ON SMALL SAMPLE PROBLEMS
IN ACTIVE LEARNING

JIANG JUN

DOCTOR OF PHILOSOPHY
CITY UNIVERSITY OF HONG KONG
MAY 2010
CITY UNIVERSITY OF HONG KONG

On Small Sample Problems in Active Learning

Submitted to
Department of Computer Science
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

by

Jiang Jun

May 2010
Abstract

In the field of machine learning, the paradigm of active learning is regarded as most similar to the mode of natural learning in which the learner asks teacher questions and then obtains the knowledge from the answer given. Active learning has been typically combined with supervised learning. In this thesis, the author addresses the small sample problems in active learning that the small number of labeled data may lead to poor classifiers during the active learning process. This important aspect of active learning has been largely ignored by researchers. By combining active learning with Support Vector Machine (SVM) to solve the small sample problems, three schemes are utilized in this thesis. The following paragraphs overview the contributions of this thesis.

The first scheme aims to improve the quality of the selected data, i.e., choose the data that convey more information to the learner. In traditional active learning, the similarity measurement between two data only relies on themselves. It may lead to low quality of the selected data, as they do not reflect information on the whole dataset. In order to select data with high quality, the motivation is to incorporate unlabeled data to improve the similarity measurement between the data. In order to support the proposed methods using this scheme, a criterion that the selected data should maximize the class entropy is suggested from the viewpoint of information theory. Based on this principle, a general data selection framework is inferred that the selected data should be as dissimilar from each other as possible. Then under the inferred framework, three similarity measurements and three novel active learning methods are proposed and derived respectively.

For the first proposed method, the weight of an edge in a graph is directly employed as the similarity between two data points connected by this edge. During the process of active learning, a single-side strategy is proposed to modify the weights of the graph. This leads to a dynamic similarity measurement which is able to unfold the similarity between data better than fixed strategies, such as the well-known Angle Diversity incorporated Active Learning with SVM (AD-ALSVM). For the second proposed method, the data are first mapped into a manifold
space which is extracted from the graph Laplacian, and then the similarities are calculated based on the mapped data. Such a strategy is useful especially when inherent manifold structures exist in the dataset. In addition to the manifold structures, geometry information can also be extracted from the graph Laplacian. The extracted geometry information can be wrapped into a kernel function for classification using SVM. The modified kernel matrix can be employed as the similarity matrix. It leads to the third proposed method in which the similarity is adaptive to the geometry structures hidden in the data selection pool.

The second common scheme is to enlarge the labeled training set by adding pseudo-labeled data which are created from the set of unlabeled data. Different from other methods, a weak label propagation method is proposed to create pseudo-labeled data, and they are combined into the labeled training set to form an enlarged training set. Instead of using a commonly deployed a sample-weighted SVM, a weighted margin SVM is applied to train the classifiers during the active learning process.

The third scheme of overcoming the small sample problems is to reduce the probability of forming a biased version space during the active learning process. One negative influence resulting from the small size of the labeled data is that the version space formed on these labeled data is prone to being biased, i.e., it does not include the target hypothesis. However, almost all the active learning methods explicitly or implicitly assume that the current version space includes the target hypothesis, such as the well-known Simple Distance-based Active Learning with SVM (SD-ALSVM), AD-ALSVM and query-by-committee. Therefore they focus on the exploitation ability of finding the target hypothesis by minimizing the current hypothesis space, while neglecting the exploration ability of finding the unexploited hypothesis space in which the target hypothesis may exist. Unlike the traditional deterministic active learning, in this thesis, simulated annealing active learning is proposed, in which both exploitation and exploration ability are invoked in the learning process, meanwhile a method is proposed to automatically regulate between these two strategies. All the proposed methods are shown to be efficient in solving the small problems in active learning in our experiments.

Finally, an adaptive active learning approach is designed by applying one of the techniques mentioned above to the phosphorylation prediction system to overcome the
problems associated with sample annotation. The experiments show that this adaptive method is able to significantly reduce the number of annotated samples and it is more efficient than the AD-ALSVM which was utilized in our previous work, where active learning approaches were applied in the phosphorylation prediction system for the first time. Hence, it leads to an effective tool to assist biologists to select the most informative sample to annotate in a large protein database.
# Table of Contents

