Figure 4: Performances of DC-RecNN with the different sizes of latent vector z on five development datasets: IE, MR, SST, SUBJ, and QC. Y-axis represents the accuracy(%), and X-axis represents dimensionality of z.

Table 5: Multiple interpretable neurons and the words/phrases captured by these neurons. The last column gives the explanations of corresponding neuron's behaviours.

process of composition. Figure 5-(a) shows a visualization. We can see in this sentence, the neuron has realized that this idiomatic collocation "in stitches" is a key pattern, which is crucial for the final sentiment prediction.

- For more complicated tasks such as semantic matching, a well-grounded understanding of the syntactic structure is crucial. In this context, we find that a meta network could capture some syntactic information. For example, the $27\text{-}th$ neuron monitors phrases constructed by light-verb. As shown in Figure 5-(b), the verb phrase "taking off" has been attended for forthcoming compositional operation, which is more useful for judging the semantic relation between the sentence pair "An airplane is taking off/A plane is landing".

6 Related Work

One thread of related work is the exploration of different kinds of compositional function over tree structures. Socher et al. [2012] proposed the recursive neural network with standard compositional function. After that, some extensions are introduced to enhance the expressive power of compositional function, such as MV-RecNN [Socher et al., 2013b], SU-RNN [Socher et al., 2013a], RNTN [Socher et al., 2013b], while these models suffer from the problem of hard-coded compositional operations and overfitting.

