Natural Language Processing in the educational environment

Alessio Miaschi $^{1,2}$

$^1$University of Pisa

$^2$ItaliaNLP Lab, Institute for Computational Linguistics “Antonio Zampolli” (ILC-CNR)

Mauriana Pesaresi Seminar, April 2
The educational environment is changing on a drastic speed, from traditional classroom teaching ecology from the adaptive individual/collaborative learning.
The educational environment is changing on a drastic speed, from traditional classroom teaching ecology from the adaptive individual/collaborative learning.

In recent years, the interest in applying NLP to education has rapidly increased.
The educational environment is changing on a drastic speed, from traditional classroom teaching ecology from the adaptive individual/collaborative learning.

In recent years, the interest in applying NLP to education has rapidly increased.

Several commercial applications already include high-stakes assessments of text and speech, writing assistants and online instructional environments.
NLP can enhance educational technology in several ways:
  ▶ automate the scoring of student texts with respect to linguistic dimensions such as grammatical correctness or organizational structure (automated essay scoring systems)
NLP for educational application

- NLP can enhance educational technology in several ways:
  - automate the scoring of student texts with respect to linguistic dimensions such as grammatical correctness or organizational structure (automated essay scoring systems)
  - track and evaluate the evolution of student’s writing skills
NLP can enhance educational technology in several ways:

- automate the scoring of student texts with respect to linguistic dimensions such as grammatical correctness or organizational structure (automated essay scoring systems)
- track and evaluate the evolution of student’s writing skills
- processing text from the web in order to personalize instructional materials to the interests of individual students
NLP for educational application

- NLP can enhance educational technology in several ways:
  - automate the scoring of student texts with respect to linguistic dimensions such as grammatical correctness or organizational structure (automated essay scoring systems)
  - track and evaluate the evolution of student’s writing skills
  - processing text from the web in order to personalize instructional materials to the interests of individual students
  - automate the generation of test questions for teachers
I. Tracking the Evolution of Written Language Competence
Definition of a NLP model to:
- track and evaluate the evolution of lower secondary school student’s writing skills
- identify relations between the evolution of written language competence and students’ background information
Definition of a NLP model to:
- track and evaluate the evolution of lower secondary school student’s writing skills
- identify relations between the evolution of written language competence and students’ background information

First Italian research:
- based on a diachronic corpus of students’ essays
- focused on the the evolution of the syntactic and lexical features and on the impact of the errors made by students
Collection of 1352 essays written by 156 Italian L1 learners during the first and second year of lower secondary school
Collection of 1352 essays written by 156 Italian L1 learners during the first and second year of lower secondary school.

Essays collected among seven different schools of Rome: 3 from the center and 4 from the suburbs.
- Collection of 1352 essays written by 156 Italian L1 learners during the first and second year of lower secondary school
- Essays collected among seven different schools of Rome: 3 from the center and 4 from the suburbs
- Each essay was:
  - manually annotated for a wide range of spelling errors
  - linguistically annotated
  - converted in a set of 147 linguistic features (lexical, morphosyntactic and syntactic)
Given a set of chronologically ordered essays written by the same student, a document $d_j$ should show a higher written quality level with respect to the ones written previously.
Given a set of chronologically ordered essays written by the same student, a document $d_j$ should show a higher written quality level with respect to the ones written previously.

Following from this assumption, we considered the problem of tracking the evolution of a student as a classification task.
Given a set of chronologically ordered essays written by the same student, a document $d_j$ should show a higher written quality level with respect to the ones written previously.

Following from this assumption, we considered the problem of tracking the evolution of a student as a classification task.

For each pair of documents, we built an $E$ event:

$$E = V_1 + V_2 + (V_1 - V_2)$$
The experiments: time intervals

<table>
<thead>
<tr>
<th>Time interval</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text at distance = 1 month</td>
<td>1087.85</td>
<td>181.14</td>
</tr>
<tr>
<td>First and penultimate text (single year)</td>
<td>498.14</td>
<td>82.85</td>
</tr>
<tr>
<td>All first and second year texts (without common prompt)</td>
<td>3301</td>
<td>550</td>
</tr>
<tr>
<td>Text at distance = 1 year</td>
<td>527</td>
<td>87.71</td>
</tr>
<tr>
<td>First and penultimate text (two years)</td>
<td>253</td>
<td>42</td>
</tr>
<tr>
<td>First and last text (single year)</td>
<td>426</td>
<td>70.85</td>
</tr>
<tr>
<td>All first and second year texts (with common prompt)</td>
<td>4999.85</td>
<td>833.14</td>
</tr>
<tr>
<td>Common prompt</td>
<td>145</td>
<td>24</td>
</tr>
<tr>
<td>First and last text (two years)</td>
<td>198.14</td>
<td>32.85</td>
</tr>
<tr>
<td>All</td>
<td>13814.71</td>
<td>2302.28</td>
</tr>
</tbody>
</table>

