

IoH-RCNN: Pursuing the Ingredients of Happiness using Recurrent Convolutional Neural Networks

Bashar Talafha¹ and Mahmoud Al-Ayyoub¹[0000-0001-9372-9076]

Jordan University of Science and Technology, Irbid 22110, Jordan
talafha@live.com, maalshbool@just.edu.jo

Abstract. Modeling human affect is non-trivial. To undertake this challenge, a novel shared task focusing on happiness is organized at an AAAI workshop. The CL-AFF Shared Task, titled “In Pursuit of Happiness”, consists of two sub-tasks on a dataset of descriptions of happy moments (taken from the HappyDB dataset), each annotated with individuals’ demographics, recollection time and relevant labels. We focus on the first sub-task, which is a semi-supervised task to determine a happy moment’s agency and social label. We present a deep learning system for this task based on Recurrent Convolutional Neural Networks (RCNN). The presented system (which we call IoH-RCNN) is trained and tested on the available dataset using 10-fold cross-validation. For predicting the agency label, the average accuracy, f1 and AUC are 85.5, 90.3 and 80.0, respectively. As for predicting the social label, the average accuracy, f1 and AUC are 91.8, 92.2 and 91.2, respectively.

Keywords: Sentiment Analysis · HCI · Affective Computing · Psychology · Deep Learning · Recurrent Neural Networks · Convolutional Neural Networks.

1 Introduction

Understanding user expression is very challenging. To tackle this issue, the organizers of the Affective Content Analysis (AffCon) workshop at AAAI 2019 propose a shared task with the goal of providing better modeling for human affect. The task, which is the first of its kind, focuses on happiness and is titled “In Pursuit of Happiness” [7].

The task offers two datasets (a small labeled one and a large unlabeled one) taken from the HappyDB dataset [1] with two sub-tasks. The first one, which is a semi-supervised problem, has the goal of predicting the agency and social labels of an account of a happy moment. On the other hand, the second one has an open-ended flavor as it requires defining new characterizations and insights for happy moments. Obviously, this is an unsupervised problem by nature [7].

We focus on the first sub-task and address it using Neural Networks (NN). Specifically, we propose a system based on the Recurrent Convolutional NN (RCNN) architecture of Lai et al. [9]. RCNN uses a recurrent structure, which is

a Bidirectional Recurrent NN (Bi-RNN) that captures the contexts for predicting agency and social labels for happy moments. Another feature of RCNN that makes it suitable for the problem at hand is the use of a max-pooling layer to determine which words are important for classification. The proposed approach is compared with several baseline methods such as Bidirectional Long Short-Term Memory (Bi-LSTM) networks, Bi-LSTM with Attention (Att-BLSTM) [14], CNN for Sentence Classification [8] and Google’s Transformer model [11, 10].

It is worth mentioning that the “In Pursuit of Happiness” tasks are quite unique. In fact, we could not find prior work addressing the semi-supervised task, which is the task under consideration in this work. Kindly, remember that the focus here is on predicting the agency and social labels of happy moments each described in a single sentence written in English and is accompanied with information related to the author (age, gender, etc.) and the happy moment (e.g., duration, concept, etc.). Thus, works focusing on modalities other than text or addressing affective analysis from a different perspective or with different objectives are outside the scope of this paper. Examples include [12, 5, 13, 3, 6]

The rest of this paper is organized as follows. In the following section, the “In Pursuit of Happiness” task is described. The proposed IoH-RCNN approach is presented in Section 3 along with its experimental evaluation. Finally, the paper is concluded in Section 4 with main findings and thoughts on future directions of this work.

2 CL-Aff Shared Task: In Pursuit of Happiness

The task at hand, which is titled “In Pursuit of Happiness” is presented in this section in a brief way. Interested readers are referred to the task description paper [7] and website¹ for more details.

The common practices when it comes to emotion analysis do not necessarily capture the experiential, contextual and agentic attributes of happy moments. Since human affect, in general, is context-driven, labeled datasets must account for these factors in generating predictive models of affect. This was the main motivation behind offering the CL-Aff shared task at the AffCon workshop of AAAI 2019. Based on the HappyDB dataset [1], the dataset offered in this task has 100k happy moments each described by a single sentence. It is comprised of two sets of training records: a small labeled set (10k records) and a large unlabeled set (70k records) and a set of testing records (17k records).

According to the organizers, the labeling is done by annotating each record with labels that identify the ‘agency’ of the author and the ‘social’ characteristic of the moment, as well as concept labels describing its theme. The two labels that will be used for classification are both binary ones. They are agency and social. The agency label is whether the author is in control or not, whereas, the social label is whether the described moment involves other people other than the author or not.

¹ <https://sites.google.com/view/affcon2019/cl-aff-shared-task>

The information made available with each happy moment include information about the author as well as information about the happy moment itself. The former include the age, gender, country, marital status, parenthood status and demographics of the author, while the latter include the reflection period information and the duration of the happy moment. Finally, each moment is annotated with 1-4 concepts, where the available concepts are 15. One example of concepts assigned to a single happy moment is: “family|education|party”.

