Answer Similarity Analysis using Convolutional Neural Networks

S M Ibrahim, Md. Mofaqkhayrul Islam Mim, Jahid Hasan Mamun, Md Tarikul Islam, Shihab Sharar, Annajiat Alim Rasel

Dept. of CSE, BRAC Uiversity
{ sm.ibrahim, mofaqkhayrul.islam, jahid.hasan.mamun, md.tarikul.islam, shihab.sharar }@g.bracu.ac.bd, annajiat@gmail.com

Day by day online education is gaining much popularity. Moreover, there has been a sudden significant increase in the adoption of online education due to the pandemic. Furthermore, a key part of online education is assessment. Because of the internet and digital data, it is almost effortless to be able to copy answers from different sources, or students may share their responses even to conceptual questions. Thus there can be lots of possibilities of getting similar answers with varying degrees of modifications. For the course instructor, it is very challenging to review the exam papers for such duplicate data, as these data are intentionally modified. This study looks at the similar answer detection in the Kaggle Question Answer dataset. This study extensively explores the dataset and utilizes machine learning models, including neural networks. For word embedding and detecting the semantic similarity of answers, Convolutional Neural Networks (CNN) have been used. A comparative study has been done to find out which neural network model gives the best performance followed by error analysis. These findings can help the evaluators in assessment in the online academic examinations.