Personality Detection from Text

Peng Zhou Yidan Weng Mahdi Rahimi

CMPS 242 Final Project

We would like to use neural networks

to detect the personality of authors

Big Five Model

- The most widely-used personality traits model in the literature
- Five broad personality dimensions:
  - Openness: inventive and curious
  - Conscientiousness: efficient and organized
  - Extraversion: outgoing, talkative, and energetic
  - Agreeableness: trustworthy, straightforward, generous
  - Neuroticism: sensitive and nervous

Overview

- We train six neural networks with the same architecture for each of the five personality traits plus the multi-label classification.
- We use different models for document modeling and report the results.

Approach

- Preprocessing: Sentence splitting, data cleaning, and sentence filtering
- Document Modeling (Feature Extraction)
    1) Hierarchical CNN + Mairesse
    2) LSTM + Mairesse
    3) Hierarchical RNN model (LSTM with Maxpooling) + Mairesse
    4) Hierarchical RNN model (2 layer LSTM) + Mairesse
    5) Mairesse baseline feature set
- Classification
    a. Fully connected network
    b. Logistic Regression

Dataset

- There are two state-of-the-art and most popular data sets for personality detection:
  - Essay dataset
  - myPersonality dataset: collected from Facebook

Essay Dataset:

- 2,467 Essays one for each author
- Collected between 1997 and 2004 and labeled with personality classes

