Deep Learning-Based Document Modeling for Personality Detection from Text

IEEE Intelligent Systems

Navonil Majumder, Instituto Politécnico Nacional
Soujanya Poria, Nanyang Technological University
Alexander Gelbukh, Instituto Politécnico Nacional
Erik Cambria, Nanyang Technological University

Big Five Personality Traits

Personality is typically formally described in terms of the Big Five personality traits, which are the following binary (yes/no) values:

- ***Extroversion* (EXT).** Is the person outgoing, talkative, and energetic versus reserved and solitary?
- ***Neuroticism* (NEU).** Is the person sensitive and nervous versus secure and confident?
- ***Agreeableness* (AGR).** Is the person trustworthy, straightforward, generous, and modest versus unreliable, complicated, meager, and boastful?
- ***Conscientiousness* (CON).** Is the person efficient and organized versus sloppy and careless?
- ***Openness* (OPN).** Is the person inventive and curious versus dogmatic and cautious?

Overview of the Method

The method includes:

- ***Preprocessing –*** sentence splitting, data cleaning and unification such as reduction to lower case
- ***Document-level feature extraction -*** used the Mairesse baseline feature set, which includes such global features as the word count and average sentence length
- ***Filtering –*** Removed sentences with noise, does not carry any personality clues
- ***Word-level feature extraction –*** represent individual words by word embedding in a continuous vector space
- ***Classification –*** used deep Convolutional Neural Network (CNN)

