Computers & Graphics 52 (2015) 93–105

Contents lists available at ScienceDirect

Computers & Graphics

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

Technical Section

# Active learning for sketch recognition ☆

Erelcan Yanık *, Tevfik Metin Sezgin

Koç University, College of Engineering, Istanbul, Turkey

ARTICLE INFO

*Article history:*
Received 15 April 2015
Received in revised form
30 July 2015
Accepted 30 July 2015
Available online 7 August 2015

*Keywords:*
Active learning
Sketch recognition
Empirical analysis
Factor analysis
ANOVA

ABSTRACT

The increasing availability of pen-based tablets, and pen-based interfaces opened the avenue for computer graphics applications that can utilize sketch recognition technologies for natural interaction. This has led to an increasing interest in sketch recognition algorithms within the computer graphics community. However, a key problem getting in the way of building accurate sketch recognizers has been the necessity of creating large amounts of annotated training data. Several authors have attempted to address this issue by creating synthetic data, or by building easy-to-use annotation tools. In this paper, we take a different approach, and demonstrate that the active learning technology can be used to reduce the amount of manual annotation required to achieve a target recognition accuracy. In particular, we show that by annotating few, but carefully selected examples, we can surpass accuracies achievable with equal number of arbitrarily selected examples. This work is the first comprehensive study on the use of active learning for sketch recognition. We present results of extensive analyses and show that the utility of active learning depends on a number of practical factors that require careful consideration. These factors include the choices of informativeness measures, batch selection strategies, seed size, and domain-specific factors such as feature representation and the choice of database. Our results imply that the Margin based informativeness measure consistently outperforms other measures. We also show that active learning brings definitive advantages in challenging databases when accompanied with powerful feature representations.



## 1. Introduction

Sketch recognition is an enabling technology that lies at the foundation of many computer graphics applications, including educational applications [1,2], graphics applications for design [3–5], shape retrieval [6], and animation [7]. A widely acknowledged problem in building accurate sketch recognition systems is the labor-intensive nature of obtaining large amounts of labeled data [8]. In this paper, we demonstrate the utility of the active learning technology in reducing the amount of manual annotation required to achieve a target recognition accuracy. The results and the approach presented in this paper provide valuable insights to the practitioners of sketch recognition as well as the broader community of computer graphics practitioners who rely on machine learning in their applications.

The sketch recognition community has attempted to address the data labeling problem by synthesizing artificial training examples from few labeled examples [8], by building custom interfaces for labeling data [9–12] or by applying automated labeling supported with a partially trained recognizer [13]. Each case requires human annotators to label data without being particularly selective about which samples are labeled. We show that, using active learning, it is possible to prioritize the labeling process in a way that allows one to build more accurate classifiers with fewer labeled instances, hence reduce the annotation effort.

Active learning is a machine learning strategy that aims to reduce the labeling effort by selecting the most informative samples from a pool of unlabeled data. The basic premise of active learning is that some training examples carry more information than others. Hence, if we can identify them among the unlabeled examples, and have them labeled by a human annotator, we can potentially converge to higher accuracies with substantially less human annotation effort.

Active learning process is initialized by training a classifier with a few labeled samples, the so-called "seed set". The learning process continues in rounds until a target validation accuracy is achieved or until we run out of resources (e.g. time or computational resources). In each round, we train a classifier with the available labeled data, and use it to classify the unlabeled examples. We then use the scores assigned to the unlabeled samples to

☆This article was recommended for publication by Beryl Plimmer

* Corresponding author.
*E-mail addresses:* eyanik13@alm.ku.edu.tr (E. Yanık), mtsezgin@ku.edu.tr (T.M. Sezgin).
*URL:* http://www.iui.ku.edu.tr (T.M. Sezgin).

http://dx.doi.org/10.1016/j.cag.2015.07.023

select the subset of most useful samples for subsequent labeling. The round ends by adding the newly labeled data to the training set and re-training the model.