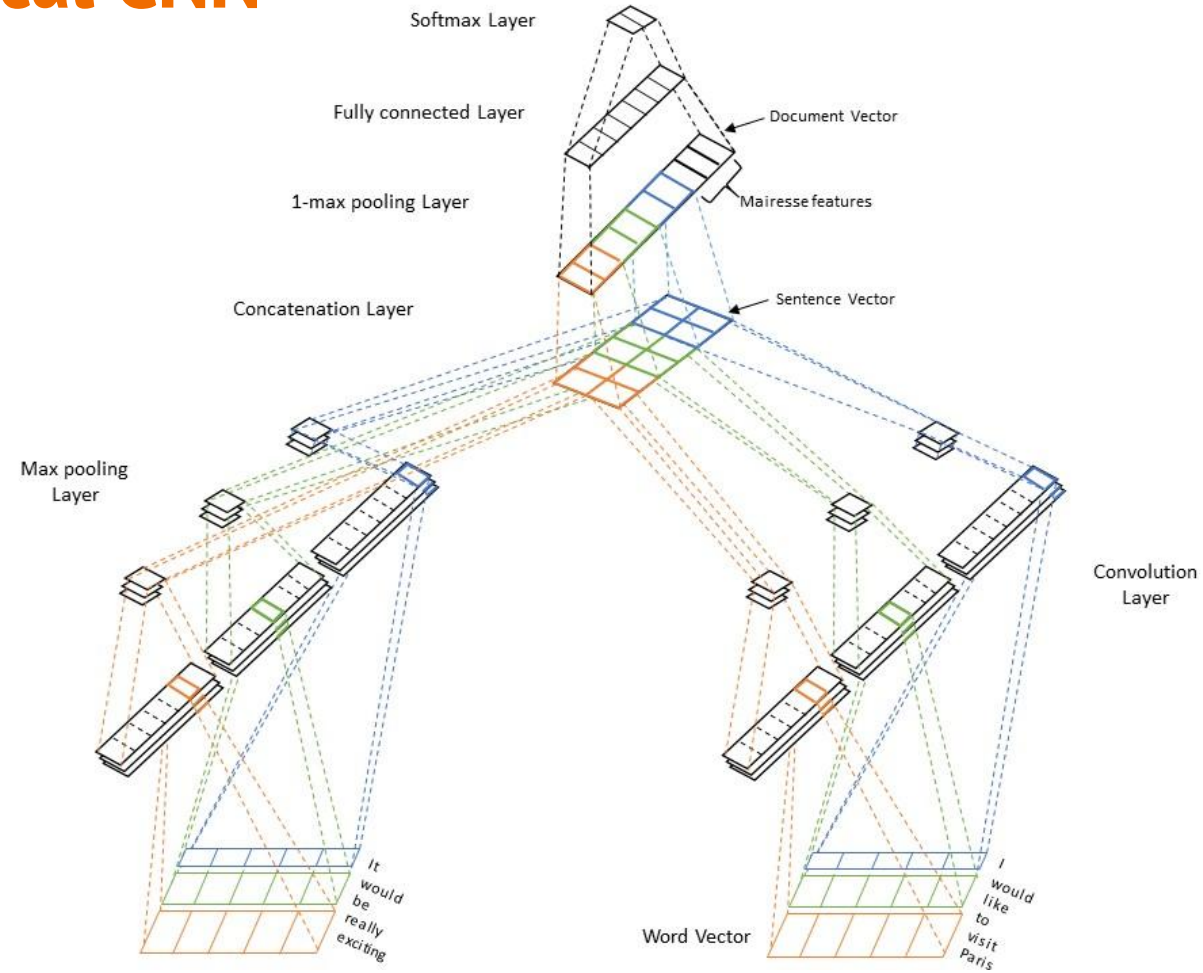
Well, here we go with the stream of consciousness essay. I used to do things like this in high school sometimes. They were pretty interesting, but I often find myself with a lack of things to say. I normally consider myself someone who gets straight to the point. I wonder if I should hit enter any time to send this back to the front. Maybe I'll fix it later. My friend is playing guitar in my room now. Sort of playing anyway. More like messing with it. He's still learning. There's a drawing on the wall next to me. Comic book characters I think, but I'm not sure who they are. It's been a while since I've kept up with comic's. I just heard a sound from ICQ. That's a chat program on the internet. I don't know too much about it so I can't really explain too well. Anyway, I hope I'm done with this by the time another friend comes over. It will be nice to talk to her again. She went home this weekend for Labor Day. So did my brother. I didn't go. I'm not sure why. No reason to go, I guess. Hmm. when did I start this. Wow, that was a long line. I guess I won't change it later. Okay, I'm running out of things to talk about. I've found that happens to me a lot in conversation. Not a very interesting person, I guess. Well, I don't know. It's something I'm working on. I'm in a class now that might help. The phone just rang. Should I get it? The guy playing the guitar answered it for me. It's for my roommate. My suitemate just came in and started reading this. I'm uncomfortable with that. He's in the bathroom now. You know, this is a really boring piece of literature. I never realized how dull most everyday thoughts are. Then again, when you keep your mind constantly moving like this, there isn't really time to stop and think deeply about things. I wonder how long this is going to be. I think it's been about ten minutes now. Only my second line. How sad. Well, not really considering how long these lines are. Anyway, I wonder what I'm going to do the rest of the night. I guess there's always homework to do. I guess we'll see. This seat is uncomfortable. My back sort of hurts. I think I'm going to have arthritis when I get older. I always thought that I wouldn't like to grow old. Not too old, I suppose. I've always been a very active person. I have a fear of growing old, I think. I guess it'll go away as I age gradually. I don't know how well I'd deal with paralysis from an accident though. As long as I have God and my friends around, I'll be okay though. I'm pretty thirsty right now. There isn't much to drink around my room. Ultimate Frisbee, I haven't played that all summer. Fun game, but tiring. I'm out of shape. I'd like to get in better shape, but I hate running. It's too dull for me. Hmmm. it's almost over now. Just a few more minutes. Let's see if I make it to the next line. Short reachable goals! Whatever. Anyway, what else do I have to do tonight. I guess I could read some. My shirt smells like dinner. It's pretty disgusting. I need to wake up for a 9:30 am class tomorrow. I remember when that wasn't early at all. Well, I made it to the next line. I'm so proud of myself. That's sarcasm, by the way. I wonder if I was suppose to right this thing as a narrative. Oh well too late now. Time for me to head out. Until next time, good bye and good luck. I don't know.

cEXT	cNEU	cAGR	cCON	cOPN
n	y	y	n	y

Word Embedding

- Use 300-dimensional GloVe to initialize the word embedding matrix
- Train the word embedding matrix along with the rest of the network