Another thread of work is the idea of using one network to direct the learning of another network [De Brabandere et al., 2016]. Naik and Mammone [1992] introduce a meta neural network to provide another network with a step size and a direction vector, which is helpful for parameter optimization. De Brabandere et al. [2016] propose the dynamic filter
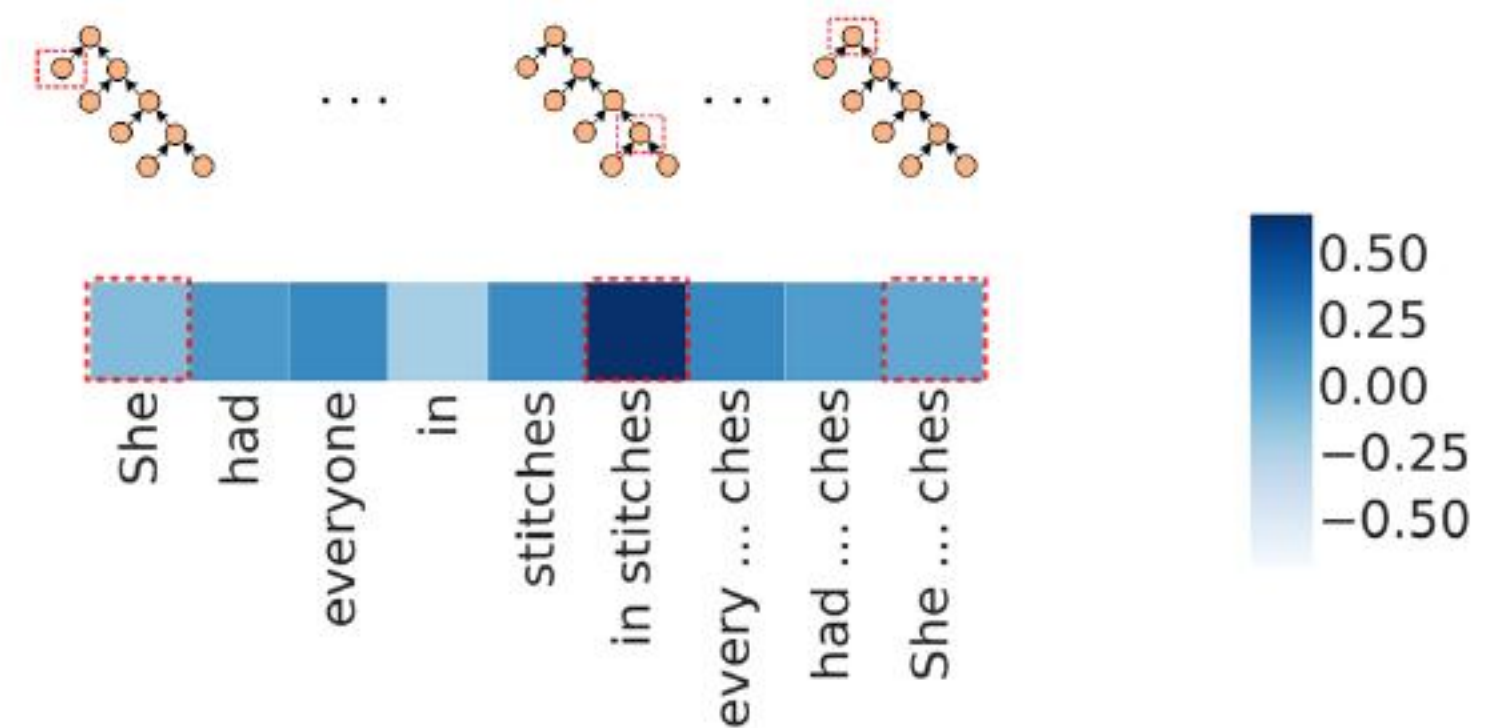
(a) The behaviour of $21\text{-}st$ neuron for sentence "She had everyone in stitches"
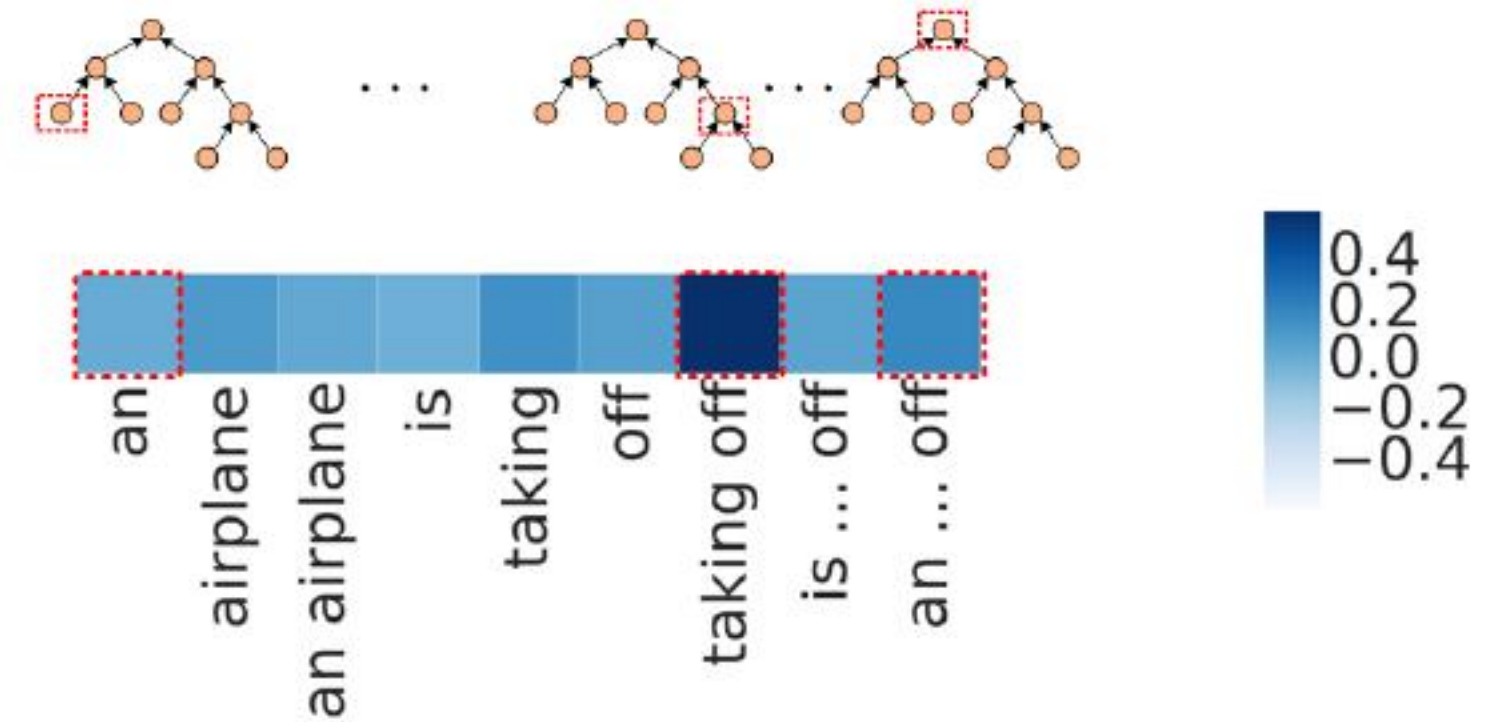
(b) The behaviour of $27\text{-}th$ neuron for sentence "An airplane is taking off"

Figure 5: The two heat maps describe the behaviours of neurons $\mathbf{z}_{21}$ and $\mathbf{z}_{27}$ from DC-TreeNN.

network to implicitly learn a variety of filtering operations. Bertinetto et al. [2016] introduce a learnet for one-shot learning, which can predict the parameters of a second network given a single exemplar. Ha et al. [2016] propose the model hypernetwork, which uses a small network to generate the weights for a larger network.

Different from these models, we employ the idea of param-