Swarm Intelligence in Educational Data Mining

Anwar Ali Yahya, Addin Osman, Ahmad Taleb
College of Computer Science and Information Systems,
Najran University,
Najran, KSA
{aaesmail,amoaddin,amtaleb}@nu.edu.sa

ABSTRACT
This paper explores the potential intersection of two fascinating and increasingly burgeoning fields: Swarm Intelligence and Educational Data Mining. A thorough review of the existing works in both fields has revealed a lack of Swarm Intelligence applications in Educational Data Mining, despite its successful applications in other data mining domain. Consequently, this paper launches the intersection of the two emerging fields by applying Swarm Intelligence techniques to an Educational Data Mining classification problem. More specifically, it proposes the use of Particle Swarm Classification to classify teachers’ classroom questions into the cognitive levels identified in Bloom's taxonomy. To do so, a dataset of questions has been collected and classified manually into Bloom's cognitive levels. Preprocessing steps have been applied to convert questions into a suitable representation. Using the dataset, the performance of Particle Swarm Classification is evaluated against several traditional machine learning approaches. The initial results provide evidences on the superiority of Particle Swarm Classification technique over the best traditional machine learning approaches.

Keywords
Educational Data Mining; Swarm Intelligence; Machine Learning; Bloom's Taxonomy.

1. INTRODUCTION
Educational data mining (EDM) is increasingly recognized as an emerging discipline focusing on the development of methods for exploring the unique types of data that come from an educational context [10] and, using these methods, to better understand students and the settings in which they learn [9]. These data come from several sources. In addition to the traditional face-to-face classroom environments, the increasing use of instrumental educational software and web-based education have created large repositories, in which large amounts of information about teaching-learning interaction are endlessly generated and ubiquitously available. EDM seeks to use these data repositories to better understand learners and learning, and to develop computational approaches that combine data and theory to transform practice to benefit learners.

EDM build off of data mining’s success in other domains such as commerce and biology [12]. In fact, it is considered as education “catching up” to other domains of data mining applications. Despite the similarity of the data mining methods used across domains, there are some important aspects that differentiate their application to education from other domains [11]. First, the objective of EDM is to improve learning which, because of difficulty of its measurement, must be estimated through proxies such as improved performance. Second, educational data have special characteristics that require a different treatment of the mining problem. More specifically, data hierarchy and non-independence becomes particularly important to account for. Third, as a result of the special characteristics of educational data, some specific data mining techniques are needed, and whilst some traditional techniques can be adapted, some cannot.

Despite its relative newness and growing, there are many applications of EDM that have been published. A comprehensive survey on the published EDM works can be found in five main survey papers [2, 3, 4, 8, 11], each of which reviews the EDM works published over a period of several years. Based on a thorough review of the EDM works reported in these survey papers, we have reported a lack of applying swarm intelligence techniques to tackle EDM problems, despite their successful applications in other data mining domains [1, 8]. Consequently, this paper launches the intersection of the swarm intelligence and EDM fields by applying swarm intelligence techniques to tackle an classification problem in EDM.

2. LITERATURE REVIEW
As mentioned above, a comprehensive survey on the existing EDM works can be found in five survey papers published to date, each of which presents different taxonomies of the reviewed works. The first taxonomy is presented by Baker [2], which classifies EDM works as follows: prediction (classification, regression, density estimation), clustering, relationship mining (association rule mining, correlation mining, sequential pattern mining, causal data mining), distillation of data for human judgment, discovery with models.

The second taxonomy appears in a survey paper by Baker [3] who suggests four key areas of EDM applications: improving student models, improving domain models, studying the pedagogical support provided by learning software, scientific research into learning and learners; and five approaches/methods: prediction, clustering, relationship mining, distillation of data for human judgment and discovery with models.

Castro [4] suggests the following taxonomy of EDM works: applications dealing with the assessment of the student’s learning performance, applications that provide course adaptation and
learning recommendations based on the student’s learning behavior, approaches dealing with the evaluation of learning material and educational web-based courses, applications that involve feedback to both teacher and students in e-learning courses, and developments for detection of atypical students’ learning behaviors.

In [13] a categorization for the main educational tasks that have employed data mining techniques is presented. These categories come from different research communities and use different data mining techniques as follows: analysis and visualization of data, providing feedback for supporting instructors, recommendations for students, predicting student’s performance, student modeling, detecting undesirable student behaviors, grouping students, social network analysis, developing concept maps, constructing courseware, planning and scheduling.

Finally a most recent survey of the EDM works is presented in [10], which extends the period described by the earlier surveys that cover from 1995 up to 2009. It presents a survey of EDM works fulfilled from 2010 up to 2013. In this survey, a sample of 240 EDM works is categorized in the following categories: student modeling, student behavior modeling, student performance modeling, assessment, student support and feedback, curriculum, domain knowledge, sequencing, and teachers support, and tools.