Hierarchical CNN



Example from essays

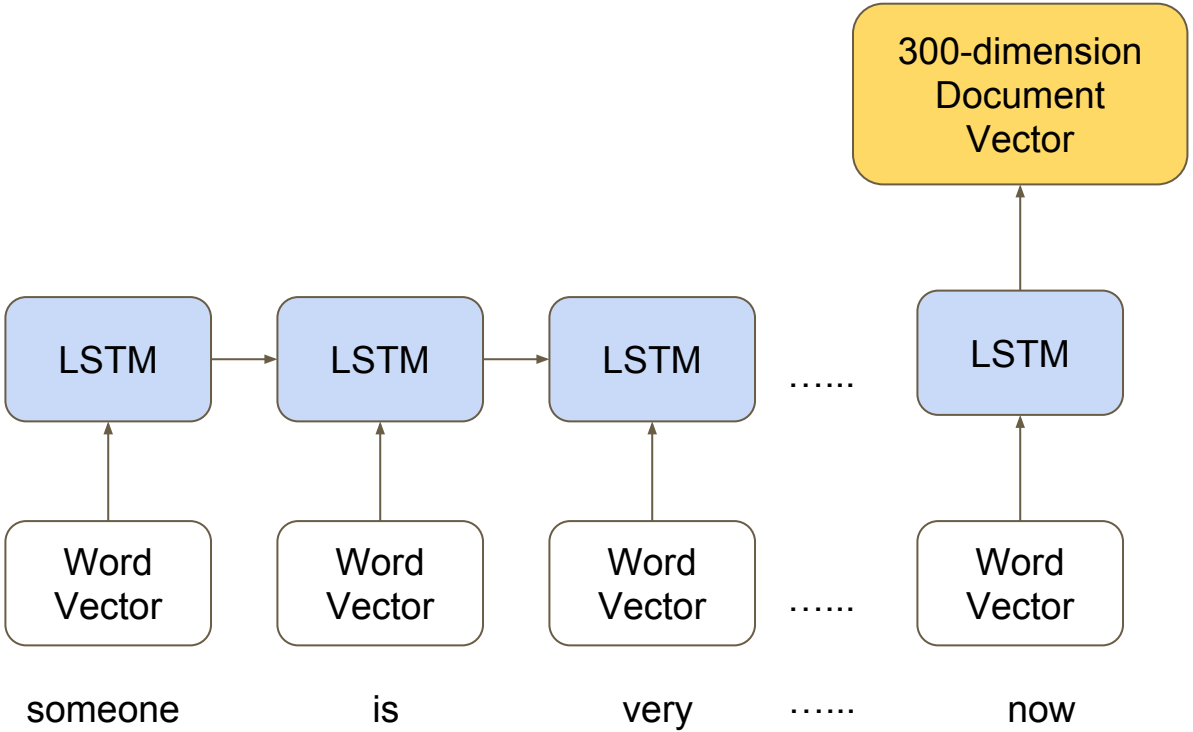
[[Someone is very loud in the hall]

[They sound like they are having fun]

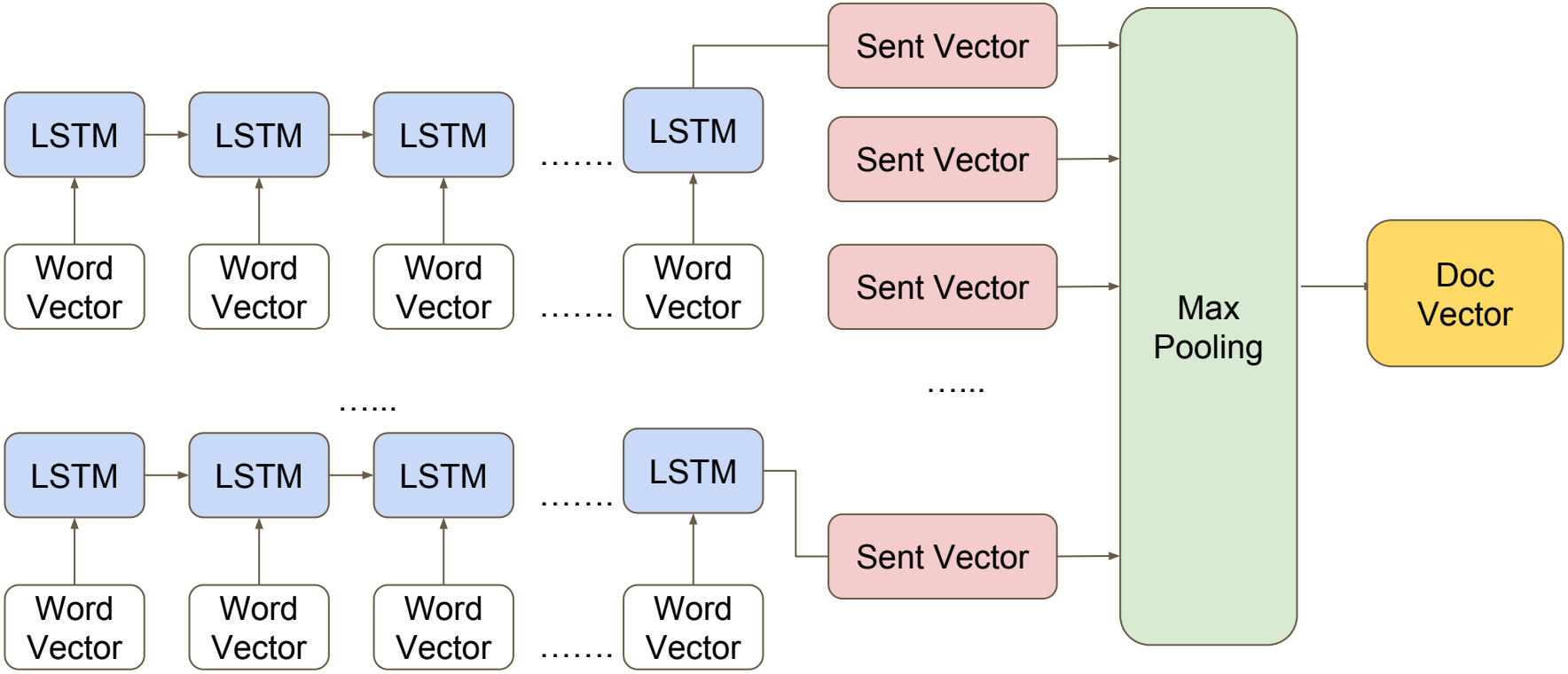
[Wish I could join them]