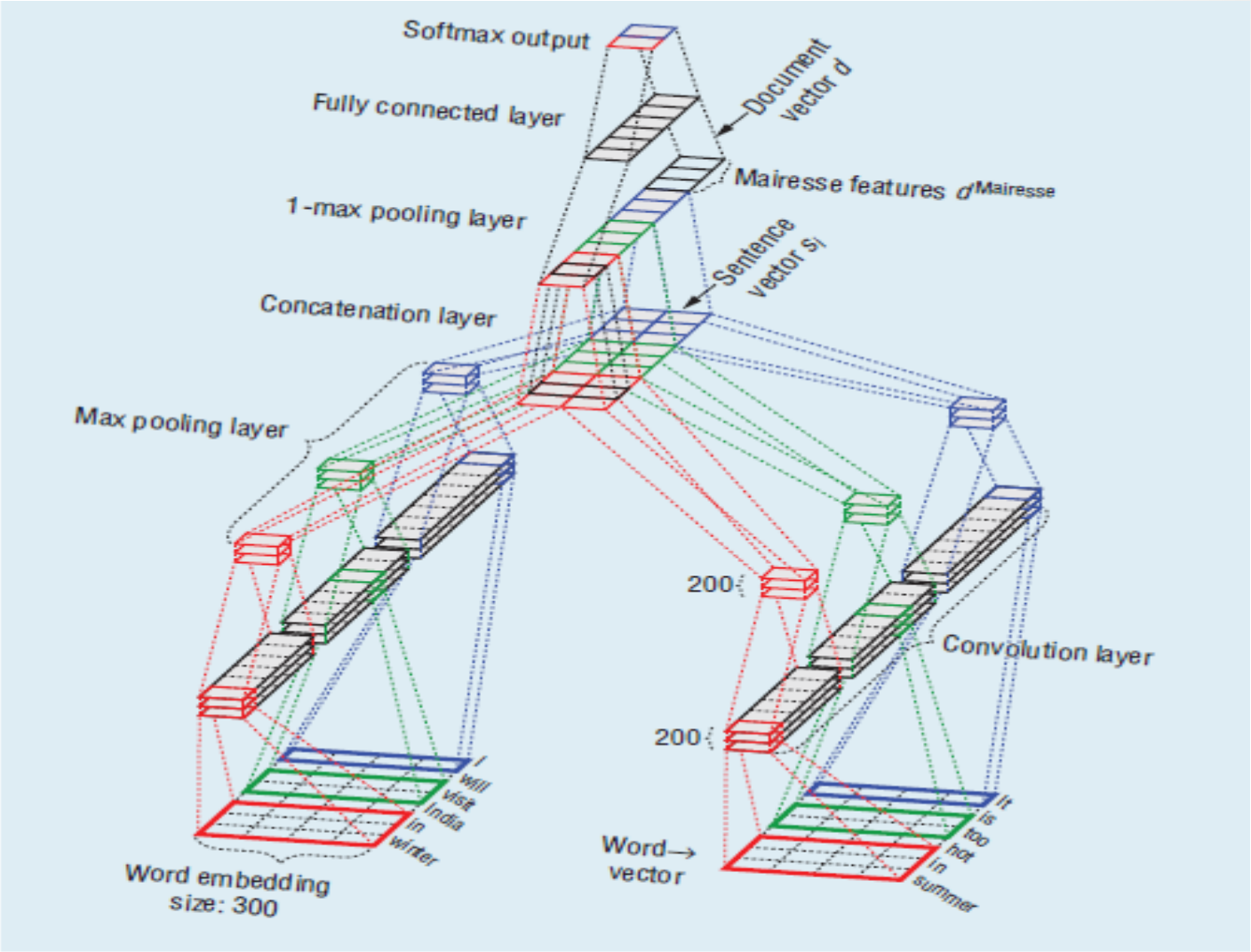

Figure 1. Architecture of our network. The network consists of seven layers. The input layer (shown at the bottom) corresponds to the sequence of input sentences (only two are shown). The next two layers include three parts, corresponding to trigrams, bigrams, and unigrams. The dotted lines delimit the area in a previous layer to which a neuron of the next layer is connected—for example, the bottom-right rectangle shows the area comprising three word vectors connected with a trigram neuron.

Table 1. Accuracy obtained with different configurations.

Document vector *d*	Filter	Classifier	Convolution filter	Personality traits				
				EXT	NEU	AGR	CON	OPN
N/A	N/A	Majority	N/A	51.72	50.02	53.10	50.79	51.52
Word *n*-grams	Not used	SVM	N/A	51.72	50.26	53.10	50.79	51.52
Mairesse[12]	N/A	SVM	N/A	55.13	58.09	55.35	55.28	59.57
Mairesse (our experiments)	N/A	SVM	N/A	55.82	58.74	55.70	55.25	60.40
Published state of the art per trait[12]	N/A	N/A	N/A	56.45	58.33	56.03	**56.73**	60.68
CNN	N/A	MLP	1, 2, 3	55.43	55.08	54.51	54.28	61.03
CNN	N/A	MLP	2, 3, 4	55.73	55.80	55.36	55.69	61.73
CNN	N/A	SVM	2, 3, 4	54.42	55.47	55.13	54.60	59.15
CNN + Mairesse	N/A	MLP	1, 2, 3	54.15	57.58	54.64	55.73	61.79
CNN + Mairesse	N/A	SVM	1, 2, 3	55.06	56.74	53.56	56.05	59.51
CNN + Mairesse	N/A	sMLP/FC	1, 2, 3	54.61	57.81	55.84	**57.30**	62.13
CNN + Mairesse	Used	sMLP/MP	1, 2, 3	**58.09**	57.33	**56.71**	56.71	61.13
CNN + Mairesse	Used	MLP	1, 2, 3	55.54	58.42	55.40	56.30	**62.68**
CNN + Mairesse	Used	SVM	1, 2, 3	55.65	55.57	52.40	55.05	58.92
CNN + Mairesse	Used	MLP	2, 3, 4	55.07	**59.38**	55.08	55.14	60.51
CNN + Mairesse	Used	SVM	2, 3, 4	56.41	55.61	54.79	55.69	61.52
CNN + Mairesse	Used	MLP	3, 4, 5	55.38	58.04	55.39	56.49	61.14
CNN + Mairesse	Used	SVM	3, 4, 5	56.06	55.96	54.16	55.47	60.67

*Bold indicates the best result for each trait.

Results/ Future Work

Table 1 in the previous slide shows the results.

- This method outperformed the state of the art for all five traits, although with different configurations for different traits
- Applying filtering and adding the document level (Mairesse) features proved to be beneficial.
- In the future, this team plans to incorporate more features and preprocessing by applying the Long Short Term Memory (LSTM) recurrent network to build both the sentence vector from a sequence of word vectors and the document vector from a sequence of sentence vectors.
- In addition, this team plans to apply their document modeling technique to other emotion related tasks, such as sentiment analysis or mood classification

Thank you..
Questions?