The shared task has two components: a semi-supervised one and an unsupervised one. For the former component, the participating teams can use both the labeled and unlabeled sets to predict the agency and social labels of each happy moment in the test set. As for the unsupervised task, it is concerned with modeling happiness. I.e., it asks for new characterizations and insights for the testing happy moments (e.g., in terms of affect, emotion, participants and content.)

The evaluation is done in a simple way. For the semi-supervised task, the evaluation metrics are simple accuracy, f1 measure and AUC (Area Under the ROC Curve). Each metric is applied to each of the two binary labels under consideration (agency and social). On the other hand, due to the open-ended nature of the unsupervised task, the evaluation metric is not defined. We note that we are only interested in addressing the semi-supervised task in this paper.

3 IoH-RCNN: Model and Evaluation

A wide variety of text classification tasks in the field of natural language processing (NLP), such as topic identification, spam filtering and sentiment analysis can be considered supervised learning problems and, thus, can be addressed with a text classifier. Many semi-supervised algorithms that use labeled and unlabeled data have been proposed to enhance supervised classification and improve the model robustness by more precise decision boundary. One of the most successful ones is Lai et al. [9]’s RCNN. In this section, we discuss our Ingredients of Happiness with RCNN (IoH-RCNN) system to address the semi-supervised task described in the previous section.

The structure of the IoH-RCNN model is presented in Figure 1. The input for the model is a sentence consisting of a sequence of words $S = w_1, \dots, w_n$ describing the happy moment along with the set of features provided by the task organizers (author gender, moment duration, context, etc.) and the output is the label value $\{0, 1\}$. Note that we built different RCNN for each problem under consideration (i.e., one of predicting agency and another one to predict social). The textual part of the input is handled through RCNN, which produces an output vector called $y^{(3)}$. As for the other parts of the input, which are the supplied (external) features, they are fed into a Feed Forward NN (FFNN), that produces an output vector called ff . The two output vectors are then concatenated and fed into the output layer. Finally, the loss function we use is softmax cross entropy.

In order to get a more precise word embedding, words are combined with their contexts. The contexts are captured by using a Bi-RNN, which allows for

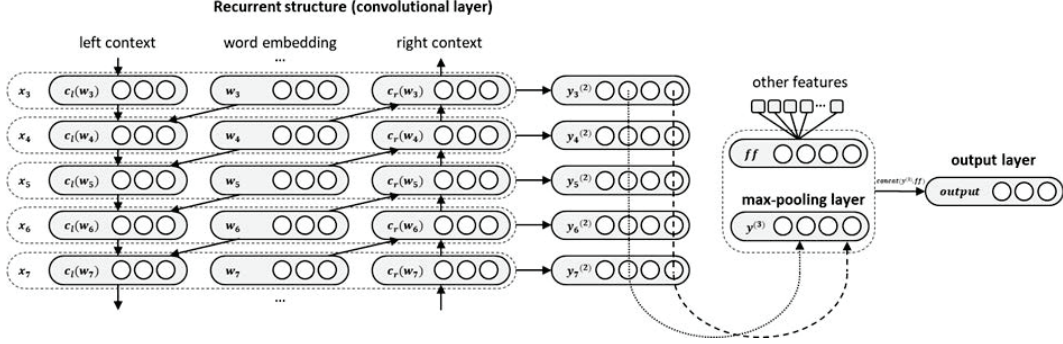


Fig. 1. Architecture of the IoH-RCNN model.

better disambiguation of each word’s meaning. This gives RCNN an advantage over traditional NN which captures context information only within a fixed window. The recurrent structure captures the left and right context information by performing a forward scan/pass followed by a backward scan/pass over the input.

To be more specific, we explain how RCNN work using the same notation used in the original paper [9]. We start with the recurrent structure. The left and right contexts of a word w_i are represented using the dense vectors $c_l(w_i)$ and $c_r(w_i)$, respectively. These vectors are computed as follows.

$$\begin{aligned} c_l(w_i) &= f(W^{(l)}c_l(w_{i-1}) + W^{(sl)}e(w_{i-1})) \\ c_r(w_i) &= f(W^{(r)}c_r(w_{i+1}) + W^{(sr)}e(w_{i+1})) \end{aligned}$$

where $e(w_i)$ is the word embedding of w_i and $f(\cdot)$ is a non-linear activation function. The matrix $W^{(l)}$ is used to transform the hidden layer (context) into the next hidden layer, while the matrix $W^{(sl)}$ is used to combine the semantic of the current word with the next word’s left context. The matrices $W^{(r)}$ and $W^{(sr)}$ are used similarly, but for the right context. In order to put things together, the left and right context vectors are concatenated with the word embedding vector to produce the representation of w_i , which is denoted as x_i .

$$x_i = [c_l(w_i); e(w_i); c_r(w_i)]$$

The word representations go through a linear transformation before being moved to the next layer.

$$y_i^{(2)} = \tanh(W^{(2)}x_i + b^{(2)})$$

The latent semantic vector $y_i^{(2)}$ is where each semantic factor will be analyzed to determine the most useful factor for representing the text.

