

NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

PH.D DEFENCE - PUBLIC SEMINAR

Title: **Robust Trait-Specific Essay Scoring Using Neural Networks and Density Estimators**

Speaker: Mr Kaveh Taghipour

Date/Time: 13 June 2017, Tuesday, 04:00 PM to 05:30 PM

Venue: Executive Classroom, COM2-04-02

Supervisor : Dr Ng Hwee Tou, Professor, School of Computing

Abstract:

Traditional automated essay scoring systems rely on carefully designed features to evaluate and score essays. The performance of such systems is tightly bound to the quality of the underlying features. However, it is laborious to manually design the most informative features for such a system. In this thesis, we develop a novel approach based on recurrent neural networks to learn the relation between an essay and its assigned score, without any feature engineering. We explore several neural network models for the task of automated essay scoring and perform some analysis to get some insights of the models. The results show that our best system, which is based on long short-term memory networks, outperforms a strong baseline by 5.6% in terms of quadratic weighted kappa.

Our proposed essay scoring system returns a holistic score, reflecting the quality of the given essay. However, this information is not enough for language learners to improve their writing skill. Providing proper feedback about different dimensions of essay writing is essential for self-learning students to improve their writing skill. Several datasets have been created in recent years to address various aspects of essay writing. However, the performance of the existing systems is still far from perfect. In this thesis, we propose a novel framework based on our neural approach to essay scoring to model these aspects in student essays, without manually designed task-specific features. Among various writing aspects, we have used our framework to model argument strength and essay organization in student essays. The experiments show that our method outperforms strong state-of-the-art systems and leads to relative error reductions of 7.0% and 13.5% (in terms of mean squared error) in argument strength and essay organization tasks, respectively.

The third task that we have tackled is to distinguish essays written by humans from computer-generated essays. Current automated essay scoring systems can assign high scores to certain types of automatically generated nonsensical essays, thus allowing the systems to be gamed. We address this problem by proposing a novel approach for detecting computer-

generated fake essays, using density estimation methods. Our method only relies on essays written by humans and does not make any prior assumptions about the computer-generated fake essays. We have evaluated our method on essays automatically generated by sampling language models and context-free grammars. The results show that current state-of-the-art automated essay scoring systems fail to detect these two types of computer-generated fake essays. However, after integrating our method, these systems detect and penalize computer-generated essays effectively and as a result, they continue to perform well on essay scoring, on both human-written essays and computer-generated essays.