**Table**: Average number of $E$ events in the 10 datasets.
We defined three different sets of experiments, using respectively:

1. the 147 linguistic features
2. + lexical complexity features
3. + lexical complexity features + annotated error features
We defined three different sets of experiments, using respectively:

1. the 147 linguistic features
2. + lexical complexity features
3. + lexical complexity features + annotated error features

Lexical complexity features: *words frequency class*, a measure of the average class frequency words in a document
We defined three different sets of experiments, using respectively:

1. the 147 linguistic features
2. + lexical complexity features
3. + lexical complexity features + annotated error features

Lexical complexity features: *words frequency class*, a measure of the average class frequency words in a document.

The annotated error features refer to the distribution of grammatical, orthographic, lexical and punctuation errors.
The experiments: results

- Leave-one-school-out Cross-validation
- Support Vector Machines as learning algorithm

<table>
<thead>
<tr>
<th>Time interval</th>
<th>1st set (F₁)</th>
<th>2nd set (F₁)</th>
<th>3rd set (F₁)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essays written at dist = 1 month</td>
<td>0.52</td>
<td>0.52</td>
<td>0.53</td>
</tr>
<tr>
<td>1st essay - second-last essay (one year)</td>
<td>0.54</td>
<td>0.54</td>
<td>0.55</td>
</tr>
<tr>
<td>Essays written at dist = 1 year</td>
<td>0.57</td>
<td>0.57</td>
<td>0.65</td>
</tr>
<tr>
<td>1st essay - second-last essay (two years)</td>
<td>0.70</td>
<td>0.70</td>
<td>0.73</td>
</tr>
<tr>
<td>1st year - 2nd year</td>
<td>0.63</td>
<td>0.67</td>
<td>0.73</td>
</tr>
<tr>
<td>All</td>
<td>0.58</td>
<td>0.56</td>
<td>0.58</td>
</tr>
</tbody>
</table>
Starting from this results, we defined a qualitative research in order to verify:

- which linguistic features contributes more to the identification of the writing skills’ evolution (*feature selection*)
Starting from this results, we defined a qualitative research in order to verify:

- which linguistic features contributes more to the identification of the writing skills’ evolution (*feature selection*)
- whether the evolution of written language competence is significantly related to the students’ background information
### Feature selection: results

<table>
<thead>
<tr>
<th>No</th>
<th>Essays written at dist = 1 month</th>
<th>Essays written at dist = 1 year</th>
<th>1st essay - second-last essay (two years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adjectives</td>
<td>Number of tokens</td>
<td>Word frequency class</td>
</tr>
<tr>
<td>2</td>
<td>Post-verbal subjects</td>
<td>Number of sentences</td>
<td>Auxiliar relations</td>
</tr>
<tr>
<td>3</td>
<td>Word frequency class</td>
<td>% chars for token</td>
<td>Auxiliar verbs</td>
</tr>
<tr>
<td>4</td>
<td>Number of tokens</td>
<td>Excess of pronouns</td>
<td>Auxiliar verbs (1st person plural)</td>
</tr>
<tr>
<td>5</td>
<td>Principal verbs (3rd person singular)</td>
<td>Grammatical errors</td>
<td>Auxiliar verbs (indicative)</td>
</tr>
<tr>
<td>6</td>
<td>Predicate adjectives</td>
<td>Principal verbs (past)</td>
<td>Number of sentences</td>
</tr>
<tr>
<td>7</td>
<td>Predicative relation</td>
<td>Number of tokens</td>
<td>Number of tokens</td>
</tr>
<tr>
<td>8</td>
<td>Principal verbs (past)</td>
<td>Word frequency class</td>
<td>Word frequency class</td>
</tr>
<tr>
<td>9</td>
<td>Dependency relations</td>
<td>% All errors</td>
<td>Subordinate clause (Degree = 1)</td>
</tr>
<tr>
<td>10</td>
<td>Prepositions</td>
<td>Predicative relations</td>
<td>Dependency relations</td>
</tr>
</tbody>
</table>

