Review on Natural Language Processing in Education Field

Hang Qi
CS 410 Tech Review Fall 2021
NetId: hangq2

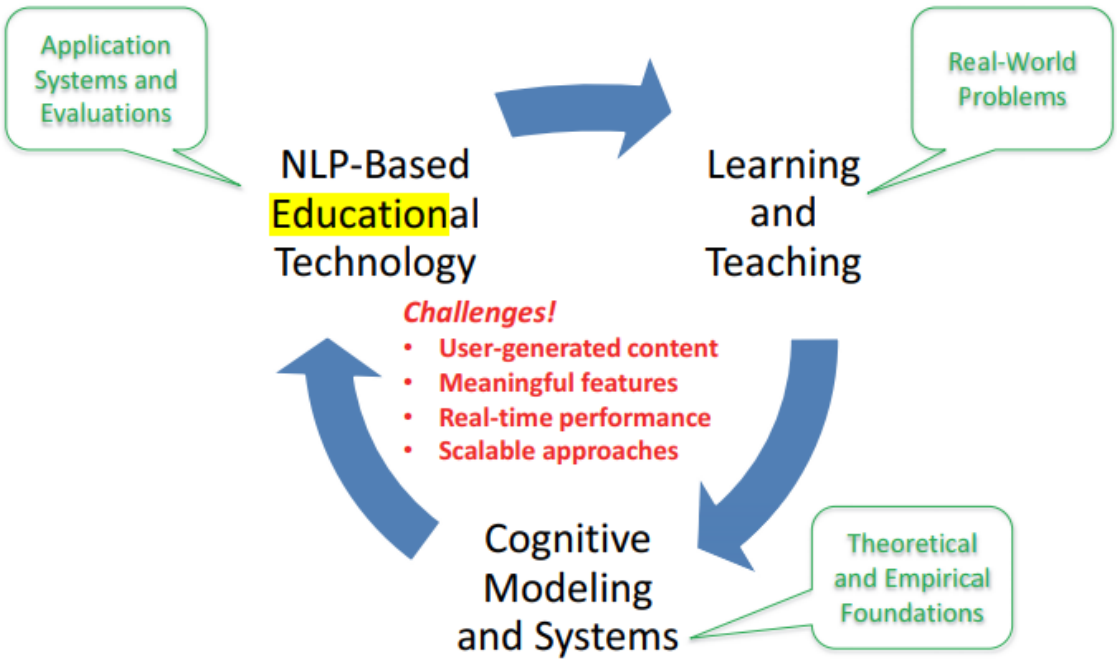
Abstract:
Natural Language Processing has been developed and studied for more than 50 years in the linguistics field. Its innovative technology has been applied to many fields today, including the education field. As the big technology companies invest more money into the field, the application becomes more mature and broad. It is everywhere in our daily studying lives. Natural Language Processing is applied in many areas of education, from auto grammar correction to voice recognition. Though there may be various ways of application, the development goes through a similar cycle: First, the problem arises through real-word need. Then, it is modified by literature theory. Lastly, Natural Language Processing Technique is then applied. Inevitably, many problems and challenges would occur during each phase of the development cycle. However, as the Natural Language Processing field combines more closely with the real-world problems, the challenge will be minimized.

Application:
First major application field is the auto scoring mechanism. Using NLP, the system can automatically find grammatical errors and analyze organizational structure.
Syntactic analysis is widely applied in checking and correcting grammatical errors. This technique has been helpful especially for ESL students and deaf students.
Semantic analysis is widely applied in analysing the meaning of content text, in both essay-long or short-answer contexts. It is also applied in topic analysis and clustering.

Second major application field is dialogue technologies. Studies have shown that students receiving human tutoring usually score higher than those receiving computer tutoring, mainly because human tutors can engage in unconstrained natural language dialogue with students. Thus, dialogue-based tutoring has been developed to decrease the gap between human and computer.

Third major application field is the text and speech processing field. The ultimate goal is to automatically generate course-related materials or even tests by searching and gathering course-related content from the web.

Development Cycle:



There are three stages in the development cycle.
On the upper right is the first stage. We discover the real-world problems through teacher and student's needs. And then we develop and apply Natural Language Processing techniques to solve the specific problem, or to meet the requirements and satisfy those needs.
For example, without tagging in the discussion forums, instructors have to go through all the problems and try to find the valuable ones that need instructors' attention. We then discover the need of using NLP technique to automatically identify the important and urgent questions that need specifically instructors' help.

On the lower bottom corner is the second stage. The cognitive models and systems are built upon theoretical foundations based on certain professional fields. The models and systems would need to match the characteristics of the specific field.
For example, certain types of models are trained to recognize and parse formally written articles, such as the Wall Street Journal. They will not be the same accurate when parsing content written in informal context, such as students' homework.

On the upper left corner is the third stage. This is the actual development and implementation stage. We have the goal of solving certain types of real world problems, then we build the

corresponding cognitive models and systems of that certain field, and lastly we can apply the NLP technologies based on what we had in the first two stages.
This cycle could be iterative. For example, when we have the NLP technique completely developed in the computer science field, we may want to revise the technique through the whole cycle in order to apply for the history field.

Challenges:
The first main challenge is that specific NLP tools are often trained only for specific topics. For example, parsing text related to the history field is hugely different from parsing text related to the computer science field. Generic methods still need to be developed.

The second main challenge is to adjust the weight of independent variables when predicting dependent variables. For example, word count may be related to the quality of essay, however, word count is not in the human grading rubric, thus not worthy to be mentioned in the evaluation.

The third main challenge is that since NLP techniques are usually applied in real time interactive applications, the algorithm needs to be fast enough that the responding time is under certain threshold.

Thoughts of Possible Future Application:
In the future, NLP could be more accurate in scoring tests and generating related materials, and also involve in more human-like dialogue with unconstrained sentencing. It will hugely ease the teachers' pressure. Teachers then can focus more on the lecturing side, other than doing repetitive useless tasks. Students can also have a fair environment, since the scoring method is standardized. And due to the expansion of scales, the tutoring resources can be more available to students.

References

Litman, Diane. "Natural language processing for enhancing teaching and learning." In *Thirtieth AAAI conference on artificial intelligence*. 2016.