In reviewing the data mining techniques used in the EDM works mentioned in these surveys, we have identified a gap in which swarm intelligence techniques can be applied, following their successful applications in other data mining domains [1, 8].

3. SWARM INTELLIGENCE

Swarm Intelligence (SI) studies the collective behavior of systems composed of many individuals interacting locally with each other and with their environment. In biological swarms, the individuals (ant, bee, termite, bird or fish) are by no means complete engineers, but instead are simple creatures with limited cognitive abilities and limited means to communicate. Yet the complete swarm exhibits intelligent behavior, providing efficient solutions for complex problems such as predator evasion and shortest path finding.

The observed success and efficiency of swarms in nature to solve difficult problems inspired researchers in computer science to develop swarm-based systems for hard optimization problems. In this respect, a distinction can be made between ant colony optimization, particle swarm optimization and prey models. Successful applications of SI techniques include the modeling of agent behavior and various optimization problems, such as the routing of packages through networks the traveling salesman problem, scheduling, robotics and data mining topic of this paper.

A detailed review of SI applications in data mining can be found in [1, 6]. According to [1], the SI applications in data mining fall into two categories: the first category consists of techniques where individuals of a swarm move through a solution space and look for solutions for the data mining task at hand. The ant colony optimization and particle swarm optimization approaches are two examples of the first category. They consist of individuals wandering through the search space in some effective manner that combines exploration and exploitation. Whilst this search space is discrete and a solution is defined implicitly for ant colony optimization approach, the search space is continuous and the locations within the search space are updated explicitly in particle swarm optimization approach. In the second category, named data organizing, swarms move data instances that are placed on a low-dimensional (typically two-dimensional) feature space in order to come to a suitable clustering or low-dimensional mapping solution of the data. In this category of clustering techniques fall ant-based sorting and prey models.

4. SWARM INTELLIGENCE IN EDUCATIONAL DATA MINING

As it has been pointed out earlier, there is a lack of applying SI techniques in EDM domain. In our opinion, this can be attributed to the relatively newness of the two fields. In light of the successful applications of SI in other data mining domains [1, 8] and the growing interest of using Particle Swarm Classification (PSC) for classification problems [5, 7, 13], this section presents an example of the intersection between SI an EDM where PSC is used for mining teachers' classroom questions.

4.1 Mining Teachers' Classroom Questions

In the field of education, questioning is widely acknowledged as an important instructional strategy. Teachers use questions for a variety of purposes: to develop interest and motivate students to become actively involved in lessons, to evaluate students’ preparation and check on homework or seatwork completion, to develop critical thinking skills and inquiring attitudes, to review and summarize previous lessons, to nurture insights by exposing new relationships.

Realizing the importance of questioning, the analysis of teacher's classroom questions has been researched extensively. The analysis of teacher's classroom questions is based on the existence of questions taxonomies. Therefore, many classification systems have been developed. Some of them consist of a limited number of general categories which can be used to classify question irrespective of context, while other classification systems were developed for a specific curriculum. Bloom's Taxonomy is the most salient example of questions taxonomy. It was developed by Benjamin Bloom, in his efforts to classify the thinking behaviours into three domains: cognitive (mental skills), affective (growth in feelings or emotional areas) and psychomotor (manual or physical skills). The cognitive domain has received much attention because of its applicability in secondary and postsecondary education. Under the cognitive domain, Bloom identified six different levels (BCL) of learning and organized them on the basis of hierarchy as follows:

- **Knowledge**: focuses on memorization, recognition, and recall of information;
- **Comprehension**: focuses on organization of ideas, interpretation of information, and translation;
- **Application**: focuses on problem solving, use of particulars and principles;
- **Analysis**: focuses on finding the underlying organization, and the division of a whole into components;
- **Synthesis**: focuses on a combination of ideas to form something new, creating something unique whether verbal or physical;
- **Evaluation**: focuses on making judgments about issues, resolving disparities or disagreements.
4.2 Teachers' Classroom Questions Dataset

In order to apply PSC technique, a data set of teachers' classroom questions has been collected from a set of courses lectures at Najran University. The questions have been annotated manually, with assistance of a pedagogical expert. To avoid the potential effect of skewed data, each BCL has 1000 questions. The collected dataset have been processed through punctuation and stop words removal, tokenization, and stemming. Table I shows samples of questions and their corresponding BCLs class.

