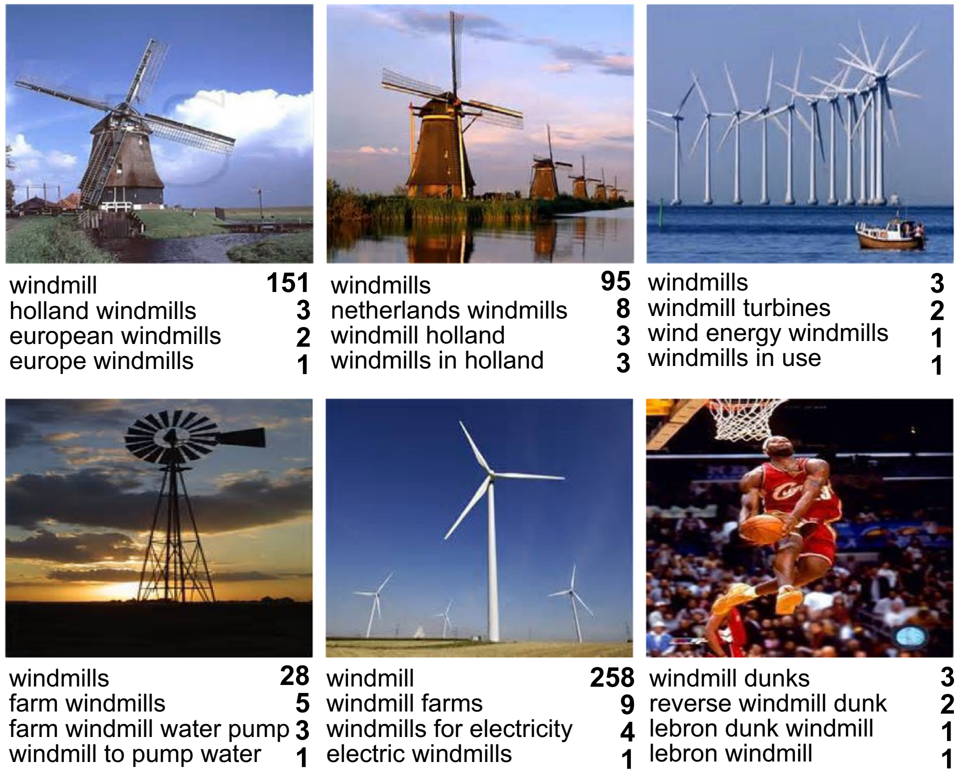

Fig. 1. Example of clickthrough data with clicks containing the term "windmill".

automatically generate annotated sets for training concept detectors. Fig. 1 shows examples of images clicked in respond to queries containing the term "windmill" from users of the Bing image search engine (Clickture-Lite dataset (Hua et al., 2013)). It is obvious that clickthrough data can provide valuable information but can also lead us to many false conclusions. Thus, the use of clickthrough data to automatically generate concept detector training sets introduces a severe *uncertainty* (i.e., label noise) for the candidate training samples (Tsikrika et al., 2009).

The uncertainty introduced by the use of clickthrough data can be quantified and taken into account during classifier training. In our previous work (Sarafis et al., 2014a) we showed that it is possible to quantify the label noise of candidate positive samples using Information Retrieval (IR) language models and then introduce it at the classifier training through appropriate weights. Experiments on a dataset collected from the search logs of the professional image search engine of Belga News Agency[1] showed that concept detectors trained using weighted SVM variants outperformed standard SVM-based detectors and, most interestingly, the performance gains were significantly greater for larger levels of label noise. We further extended our work in Sarafis et al. (2014b, 2015) with techniques for automatic training set and sample weight generation. Three IR models were used to select the positive samples along with a relevance score: Vector Space Models (Salton et al., 1975), BM25 (Robertson, 2009), and Language Models (Hiemstra, 1998). The relevance scores were used to calculate appropriate weights for three, batch training, weighted SVM variants: Fuzzy SVM (Lin and Wang, 2002), Power SVM (Yu and Power, 2012), and Bilateral-weighted Fuzzy SVM (Wang et al., 2005). The combinations of IR models and SVM algorithms were evaluated for 40 concepts on Clickture-Lite dataset and the experimental results showed that the Fuzzy SVM weighting scheme is the most appropriate for this task. The combination of Fuzzy SVM with sample weights calculated from Vector Space Models (VSM) demonstrated a performance improvement of 110% relatively to standard SVM concept detectors. Moreover, the Fuzzy SVM concept detectors demonstrated high robustness towards various levels of artificially controlled label noise in the automatically constructed training sets.

Based on these results, it becomes evident that modeling individual sample uncertainty using the number and the variety of the user clicks for an image leads to significant performance improvements. However, in order to assess the importance of each training sample, the previously proposed approaches regarded clickthrough data as a static collection of search logs ignoring the fact that the data are not known in advance and that new user feedback is constantly produced. Meanwhile, batch training algorithms may suffer from memory limitations due to training set sizes and can only be updated through a retraining procedure that is generally more computationally expensive and not efficient for large scale problems (Bottou, 2010; Kivinen et al., 2004).

On the other hand, treating clickthrough data as streams has many advantages. First, online methods are better suited for large scale machine learning problems since there is no need for the training sets to be loaded in the computer memory all at once. Additionally, search engines that employ online concept detector updating are able to constantly evolve according to user behavior or new user information needs. For example, one can start to train detectors for trending concepts as soon as they are identified. Moreover, the trained models can be updated and adapt to concept changes guided by the users' implicit feedback.

In this paper, we propose and evaluate methods for online training of concept detectors that are able to evolve and improve while taking into account the inherent label uncertainty of the automatically generated annotations. The premise of the method is that the sample repetitions, as dictated by the number of clicks, can lead to an *indirect importance weighting* in the feature space.

[1] The data were collected in the context of Vitalas project http://vitalas.ercim.eu/.