1 Abstract

Binary classification has been applied in an attempt to automatize the detection of engineering student misconceptions in the topics of diffusion and heat transfer. According to Educational Engineering experts, analyzing the verbs used by students to describe science concepts can systematically assess the presence of misconceptions. Modern misconception assessment is based on the ontological categorizations of concepts where each category is accompanied by the different verbs used to describe these concepts. This systematic analysis is called the predicate test. The predicate test has proven to be successful for conceptual change assessment but has not been automated with the use of modern computing systems.

The main goal of this research is to show that conceptual change assessment can be considered a task of text classification. The J48 binary classifier has been used to build a prediction model using the results of predicate tests, which have been previously annotated by Educational Engineering experts.

2 Ontologies and Conceptual Change

- Student and teacher interpretation of concepts is based on the ontological categorization of ideas or concepts into the categories represented in Figure 1 [1]
- Incorrect categorization of concepts indicates the presence of a misconception and identifies the need for conceptual change [1]

![Figure 1: Distributed concept ontologies proposed by Chi.](Image)

Chi’s conceptual change theory is based on three assumptions:

1. Entities in the world can be ontologically categorized
2. The nature of physics science concepts dictates their categorization into constraint-based interaction concepts
3. Students hold naïve preconceptions aligned with substance-based descriptions of concepts

3 Predicate Tests and Misconceptions

- Compares student and expert predicate use to detect misconceptions [1]
- Verbs used by domain experts to describe concepts are known as ontological attributes of that concept [1]
- The predicate test identifies ontological boundaries between student and expert predicate by analyzing the ontological attributes used to describe concepts [1]

4 Text Classification for Predicate Test

- Our dataset consists of descriptions to multiple-choice question answers about diffusion and heat transfer
- Our dataset contains 40 examples of the emergent process class and 100 examples of the direct process class
- J48 [2] was trained with our expert annotated dataset where ontological attributes are categorized into the emergent process and sequential process categories
- Main research question:
  - Is it possible to automatize predicate tests with text classification?

5 Classification Results

- 95 instances were classified into the emergent process class with 79 true positives
- 105 instances were classified into the sequential process class with 84 true positives
- Table 1 and 2 show our classifier results and confusion matrix respectively.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergent</td>
<td>0.832</td>
<td>0.79</td>
</tr>
<tr>
<td>Sequential</td>
<td>0.8</td>
<td>0.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class</th>
<th>Emergent</th>
<th>Sequential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergent</td>
<td>79</td>
<td>21</td>
</tr>
<tr>
<td>Sequential</td>
<td>16</td>
<td>84</td>
</tr>
</tbody>
</table>

6 Conclusions and Future Work

- Our results provide empirical evidence that binary classification can be used to successfully automate the predicate test
- Sensitivity or recall describes how 79% of emergent class and 84% of sequential class instances were correctly classified
- Our quantitative results show that the predicate test can be automated using supervised learning techniques
- Enhancing our feature set may improve classifier results
- Bags of words theory will be applied to further experiment with additional preprocessing techniques
- We will experimenting with multi-label classification instead of multi-class classification

7 References