[Is Ben thinking about me right now]]

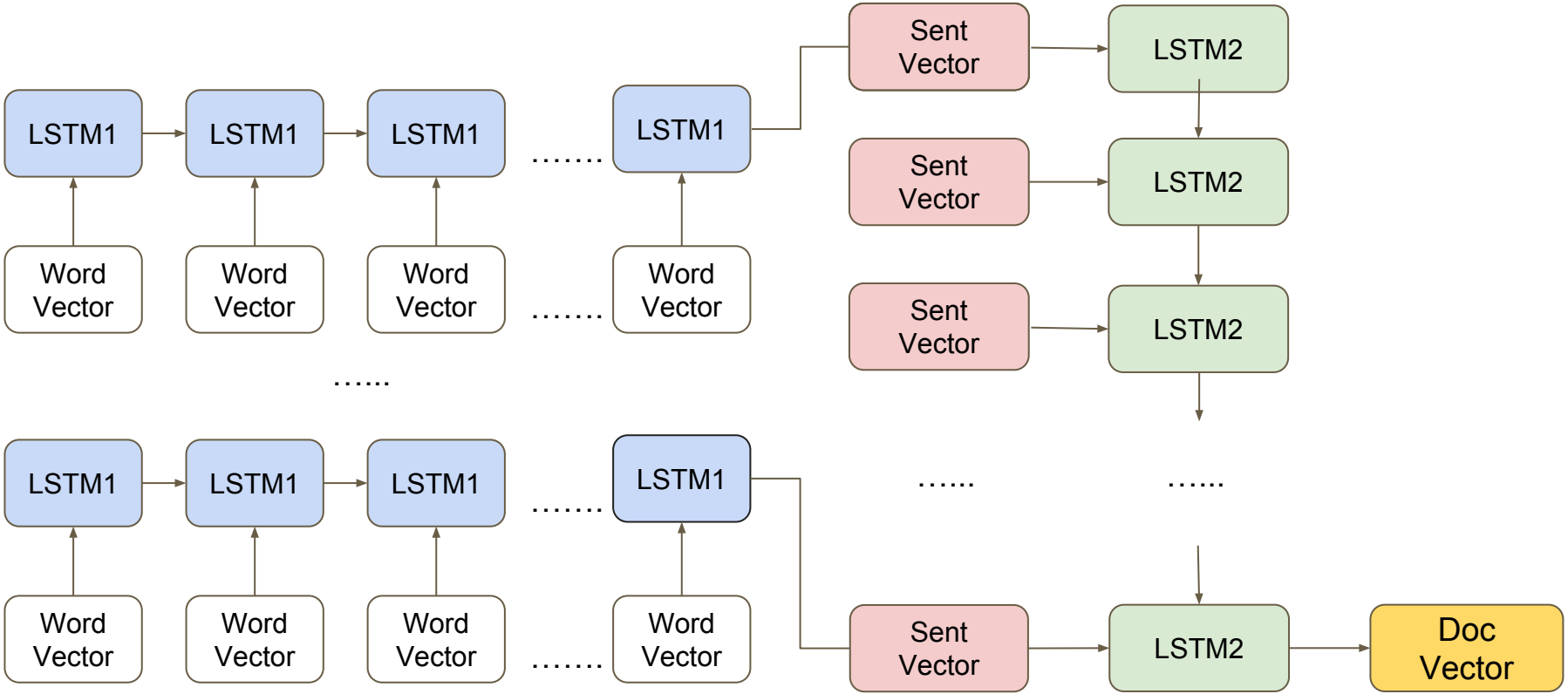
LSTM


300-dimension
Document
Vector
LSTM
LSTM
LSTM
.....
LSTM
Word
Vector
Word
Vector
Word
Vector
.....
Word
Vector
someone
is
very
.....
now