<table>
<thead>
<tr>
<th>BCL</th>
<th>Question example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>Can you define the fourth normal form</td>
</tr>
<tr>
<td>Comprehension</td>
<td>can you explain how we can implement these operations on an array</td>
</tr>
<tr>
<td>Application</td>
<td>how can we use the table generated by the dynamic programming algorithm</td>
</tr>
<tr>
<td>Analysis</td>
<td>can this algorithm be classified as a stable algorithm</td>
</tr>
<tr>
<td>Synthesis</td>
<td>how to develop a bottom-up version of merge-sort algorithm</td>
</tr>
<tr>
<td>Evaluation</td>
<td>can you decide which grammar should be used</td>
</tr>
</tbody>
</table>

4.3 PSC for Mining Teachers' Classroom Questions

The main idea of applying PSC to classify teachers' classroom questions is that each BCL is totally identified by a centroid in the particle space. Thus, given the questions dataset with 6 BCL classes and N input parameters, the classification problem can be viewed as finding the optimal coordinates of each BCL centroid in N-dimensional space. We start with a swarm of particles, each one representing potential centroids, and then iterate through a training stage to obtain the best centroid for each class. During the subsequent decision stage, the centroid is evaluated by a fitness function to measure the classification accuracy. The ith individual in the swarm is encoded in vector form as follows.

\[
\{ P_i^{BCL}, P_i^{RCL}, V_i^{BCL}, V_i^{RCL} \} \quad (1)
\]

In this formula, the position of the given BCL centroid given by N real numbers representing its coordinates in problem space:

\[
P_i^{BCL} = \{ P_i^{BCL}, \ldots, P_i^{BCL} \} \quad (2)
\]

and

\[
P_i^{RCL} = \{ P_i^{RCL}, \ldots, P_i^{RCL} \} \quad (3)
\]

Similarly the velocity of the the given BCL centroid is made up of N real numbers representing its velocity components in problem space:

\[
V_i^{BCL} = \{ V_i^{BCL}, \ldots, V_i^{BCL} \} \quad (4)
\]

and

\[
V_i^{RCL} = \{ V_i^{RCL}, \ldots, V_i^{RCL} \} \quad (5)
\]

The fitness function \( \psi \) is computed as the sum over all training set instances of the Euclidean distance between \( x_j \) and \( P_i^{BCL} \), which is the current potential centroid of the given BCL. This sum is divided by \( D_{Train} \), which is the number of instances composing the training set:

\[
\psi(i) = \frac{1}{D_{Train}} \sum_{j=1}^{D_{Train}} d(x_j, P_i^{BCL}) \quad (6)
\]

Given this fitness function, the problem becomes a typical minimization problem. During the training stage, the smaller this fitness is, the more representative is the potential centroid. At the end of the stage, this value is stored to be reused during the testing stage as the centroid representing the BCL class.

4.4 Experimental Results

The PSC has been applied, as described above to find the centroid for each BCL and PSC performance is evaluated in term of F1 using 10-fold cross validation and the following PSC parameters: \( n = 50, T_{max} = 1000, \nu_{max} = 0.05, \nu_{min} = -0.05, c_1 = 2.0, c_2 = 2.0, w_{max} = 0.9, w_{min} = 0.4 \).

In addition to the PSC experiments, four traditional machine learning techniques: k-Nearest Neighbors (kNN), Naïve Bayes (NB), Support Vector Machines (SVM), and Rocchio Algorithm (RA), and four term selection approaches, Term Frequency(TF), Mutual Information(MI), Information Gain(IG), and Chi square (\( \chi^2 \)) were experimented. Table 2 presents a summary of the results of these experiments, which have been presented in a detailed form in [14, 15]. As shown in Table 2, the combination of RA and IG led to the best performance.

<table>
<thead>
<tr>
<th>ML Technique</th>
<th>TF</th>
<th>MI</th>
<th>IG</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN</td>
<td>0.685</td>
<td>0.610</td>
<td>0.670</td>
<td>0.642</td>
</tr>
<tr>
<td>NB</td>
<td>0.675</td>
<td>0.595</td>
<td>0.617</td>
<td>0.621</td>
</tr>
<tr>
<td>SVM</td>
<td>0.720</td>
<td>0.637</td>
<td>0.730</td>
<td>0.716</td>
</tr>
<tr>
<td>RA</td>
<td>0.726</td>
<td>0.721</td>
<td>0.746</td>
<td>0.715</td>
</tr>
</tbody>
</table>

The performance of PSC is compared against the performance of RA using IG as a term selection approach. Figure 1, depicts the performance of PSC and RA in term of F1 using different number of terms selected by IG approach for the Comprehension and Evaluation BCL. As shown in Figure 1, PSC is able to outperform RA in many experimental cases.
Conclusion
Motivated by the lack of SI applications in EDM and the successful applications of SI in several data mining domains, this paper launches the intersection of the two emergent domains by applying SI to tackle a classification problem in EDM. More specifically, it proposes the use of PSC to classify teachers’ classroom questions into different cognitive levels identified in Bloom’s taxonomy. The initial PSC results show its promising success comparing with the traditional machine learning techniques. The future plan of this work will focus on further enhancement of PSC performance by tuning its parameters and exploring its variants.

6. REFERENCES