**Table:** Ranking of the first 10 features for three different time intervals.
### Feature selection: results

<table>
<thead>
<tr>
<th>Features</th>
<th>1st essay - second-last essay (two years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical errors</td>
<td>0.74</td>
</tr>
<tr>
<td>Ortographic errors</td>
<td>0.72</td>
</tr>
<tr>
<td>Lexical errors</td>
<td>0.70</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>0.68</td>
</tr>
</tbody>
</table>

**Table:** Classification results using different sets of annotated error features.
Using the confidence of our classifier, we tried to identify the relation between the evolution of written language competence and the students’ background information.
Using the confidence of our classifier, we tried to identify the relation between the evolution of written language competence and the students’ background information.

Assumption: at a higher confidence interval could correspond a notable evolution of the student’s writing skills.
Using the confidence of our classifier, we tried to identify the relation between the evolution of written language competence and the students' background information.

Assumption: at a higher confidence interval could correspond a notable evolution of the student’s writing skills.

Once computed the confidence intervals, we split the students according to:
- Center/Suburb of Rome
- Confidence intervals
## Writing competence and background information

### Essays written at distance

<table>
<thead>
<tr>
<th>Urban area</th>
<th>Essays written at dist = 1 month</th>
<th>1st essay - second-last essay (two years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center</td>
<td>0.579</td>
<td>0.629</td>
</tr>
<tr>
<td>Suburbs</td>
<td>0.513</td>
<td>0.670</td>
</tr>
</tbody>
</table>

### Confidence

<table>
<thead>
<tr>
<th>Confidence</th>
<th>% foreign students</th>
<th>% bilingual students</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>26.08</td>
<td>65.21</td>
</tr>
<tr>
<td>Low</td>
<td>10.54</td>
<td>46.6</td>
</tr>
</tbody>
</table>
Conclusion

- Investigated the possibility to define the evolution of students’ writing skills as a classification task
Conclusion

- Investigated the possibility to define the evolution of students’ writing skills as a classification task
- Studied the contribute of each linguistic feature in the identification of the writing skills’ evolution

Future developments:
- experiments using wide time intervals and geographical areas
- integration of the computational model in teaching tools (MOOC platforms, etc.)
Investigated the possibility to define the evolution of students’ writing skills as a classification task

Studied the contribute of each linguistic feature in the identification of the writing skills’ evolution

Identified some relations between the evolution of written language competence and the students’ background information

Future developments:
- experiments using wide time intervals and geographical areas
- integration of the computational model in teaching tools (MOOC platforms, etc.)
Conclusion

- Investigated the possibility to define the evolution of students’ writing skills as a classification task
- Studied the contribute of each linguistic feature in the identification of the writing skills’ evolution
- Identified some relations between the evolution of written language competence and the students’ background information
- Future developments:
  - experiments using wide time intervals and geographical areas
  - integration of the computational model in teaching tools (MOOC platforms, etc.)
II. Identifying prerequisite relationships among learning objects
In the age of e-learning, many instructors are facing the hard task of building web-based courses.
In the age of e-learning, many instructors are facing the hard task of building web-based courses. The primary target is to share the knowledge, through the repository of Learning Objects on the web.
In the age of e-learning, many instructors are facing the hard task of building web-based courses. The primary target is to share the knowledge, through the repository of Learning Objects on the web. In those repository there isn’t correlation between materials.
Introduction

- In the age of e-learning, many instructors are facing the hard task of building web-based courses.
- The primary target is to share the knowledge, through the repository of Learning Objects on the web.
- In those repository there isn’t correlation between materials.
- In order to generate automatically chains of relations between LOs, it is necessary to infer prerequisite relations among concepts.
Prerequisite: a definition

What should one know/learn before starting to learn a new area such as deep learning?
What should one know/learn before starting to learn a new area such as deep learning?

A prerequisite is usually a concept or requirement before one can proceed to a following one.
What should one know/learn before starting to learn a new area such as deep learning?

