The BoW representation had a dev set performance of 0.867 while the CNN model had a dev set performance of 0.827. The difference in their maximum dev set performances is slight, and neither model achieves a "good" score on the dev set. However, the BoW representation performs slightly better than the CNN model. There are a multitude of reasons why the CNN model doesn't perform as well as the BoW representation that are specific to CNNs e.g. learning rate, not enough data, etc. When comparing the two models on this dataset, the fact that the BoW representation that doesn't consider the order of the input words outperforms (albeit slightly) the CNN model that does consider the structure of the words could mean that the structure in each text is not relevant to its classification with these given labels.