Abstract ....................................................................................................................................................... i  
Acknowledgements ........................................................................................................................................ iv  
Table of Contents ........................................................................................................................................ v  
List of Figures .................................................................................................................................................. viii  
List of Tables ................................................................................................................................................... xi  
Index of Abbreviations ...................................................................................................................................... xii  
Chapter 1 Introduction ..................................................................................................................................... 1  
  1.1. Active Learning ...................................................................................................................................... 1  
  1.2. Version Space Reduction based Active Learning with SVM ................................................................. 4  
    1.2.1. Theoretical Foundation ...................................................................................................................... 4  
    1.2.2. Single Mode Active Learning with SVM ........................................................................................... 5  
    1.2.3. Batch Mode Active Learning with SVM ........................................................................................... 8  
  1.3. Other Approaches of Active Learning with SVM .................................................................................. 10  
    1.3.1. Statistical based Active Learning ....................................................................................................... 10  
    1.3.2. Min-max based Active Learning with SVM ...................................................................................... 10  
    1.3.3. Multi-view Based Active Learning with SVM .................................................................................. 11  
  1.4. Mixed Active Learning ............................................................................................................................ 11  
    1.4.1. Hybrid Active Learning ...................................................................................................................... 11  
    1.4.2. Semi-supervised Active Learning ...................................................................................................... 12  
  1.5. Thesis Overview & Contributions .......................................................................................................... 13  
  1.6. Thesis Organization ................................................................................................................................. 16  
Chapter 2 Class Entropy Based Batch Mode Active Learning with SVM ....................................................... 18  
  2.1. Introduction .......................................................................................................................................... 18  
    2.1.1. Motivation ........................................................................................................................................... 18  
    2.1.2. Related Works .................................................................................................................................... 19  
    2.1.3. Main Contribution ............................................................................................................................... 20  
  2.2. Class Entropy based Batch Mode Data Selection Criterion ....................................................................... 21  
    2.2.1. Preliminary ......................................................................................................................................... 21  
    2.2.2. Algorithm Development and its Theoretical Foundation ................................................................. 23  
    2.2.3. Practical Ramification of the Algorithm .............................................................................................. 30  
  2.3. Comparison to Other Batch Mode Active Learning Algorithms ............................................................... 35  
    2.3.1. Comparison to Angle Diversity Incorporated Active Learning with SVM ......................................... 37  
    2.3.1. Comparison to Min-Max based Active Learning with SVM ............................................................... 38  
  2.4. Three Methods based on the Proposed Framework ................................................................................ 39  
Chapter 3 Dynamic Batch Mode Active Learning with SVM ........................................................................... 40  
  3.1. Introduction .......................................................................................................................................... 40  
  3.2. The Proposed Algorithm ........................................................................................................................ 42  
    3.2.1. Motivation ........................................................................................................................................... 42
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1-1</td>
<td>Comparison between active learning and other machine learning methods</td>
<td>2</td>
</tr>
<tr>
<td>1.1-2</td>
<td>Framework of active learning</td>
<td>3</td>
</tr>
<tr>
<td>1.2-1</td>
<td>Illustration of simple distance-based ALSVM</td>
<td>6</td>
</tr>
<tr>
<td>1.2-2</td>
<td>One example of simple batch querying with “a”, “b” and “c” samples with pure SD-ALSVM</td>
<td>9</td>
</tr>
<tr>
<td>1.2-3</td>
<td>One example of batch querying with “a”, “b” and “c” samples by incorporating angle diversity into SD-ALSVM</td>
<td>9</td>
</tr>
<tr>
<td>1.6-1</td>
<td>Organization of the thesis</td>
<td>17</td>
</tr>
<tr>
<td>2.1-1</td>
<td>An example of active learning with SVM</td>
<td>19</td>
</tr>
<tr>
<td>2.2-1</td>
<td>Framework of class entropy based batch mode active learning</td>
<td>29</td>
</tr>
<tr>
<td>2.2-2</td>
<td>Framework of sequential class entropy based batch mode active learning</td>
<td>30</td>
</tr>
<tr>
<td>2.2-3</td>
<td>Framework of average-based sequential class entropy based batch mode active learning</td>
<td>31</td>
</tr>
<tr>
<td>2.2-4</td>
<td>Framework of maximum-based sequential class entropy based batch mode active learning</td>
<td>33</td>
</tr>
<tr>
<td>2.2-5</td>
<td>Framework of relaxed class entropy based batch mode active learning</td>
<td>34</td>
</tr>
<tr>
<td>2.3-1</td>
<td>Framework of angle diversity incorporated active learning with SVM</td>
<td>36</td>
</tr>
<tr>
<td>2.3-2</td>
<td>Framework of relaxed class entropy based batch mode active learning</td>
<td>38</td>
</tr>
<tr>
<td>3.2-1</td>
<td>Illustration of the relationship of the data points</td>
<td>42</td>
</tr>
<tr>
<td>3.2-2</td>
<td>A set of the data points</td>
<td>44</td>
</tr>
<tr>
<td>3.2-3</td>
<td>Illustration of the negative effects brought by the must-link constraints</td>
<td>47</td>
</tr>
<tr>
<td>3.2-4</td>
<td>Framework of dynamic batch mode active learning</td>
<td>49</td>
</tr>
<tr>
<td>3.3-1</td>
<td>Experiment I results (on benchmark results)</td>
<td>52</td>
</tr>
<tr>
<td>3.3-2</td>
<td>Experiment I results (on benchmark results)</td>
<td>53</td>
</tr>
<tr>
<td>3.3-3</td>
<td>Illustration of Corel image database</td>
<td>55</td>
</tr>
<tr>
<td>3.3-4</td>
<td>Experiment II results (image retrieval on Corel database)</td>
<td>58</td>
</tr>
</tbody>
</table>
Figure 3.3-5: Experiment II results (image retrieval on Corel database)............................... 59