A prerequisite is usually a concept or requirement before one can proceed to a following one.

A concept $C_1$ is a prerequisite to another concept $C_2$ if the knowledge of $C_1$ is necessary to understand $C_2$. 

Some issues

- The prerequisite relation exists as a natural dependency among concepts in cognitive processes.
Some issues

- The prerequisite relation exists as a natural dependency among concepts in cognitive processes
- Prerequisite relations can differ according to different domains
Some issues

- The prerequisite relation exists as a natural dependency among concepts in cognitive processes.
- Prerequisite relations can differ according to different domains.
- Discovering prerequisite relations among concepts is usually done manually by domain experts → inefficient and expensive.
Early works explored Wikipedia as a resource for detecting prerequisite relations

AL-CPL Dataset (Liang et al., 2018): collections of concept pairs on four different domains:
- Data Mining
- Geometry
- Physics
- Precalculus
Early works explored Wikipedia as a resource for detecting prerequisite relations.

Classify prerequisite relations using Wikipedia articles and their linkage structure.

AL-CPL Dataset (Liang et al., 2018): collections of concept pairs on four different domains:
- Data Mining
- Geometry
- Physics
- Precalculus
Early works explored Wikipedia as a resource for detecting prerequisite relations.

Classify prerequisite relations using Wikipedia articles and their linkage structure.

AL-CPL Dataset (Liang et al., 2018): collections of concept pairs on four different domains:

- Data Mining
- Geometry
- Physics
- Precalculus
## AL-CPL Dataset (Liang et al., 2018)

### English

<table>
<thead>
<tr>
<th>Domain</th>
<th>Concepts</th>
<th>Pairs</th>
<th>Prerequisites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Mining</td>
<td>120</td>
<td>826</td>
<td>292</td>
</tr>
<tr>
<td>Geometry</td>
<td>89</td>
<td>1681</td>
<td>524</td>
</tr>
<tr>
<td>Physics</td>
<td>153</td>
<td>1962</td>
<td>487</td>
</tr>
<tr>
<td>Precalculus</td>
<td>224</td>
<td>2060</td>
<td>699</td>
</tr>
</tbody>
</table>

### Italian

<table>
<thead>
<tr>
<th>Domain</th>
<th>Concepts</th>
<th>Pairs</th>
<th>Prerequisites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Mining</td>
<td>75</td>
<td>429</td>
<td>154</td>
</tr>
<tr>
<td>Geometry</td>
<td>73</td>
<td>1338</td>
<td>430</td>
</tr>
<tr>
<td>Physics</td>
<td>131</td>
<td>1651</td>
<td>409</td>
</tr>
<tr>
<td>Precalculus</td>
<td>176</td>
<td>1504</td>
<td>502</td>
</tr>
</tbody>
</table>

**Table:** Dataset statistics.
Our approach

Given a pair of concepts \((A, B)\), predict whether or not \(B\) is a prerequisite of \(A\)
Our approach

- Given a pair of concepts \((A, B)\), predict whether or not \(B\) is a prerequisite of \(A\)
- Using for each concept, the corresponding Wikipedia page
Our approach

- Given a pair of concepts \((A, B)\), predict whether or not \(B\) is a prerequisite of \(A\)
- Using for each concept, the corresponding Wikipedia page
- Training deep learning models using only:
  - a pre-trained word-embeddings lexicon (page features)
  - a set of linguistic features extracted from the Wikipedia pages (combined features)
Word embedding: categorize semantic similarities between linguistic items based on their distributional properties in large samples of language data.
Word embedding: categorize semantic similarities between linguistic items based on their distributional properties in large samples of language data

Lexicon of 128 dimensions built with **word2vec** (Mikolov et al., 2013) and starting from:

- itWac: 2-billion-word Italian corpus
- ukWac: 2-bilion-word English corpus
linguistic characteristics from the combination of $A$ and $B$ Wikipedia pages
Combined Features

- Linguistic characteristics from the combination of $A$ and $B$ Wikipedia pages
- 16 different text-based features, among which:
  - if $B/A$ appears in $A/B$ content
  - the Jaccard similarity between $A$ and $B$
  - the RefD metric between $A$ and $B$ as proposed by Liang et al. (2015)
Classifiers