Despite its theoretical appeal, recent empirical results show that active learning does not always yield the expected benefits in practical real world problem settings [14]. For example, Schein and Ungar report inconsistent and negative results for active learning [15]. Likewise, Gasperin reports that none of the experimented active learning methods reached a remarkable performance although they converge on different sets of training examples from each other [16]. Guo and Schuurmans also point out that active learning methods perform poorly with respect to random learning, which is the strategy of selecting samples randomly from a pool of unlabeled examples [17]. Therefore, there is a practical and real need for analyzing the empirical performance of active learning in various settings in order to understand if active learning is effective at all, and if so under which conditions. In this paper, we present such an analysis for the sketch recognition domain to identify the set of practical issues one should be aware of when using active learning, and investigate how these issues affect active learning performance.

Specifically, we investigate the performance of active learning under combinations of a large variety of informativeness measures and batch selection strategies, as well as factors such as feature representation, database and seed set size for sketch recognition. Our analysis results constitute a detailed and practical guide for active learning users for sketch recognition and provide valuable insights for machine learning practitioners in the computer graphics community. Our main contributions can be summarized as follows:

- We present a set of carefully designed experiments and a battery of accompanying statistical tests, which will serve as a roadmap to follow for practitioners of active learning who wish to perform factor analysis.
- We present the first extensive empirical analysis on active learning for sketch recognition, and provide a detailed discussion of the analysis results.
- We determine the best performing and reliable informativeness measure for sketch recognition.
- We show that starting with a large seed set yields better active learning performance for the single classifier approach.
- We show that the use of active learning brings definitive advantages in challenging databases when accompanied with powerful feature representations.

This paper is organized as follows: First, we introduce informativeness measures and batch selection strategies that are included in our analysis. In Section 3, we first describe the databases and the feature representations used in our experiments, then describe the details of our experimental design. In Section 4, we describe the deficiency measure employed in our analysis and then present the analysis methodology. We present the analysis results with a discussion in Section 5. Finally, we conclude with related work and a summary of future research directions.

## 2. Active learning methods

There are two essential steps in active learning: measuring informativeness of unlabeled samples and selecting batches of collectively informative samples which are mutually non-redundant. In this section, we describe informativeness measures and batch selection strategies that are used in our experiments.

### 2.1. Informativeness measures

There are two main approaches for measuring informativeness: the *single classifier* approach and the *query by committee* (QBC) approach. Measures of informativeness are based on the rationale that samples that a classifier cannot confidently classify, or a group of classifiers disagree on can potentially supply more information when labeled. We list the informativeness measures included in our experiments in Table 1. Four of the measures follow the single classifier approach, in which decisions are based on a single classifier's prediction on a sample. The other four measures follow the query by committee (QBC) approach, in which the disagreement of the committee members on the label of a sample is used to derive informativeness.

### 2.2. Batch selection strategies

Active learning requires classifiers to be retrained as more labeled data gets added to the training set. Since training is costly, newly labeled examples are usually added in batches, rather than one by one. Although adding samples in batches reduces computational requirements, it bears the risk of adding samples which carry mutually redundant information. In particular, two samples which are extremely informative when taken individually may actually contain similar and redundant information, so including them both in the training data may not yield extra advantage over having just one or the other. Hence, we should avoid sets containing mutually redundant samples. Several batch selection strategies have been proposed in the literature to avoid this problem, and we included four of them in our experiments.

Our empirical analysis includes the following batch selection strategies: Default selection, Global-FV strategy, Global-PE strategy

**Table 1**
Brief description of the informativeness measures used in our analysis.

| | Informativeness measures | A sample is considered informative when: |
|---|---|---|
| **Single classifer approach** | Entropy based Selection[a] [18,19] | The entropy is high on class probabilities of a sample. |
| | Least Confident based Selection [18,20] | The most likely class probability of a sample has a low value. |
| | Margin based Selection[a] [18] | The difference of the most and the second most likely class probabilities of a sample has a low value. |
| | Körner–Wrobel Selection[a] [21] | The Körner–Wrobel value computed for the sample is low. It is a combination of Least Confident and Margin based selection strategies. |
| **Query by committee approach** | Kullback Leibler Divergence based Selection [21,22] | KL-Divergence among the committee on a sample is high. |
| | Jensen Shannon Divergence based Selection [21,23] | JS-Divergence among the committee on a sample is high. |
| | Vote Entropy based Selection [21,24] | The entropy of the class label votes of the committee is high. |
| | Weighted Vote Entropy based Selection [21] | The weighted entropy of the class label votes of the committee is high. |

[a] The method has implementation also for the query by committee approach, in the literature, but we only include the single classifier version.