Figure 4.2-1: Comparison between the representations in the original space and manifold space. 63

Figure 4.2-2: The framework of constructing manifold space .................................................... 65

Figure 4.2-3: Framework of manifold diversity incorporated active learning ............................. 67

Figure 4.2-4: Framework of angle diversity incorporated active learning with SVM .................. 68

Figure 4.3-1: Experiment I results (on benchmark results)...................................................... 74

Figure 4.3-2: Experiment II results (image retrieval on Corel database).................................. 76

Figure 5.2-1: Illustration of the motivation of the proposed algorithm....................................... 80

Figure 5.2-2: A toy data set in which one triangle structure is hidden...................................... 84

Figure 5.3-1: Comparison of the DMCE-ALSVM, MCE-ALSVM, Min-Max, and AD-ALSVM method (on benchmark results)................................................................. 89

Figure 5.3-2: Comparison of the DACE-ALSVM, ACE-ALSVM, Min-Max, and AD-ALSVM method (on benchmark results)................................................................. 90

Figure 5.3-3: Comparison of the DMCE-ALSVM, MCE-ALSVM, Min-Max, and AD-ALSVM method (image retrieval on Corel database)............................................... 93

Figure 5.3-4: Comparison of the DACE-ALSVM, ACE-ALSVM, Min-Max, and AD-ALSVM method (image retrieval on Corel database)............................................... 94

Figure 6.2-1: Illustration of the goal of the proposed method.................................................... 98

Figure 6.2-2: Illustration of proposed framework........................................................................ 99

Figure 6.2-3: Illustration of weight margin SVM....................................................................... 102

Figure 6.2-4: Procedures of linear neighborhood propagation.................................................. 105

Figure 6.2-5: Procedures of weak neighborhood propagation.................................................. 109

Figure 6.2-6: Calculation of the confidence of the pseudo-labeled data in the linear neighborhood propagation method................................................................. 110

Figure 6.2-7: Calculation of the confidence of the pseudo-labeled data in the weak label propagation method.............................................................................................. 111

Figure 6.2-8: Comparison between the data selection criterion with and without incorporating degree............................................................................................................. 113

Figure 7.2-1: Analogy between active learning and optimization problem................................. 128
Figure 7.2-2: The pseudo-code for simulated annealing ................................................................. 129
Figure 7.2-3: The underlying idea of simulated annealing ............................................................ 130
Figure 7.2-4: The logical order of some data selection criteria .................................................. 132
Figure 7.2-5: The procedures of the adaptive data selection criterion ..................................... 133
Figure 7.2-6: The variation of the angle between the decision hyperplanes corresponding to the 
two successive classifiers ........................................................................................................... 134
Figure 7.2-7: The variation of the angle between the decision hyperplanes corresponding to the 
successive two classifiers ....................................................................................................... 136
Figure 7.2-8: Framework of simulated annealing simple distance based active learning with 
SVM (SASD-ALSVM) .................................................................................................................. 137
Figure 7.2-9: Framework of simulated annealing angle diversity incorporated active learning 
with SVM (SAAD-ALSVM) ....................................................................................................... 138
Figure 7.3-1: Illustration of the Tencluster dataset ...................................................................... 139
Figure 7.3-2: Results of the experiments on the artificial dataset .............................................. 140
Figure 7.3-3: Results of comparison between SASD-ALSVM and SD-ALSVM ....................... 145
Figure 7.3-4: Results of comparison between SAAD-ALSVM and AD-ALSVM ....................... 147
Figure 8.2-1: Framework of maximum-based sequential class entropy based batch mode active 
learning with SVM ................................................................................................................... 154
Figure 8.3-1: The comparison of A-ALSVM, AD-ALSVM and Random SVM method ............ 160
List of Tables