- We tested three neural network models, according to the different type of features:
  - two LSTM sub-networks joined by concatenation (page features)
We tested three neural network models, according to the different type of features:

- two LSTM sub-networks joined by concatenation (page features)
- feedforward neural network (combined features)
We tested three neural network models, according to the different type of features:

- two LSTM sub-networks joined by concatenation (page features)
- feedforward neural network (combined features)
- combination of the two models
Experimental Settings

- Evaluated our approach predicting in-domain and cross-domain prerequisite relations.
Experimental Settings

- Evaluated our approach predicting in-domain and cross-domain prerequisite relations
- Balanced the training and testing sets by oversampling the minority class
Experimental Settings

- Evaluated our approach predicting in-domain and cross-domain prerequisite relations
- Balanced the training and testing sets by oversampling the minority class
- Zero Rule algorithm as baseline and F-Score as metric for evaluation
## Results

### Table: In-domain results.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Baseline</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Mining</td>
<td>66.66</td>
<td>72.45</td>
<td>64.25</td>
<td>77.91</td>
</tr>
<tr>
<td>Geometry</td>
<td>67.86</td>
<td>86.89</td>
<td>85.27</td>
<td>90.01</td>
</tr>
<tr>
<td>Physics</td>
<td>75.22</td>
<td>79.28</td>
<td>76.26</td>
<td>85.08</td>
</tr>
<tr>
<td>Precalculus</td>
<td>66.66</td>
<td>90.53</td>
<td>89.02</td>
<td>93.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain</th>
<th>Baseline</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Mining</td>
<td>66.66</td>
<td>88.81</td>
<td>73.29</td>
<td>89.66</td>
</tr>
<tr>
<td>Geometry</td>
<td>68.82</td>
<td>92.43</td>
<td>89.66</td>
<td>95.69</td>
</tr>
<tr>
<td>Physics</td>
<td>75.17</td>
<td>83.49</td>
<td>80.72</td>
<td>88.54</td>
</tr>
<tr>
<td>Precalculus</td>
<td>66.66</td>
<td>92.48</td>
<td>90.90</td>
<td>94.95</td>
</tr>
</tbody>
</table>

**Italian**

**English**
## Results

<table>
<thead>
<tr>
<th>Domain</th>
<th>Italian</th>
<th>Baseline</th>
<th>#2</th>
<th>#3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Mining</td>
<td>66.66</td>
<td>37.09</td>
<td>30.36</td>
<td></td>
</tr>
<tr>
<td>Geometry</td>
<td>67.86</td>
<td>79.53</td>
<td>76.33</td>
<td></td>
</tr>
<tr>
<td>Physics</td>
<td>75.22</td>
<td>71.56</td>
<td>69.6</td>
<td></td>
</tr>
<tr>
<td>Precalculus</td>
<td>66.66</td>
<td>83.66</td>
<td>83.4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain</th>
<th>English</th>
<th>Baseline</th>
<th>#2</th>
<th>#3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Mining</td>
<td>66.66</td>
<td>50.89</td>
<td>38.78</td>
<td></td>
</tr>
<tr>
<td>Geometry</td>
<td>68.82</td>
<td>80.41</td>
<td>82.53</td>
<td></td>
</tr>
<tr>
<td>Physics</td>
<td>75.17</td>
<td>74.74</td>
<td>63.67</td>
<td></td>
</tr>
<tr>
<td>Precalculus</td>
<td>66.66</td>
<td>87.14</td>
<td>84.41</td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Cross-domain results.
Further Work

- Repeat the experiments using only one domain in training and testing
Further Work

- Repeat the experiments using only one domain in training and testing
- Repeat the experiments with different classification methods
Further Work

- Repeat the experiments using only one domain in training and testing
- Repeat the experiments with different classification methods
- Design active learning strategies, in order to understand how much information we need to obtain good results for each domain
III. Conclusion
NLP can improve educational technology in several ways
NLP can improve educational technology in several ways

Different applications and perspectives, in order to address the needs of teachers and learners
Conclusion

- NLP can improve educational technology in several ways
- Different applications and perspectives, in order to address the needs of teachers and learners
- Future developments:
  - From prerequisite relations identification to personalized recommendations and intelligent tutoring systems
  - Text generation model for the educational scenario
Thanks for your attention!

WELL LOOK AT THAT

IT’S PIZZA O’CLOCK