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

www.italianlp.it

- The educational environment is changing on a drastic speed, from traditional classroom teaching ecology from the adaptive individual/collaborative learning

NLP for educational application

- 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

NLP for educational application

- 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 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)

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 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

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

Introduction

- 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

CItA Corpus (Barbagli et al., 2016)

- Collection of 1352 essays written by 156 Italian L1 learners during the first and second year of lower secondary school

CItA Corpus (Barbagli et al., 2016)

- 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

CItA Corpus (Barbagli et al., 2016)

- 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)

Approach and experiments definition

- 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

Approach and experiments definition

- 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

Approach and experiments definition

- 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

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

Table: Average number of E events in the 10 datasets.

The experiments: linguistic 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

The experiments: linguistic 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

The experiments: linguistic 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
- The annotated error features refer to the distribution of grammatical, ortographic, lexical and punctuation errors

ITALIAN NATURAL LANGUAGE PROCESSING LAB@ILC
www.italianlp.it

The experiments: results

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

Time interval	1st set (F_1)	2nd set (F_1)	3rd set (F_1)
Essays written at dist = 1 month	0.52	0.52	**0.53**
1st essay - second-last essay (one year)	0.54	0.54	**0.55**
Essays written at dist = 1 year	0.57	0.57	**0.65**
1st essay - second-last essay (two years)	0.70	0.70	**0.73**
1st year - 2nd year	0.63	0.67	**0.73**
All	0.58	0.56	**0.58**

Qualitative research

- 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*)

Qualitative research

- 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

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

Table: Ranking of the first 10 features for three different time intervals.

Feature selection: results

Features	1st essay - second-last essay (two years)
Grammatical errors	**0.74**
Ortographic errors	0.72
Lexical errors	0.70
Punctuation errors	0.68

Table: Classification results using different sets of annotated error features.

Writing competence and 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

Writing competence and 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

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

Writing competence and background information

Urban area	Essays written at dist = 1 month	1st essay - second-last essay (two years)
Center	0.579	0.629
Suburbs	0.513	0.670

Confidence	% foreign students	% bilingual students
High	26.08	65.21
Low	10.54	46.6

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

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

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

Introduction

- In the age of e-learning, many instructors are facing the hard task of building web-based courses

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 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 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?

Prerequisite: a definition

- 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

Prerequisite: a definition

- 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

Data collections

- Early works explored Wikipedia as a resource for detecting prerequisite relations

Data collections

- Early works explored Wikipedia as a resource for detecting prerequisite relations
- Classify prerequisite relations using Wikipedia articles and their linkage structure

Data collections

- 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			
Domain	**Concepts**	**Pairs**	**Prerequisites**
Data Mining	120	826	292
Geometry	89	1681	524
Physics	153	1962	487
Precalculus	224	2060	699
Italian			
Domain	**Concepts**	**Pairs**	**Prerequisites**
Data Mining	75	429	154
Geometry	73	1338	430
Physics	131	1651	409
Precalculus	176	1504	502

Table: Dataset statistics.

- 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)

ITALIAN NATURAL LANGUAGE PROCESSING LAB@ILC
www.italianlp.it

Page Features

- Word embedding: categorize semantic similarities between linguistic items based on their distributional properties in large samples of language data

Page Features

- 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)

Classifiers

- 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)

Classifiers

- 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

Italian				
Domain	**Baseline**	**#1**	**#2**	**#3**
Data Mining	66.66	72.45	64.25	**77.91**
Geometry	67.86	86.89	85.27	**90.01**
Physics	75.22	79.28	76.26	**85.08**
Precalculus	66.66	90.53	89.02	**93.91**
English				
Domain	**Baseline**	**#1**	**#2**	**#3**
Data Mining	66.66	88.81	73.29	**89.66**
Geometry	68.82	92.43	89.66	**95.69**
Physics	75.17	83.49	80.72	**88.54**
Precalculus	66.66	92.48	90.90	**94.95**

Table: In-domain results.

Results

Italian			
Domain	**Baseline**	**#2**	**#3**
Data Mining	**66.66**	37.09	30.36
Geometry	67.86	**79.53**	76.33
Physics	**75.22**	71.56	69.6
Precalculus	66.66	**83.66**	83.4
English			
Domain	**Baseline**	**#2**	**#3**
Data Mining	**66.66**	50.89	38.78
Geometry	68.82	80.41	**82.53**
Physics	**75.17**	74.74	63.67
Precalculus	66.66	**87.14**	84.41

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

Conclusion

- NLP can improve educational technology in several ways

Conclusion

- 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!