Table 3.3-1: Description of datasets .................................................................................................................. 51
Table 3.3-2: The average precision of the top 100 returned images ................................................................. 57
Table 4.3-1: Description of datasets .................................................................................................................. 71
Table 5.3-1: Description of datasets .................................................................................................................. 88
Table 6.3-1: The setting of PD-ALSVM ............................................................................................................. 116
Table 6.3-2: Description of datasets .................................................................................................................. 117
Table 7.2-1: Analogy between active learning and optimization problem ......................................................... 128
Table 7.3-1: Description of benchmark datasets ............................................................................................... 141
Table 8.3-1: Six sub-datasets employed in experiments ...................................................................................... 158
Table 8.3-2: Comparison between A-ALSVM and AD-ALSVM ........................................................................ 161
Table 8.3-3: Comparison of A-ALSVM, AD-ALSVM and Random SVM (Ac) .................................................. 162
Table 8.3-4: Comparison of A-ALSVM, AD-ALSVM and Random SVM (CC) .................................................. 162
Table 8.3-5: Comparison of A-ALSVM, AD-ALSVM and Random SVM (Sn) .................................................. 164
Table 8.3-6: Comparison of A-ALSVM, AD-ALSVM and Random SVM (Sp) .................................................. 164
## Index of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-ALSVM</td>
<td>Adaptive Active Learning with SVM</td>
</tr>
<tr>
<td>ACE-AL</td>
<td>Average-based sequential Class Entropy based batch mode Active Learning</td>
</tr>
<tr>
<td>ACE-ALSVM</td>
<td>Average-based sequential Class Entropy based batch mode Active Learning with SVM</td>
</tr>
<tr>
<td>AD-ALSVM</td>
<td>Angle Diversity incorporated Active Learning with SVM</td>
</tr>
<tr>
<td>ALSVM</td>
<td>Active Learning with SVM</td>
</tr>
<tr>
<td>CDB-ALSVM</td>
<td>Cannot-link constraint DB-ALSVM</td>
</tr>
<tr>
<td>CE-ALSVM</td>
<td>Class Entropy based batch mode Active Learning with SVM</td>
</tr>
<tr>
<td>DACE-ALSVM</td>
<td>Degree incorporated Average-based sequential Class Entropy based batch mode Active Learning with SVM</td>
</tr>
<tr>
<td>DB-ALSVM</td>
<td>Dynamic Batch mode Active Learning with SVM</td>
</tr>
<tr>
<td>DMCE-ALSVM</td>
<td>Degree incorporated Max-based sequential Class Entropy based batch mode Active Learning with SVM</td>
</tr>
<tr>
<td>MCE-AL</td>
<td>Maximum-based sequential Class Entropy based batch mode Active Learning</td>
</tr>
<tr>
<td>MCE-ALSVM</td>
<td>Maximum-based sequential Class Entropy based batch mode Active Learning with SVM</td>
</tr>
<tr>
<td>MD-AL</td>
<td>Manifold Diversity incorporated Active Learning</td>
</tr>
<tr>
<td>MD-ALSVM</td>
<td>Manifold Diversity incorporated Active Learning with SVM</td>
</tr>
<tr>
<td>MDB-ALSVM</td>
<td>Must-link constraint DB-ALSVM</td>
</tr>
<tr>
<td>NDB-ALSVM</td>
<td>Normal constraint DB-ALSVM</td>
</tr>
<tr>
<td>PD-AL</td>
<td>Pseudo-labeled Data incorporated Active Learning</td>
</tr>
<tr>
<td>PD-ALSVM</td>
<td>Pseudo-labeled Data incorporated Active Learning with SVM</td>
</tr>
<tr>
<td>SAAD-ALSVM</td>
<td>Simulated Annealing AD-ALSVM</td>
</tr>
<tr>
<td>SASD-ALSVM</td>
<td>Simulated Annealing SD-ALSVM</td>
</tr>
<tr>
<td>SD-ALSVM</td>
<td>Simple Distance-based ALSVM</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>UPD-ALSVM</td>
<td>Unbalanced PD-ALSVM</td>
</tr>
</tbody>
</table>