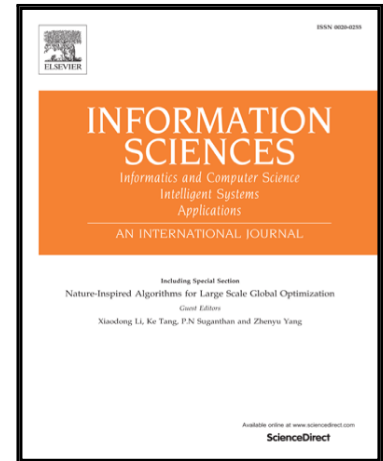


Accepted Manuscript

Query-by-committee improvement with diversity and density in batch active learning

Seho Kee, Enrique del Castillo, George Runger

PII: S0020-0255(18)30370-0
DOI: [10.1016/j.ins.2018.05.014](https://doi.org/10.1016/j.ins.2018.05.014)
Reference: INS 13642



To appear in: *Information Sciences*

Received date: 1 September 2017
Revised date: 27 November 2017
Accepted date: 2 May 2018

Please cite this article as: Seho Kee, Enrique del Castillo, George Runger, Query-by-committee improvement with diversity and density in batch active learning, *Information Sciences* (2018), doi: [10.1016/j.ins.2018.05.014](https://doi.org/10.1016/j.ins.2018.05.014)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Query-by-committee improvement with diversity and density in batch active learning

Seho Kee^{a,*}, Enrique del Castillo^b, George Runger^c

^a*Department of Industrial Engineering, Arizona State University, AZ*

^b*Department of Industrial & Manufacturing Engineering, Pennsylvania State University, PA*

^c*Department of Biomedical Informatics, Arizona State University, AZ*

Abstract

Active learning has gained attention as a method to expedite the learning curve of classifiers. To this end, uncertainty sampling is a widely adopted strategy that selects instances closer to the decision boundary. However, uncertainty sampling alone may not be sufficient in batch active learning due to the redundancy of instances and its susceptibility to outliers. In this study, we utilize query-by-committee (QBC) for uncertainty and demonstrate that its performance can be improved by introducing diversity and density in instance utility. Test results show that uncertainty sampling by QBC can be significantly improved with diversity and density incorporated in instance selection. Furthermore, we investigate several distance measures for use in diversity and density and show that random forest dissimilarity can be an effective distance measure in batch active learning. The effects of the characteristics of the data on the results are also analyzed.

Keywords: batch active learning, density, diversity, query-by-committee, random forest

1. Introduction

Active learning (AL) [44, 38, 18] has gained popularity in pursuit of efficient training of classifiers under a limited budget for labeling instances. AL algorithms actively select instances that show more potential contribution to improving the model performance and label them first. Because the objective of AL is to achieve higher performance of classifiers more quickly, instances chosen to be labeled (the so called query instances) selected using uncertainty sampling [34] are likely to be near the decision boundary.

However, uncertainty sampling alone may not be sufficient in batch active learning (BAL). Unlike serial AL where labels are applied one instance at a time, BAL queries multiple instances for labels. In BAL, one should consider the smoothness assumption that neighboring instances are likely to share the prediction results. If one forms a query only from the most uncertain instances, such a query is susceptible to redundancy of instances as neighbors of the most uncertain instance will also likely show high uncertainty. To prevent redundant instances in queries, diversity, which increases the pairwise distances among query instances, has been employed in instance selection to ensure query instances are not too close [10, 30, 55].

*Corresponding author

Email addresses: seho.ke@asu.edu (Seho Kee), exd13@psu.edu (Enrique del Castillo), george.runger@asu.edu (George Runger)

Download English Version:

<https://daneshyari.com/en/article/6856394>

Download Persian Version:

<https://daneshyari.com/article/6856394>

[Daneshyari.com](https://daneshyari.com)