2 METHOD

The proposed model is a single, end-to-end ranking-based algorithm with a multi-layer perceptron. The model input is a pair of headlines and the article bodies. We use stance labels to help compute the value of the objective during the training phase. The multi-layer perceptron with two hidden layers produces a value u^+ for the true stance and three values u_1^-, u_2^-, u_3^- for three false stances with the goal to satisfy this constraint: $u^+ \geq \max\{u_1^-, u_2^-, u_3^-\} + 1$. The objective of the loss function is to maximize the value difference between the true and false stances. During testing, given a news body and its headline the model assigns a particular stance (predicted label) to it, on which the model produces the highest value.

2.1 Feature Space

We extract TF-IDF features to represent both headlines and bodies of the news articles: TF-IDF of the headlines, and TF-IDF of the bodies. We also consider cosine similarity between the headline and the body of each article, and those extracted by the official baseline as our features. All features are concatenated to our input vector, denoted as $\mathbf{v}$, that is then fed into the multi-layer perception.

2.2 Multi-layer Perception

The multi-layer perception consists of two hidden layers with 100 units and 4 units, respectively. The second hidden layer produces a four-dimensional vector $\mathbf{u}$ with each dimension indicating a value for the corresponding stance respectively,

$$\mathbf{u} = \mathbf{W}_2^\top \cdot \mathrm{ReLu}(\mathbf{W}_1^\top \cdot \mathbf{v} + \mathbf{b}_1) + \mathbf{b}_2 , \tag{1}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are the weight matrices, $\mathbf{b}_1$ and $\mathbf{b}_2$ are the bias vectors in the first and the second hidden layer, respectively; and the rectified linear unit (ReLu) is applied to the hidden layers as the non-lineaer activation function. Then the ranking loss function can be defined as follows:

$$L(\mathbf{u}) = \max(0, 1 - u^+ + \max(u_1^-, u_2^-, u_3^-)) , \tag{2}$$

where $u^+ = u_i$ with i being the index of the true stance for the news, and $\{u_1^-, u_2^-, u_3^-\} \equiv \{u_j\}_{j=1}^4 / \{u_i\}$. Here $\{u_j\}_{j=1}^4$ are the elements in $\mathbf{u}$. The goal of the training is to enlarge the difference between u^+ and the maximum in $\{u_1^-, u_2^-, u_3^-\}$ for all the news articles.

2.3 Experimental Setting

Following the official setting, we split the data set (75,385 news articles) into training, validation and testing subsets. [5] We apply dropout and include an L2 regularization for the MLP weights in the loss function to mitigate overfitting. We train in mini-batches over the entire training set with the Adadelta optimizer.

The evaluation is based on a weighted, two-layer scoring system. The "related"/"unrelated" classification is given 25% score weighting in the *accuracy* evaluation metric because it is easier. The classification between "agree", "disagree" and "discuss" is given 75% score weighting since it is more difficult and relevant to fake news detection. The scoring system produces a *relative score* based on the differentially weighted scoring metric.

[5]Data set is publicly available from https://github.com/FakeNewsChallenge/fnc-1.

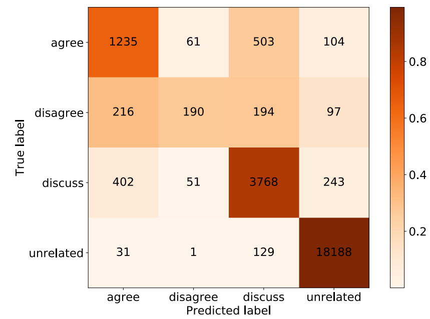

Figure 1: The heat map of detection results.

Table 1: Performance comparison with the state-of-art algorithms on the FNC-1 competition dataset.

Team	Accuracy (%)				Relative Score(%)
	agree	disagree	discuss	unrelated	
Official baseline	9.09	1.00	79.66	97.98	75.20
Chips Ahoy!	55.96	0.29	70.30	98.99	80.21
UCLMR	44.04	6.60	81.38	97.90	81.72
Athene	44.72	9.47	80.89	99.25	81.97
Talos	58.54	1.87	76.19	98.70	82.02
This work	64.90	27.26	84.41	99.12	86.66

3 RESULTS

As for the FNC-1 official evaluation metric, our method achieves an overall relative score of 86.66%. Fig. 1 demonstrates the performance with a heat map and Table 1 compares our method with the state-of-the-art. Our method significantly achieves the best performance and improves detection accuracy. When separating "unrelated" pairs of headlines and bodies, we got over 99% accuracy. Besides, the detection accuracy for "discuss" is improved by 3.72%. Importantly, our method successfully improves the pretty limited accuracy of "agree" and "disagree" by 10.86% and 187.86%, respectively. This is the most attracting point as "agree" and "disagree" are the most difficult to detect by previous methods and the most relevant to the superordinate goal of automating the stance detection process.

4 CONCLUSION

We tackle the news stance detection task using a ranking-based method rather than classification-based algorithms. Given a pair of headlines and article bodies, the ranking-based method can compare and maximize the difference between the true and false stances. Though the architecture is simple, it successfully improves stance detection performance to a large extent.

5 ACKNOWLEDGMENTS

This research was supported by the China Scholarship Council (CSC) for award No. 201704910908.

REFERENCES

[1] William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*. 1163–1168.

[2] Benjamin Riedel, Isabelle Augenstein, Georgios P Spithourakis, and Sebastian Riedel. 2017. A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. *arXiv preprint arXiv:1707.03264* (2017).

[3] Jen Weedon, William Nuland, and Alex Stamos. 2017. Information operations and Facebook. *version* 1 (2017), 27.