The previous paragraph explained the recurrent structure of RCNN. Now, we explain the CNN part, which views recurrent structure as the convolutional

layer. After calculating word representations and latent semantic vector, a max-pooling operation is applied to capture the information throughout the entire text.

$$y^{(3)} = \max_{i=1}^n y_i^{(2)}$$

Here, the max operation is performed in an element-wise fashion.

3.1 Empirical Evaluation

The proposed system (IoH-RCNN) is implemented on Google’s TensorFlow. It is evaluated and tested on the provided data using 10-fold cross validation. The results for the agency label are as follows.

- Average accuracy= 85.5
- Average f1= 90.3
- Average AUC= 80.0

As for the Social label, the results are as follows.

- Average accuracy= 91.8
- Average f1= 92.2
- Average AUC= 91.2

As mentioned earlier, we compare IoH-RCNN with three baseline systems: Bi-LSTM, Att-BLSTM and CNN. We experiment with different configurations for the parameters of these systems and report the best results. The configurations that produce the best results are reported in Table 1.

Table 1. Configurations of baseline systems.

Bi-LSTM	Embedding layer Bidirectional RNN layer Concat all the outputs from RNN layer Fully-connected layer embedding_size = 256 num_hidden = 256 num_layers = 2 learning_rate = 1e-3
Att-BLSTM	embedding_size = 256 num_hidden = 256 num_layers = 2 learning_rate = 1e-3 Attention = Bahdanu [2]
CNN	embedding_size = 128 learning_rate = 1e-3 filter_sizes = [3, 4, 5] num_filters = 100

A comparison of the four models’ accuracies is shown in Table 2. The table shows the superiority of IoH-RCNN over baseline systems in predicting both the agency and social labels.

Table 2. Comparing the average accuracy of IoH-RCNN with baseline systems.

	Agency	Social
IoH-RCNN	85.5	91.8
Bi-LSTM	82.6	89.9
Att-BLSTM	84.3	89.2
CNN	85.2	89.8

We note that we also experiment with other innovative approaches such as the Transformer model. However, the results we obtain are not high enough compared with those of the IoH-RCNN system.

4 Conclusion

In this paper, we addressed the semi-supervised sub-task of the CL-AFF Shared Task, titled “In Pursuit of Happiness”. Specifically, the problem is: given a small set of happy moments descriptions labeled with agency and social labels and a large unlabeled set of happy moments descriptions, the goal is to build a model to predict the agency and social labels of the test set. The happy moments are annotated with information related to the author (age, gender, etc.) and the happy moment (e.g., duration, concept, etc.). The model we presented to solve this problem is based on the exciting RCNN. The obtained results were high (with accuracies of 85.5 and 91.8 for the agency the social labels, respectively) compared with other methods such as Bi-LSTM, Attn-BLSTM, CNN and Transformer (whose best accuracies for the agency the social labels were 85.2 and 89.9, respectively). In the future, we plan on exploring more cutting edge techniques such as Ensemble Classification, Transfer Learning and Bidirectional Encoder Representations from Transformers (BERT) [4].

Acknowledgment

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

1. Asai, A., Evensen, S., Golshan, B., Halevy, A., Li, V., Lopatenko, A., Stepanov, D., Suhara, Y., Tan, W.C., Xu, Y.: Happydb: A corpus of 100,000 crowdsourced happy moments. In: Proceedings of LREC 2018. European Language Resources Association (ELRA), Miyazaki, Japan (May 2018)

2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
3. Cambria, E., Fu, J., Bisio, F., Poria, S.: Affectivespace 2: Enabling affective intuition for concept-level sentiment analysis. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Ding, H., Jiang, T., Riloff, E.: Why is an event affective? classifying affective events based on human needs. In: The Workshops of the The Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018. pp. 8–15 (2018)
6. Jaidka, K., Chhaya, N., Wadbude, R., Kedia, S., Nallagatla, M.: Batframe: An unsupervised approach for domain-sensitive affect detection. In: International Conference on Computational Linguistics and Intelligent Text Processing. pp. 20–34. Springer (2017)
7. Jaidka, K., Mumick, S., Chhaya, N., Ungar, L.: The CL-Aff Happiness Shared Task: Results and Key Insights. In: Proceedings of the 2nd Workshop on Affective Content Analysis @ AAAI (AffCon2019). Honolulu, Hawaii (January 2019)
8. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
9. Lai, S., Xu, L., Liu, K., Zhao, J.: Recurrent convolutional neural networks for text classification. In: AAAI. vol. 333, pp. 2267–2273 (2015)
10. Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A.N., Gouws, S., Jones, L., Kaiser, L., Kalchbrenner, N., Parmar, N., et al.: Tensor2tensor for neural machine translation. arXiv preprint arXiv:1803.07416 (2018)
11. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems. pp. 5998–6008 (2017)
12. Yang, J., She, D., Lai, Y.K., Yang, M.H.: Retrieving and classifying affective images via deep metric learning. In: AAAI (2018)
13. Zhou, H., Huang, M., Zhang, T., Zhu, X., Liu, B.: Emotional chatting machine: Emotional conversation generation with internal and external memory. arXiv preprint arXiv:1704.01074 (2017)
14. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). vol. 2, pp. 207–212 (2016)