LSTM with Maxpooling



Hierarchical LSTM

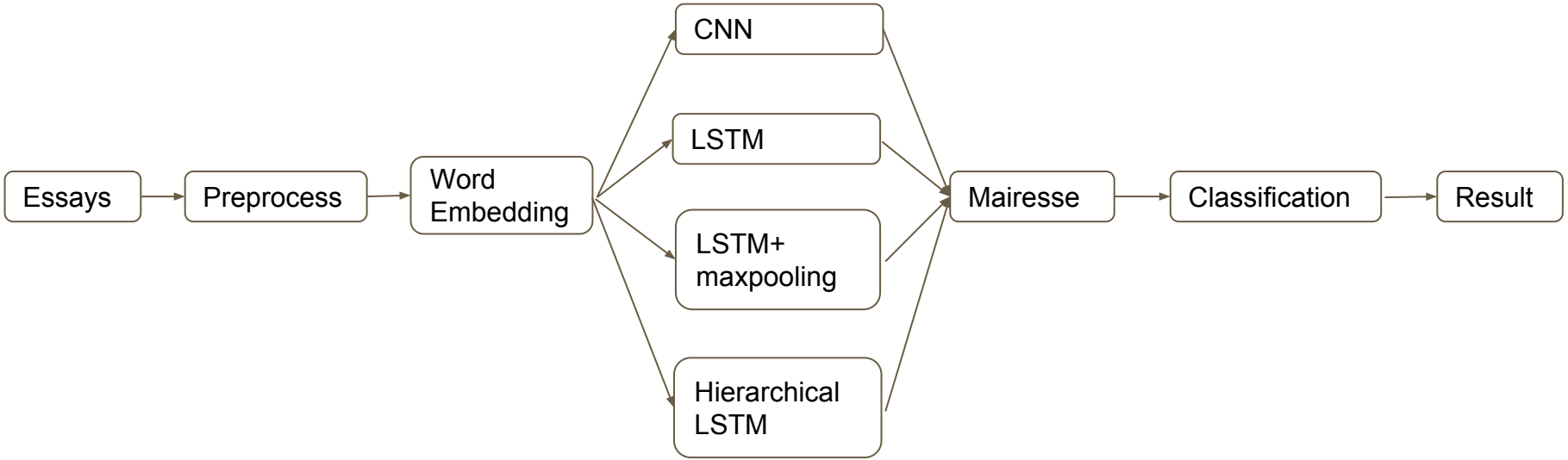


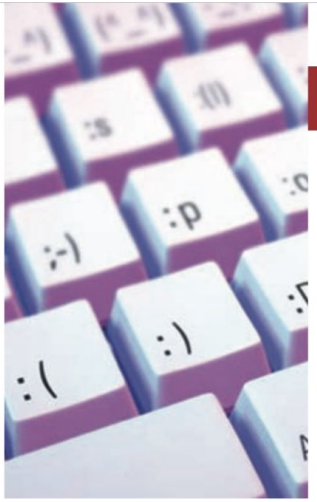
Mairesse and Classification

- Document-level features
  - Mairesse is created by François Mairesse which comprises the features of Linguistic Inquiry, Word Count, Medical Research Council, utterance-type, prosodic.
  - Examples: word count, average number of words per sentence, total number of pronouns, past tense verbs, present tense verbs, future tense verbs, letters, phonemes, syllables, questions, assertions in whole essays.
  - **84** features (from Mairesse) + **300** features (from every framework) = **384** features
  - Personality Recognizer v1.03

- Classification:Multilayer Perceptron and Logistic Regression

Flow Chart Review





AFFECTIVE COMPUTING AND SENTIMENT ANALYSIS

Editor: **Erik Cambria,** Nanyang Technological University, Singapore, cambria@ntu.edu.sg

Deep Learning-Based Document Modeling for Personality Detection from Text

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

Published in IEEE Intelligent Systems in 2017

Results in the Paper

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 in our Project

Document Vectorization	Classifier	EXT	NEU	AGR	CON	OPN	Multi-label (All Five)
Published state of the art per trait	N/A	56.45	58.33	56.03	56.73	60.68	N/A
Paper's best	Different	58.09	59.38	56.71	57.30	62.68	N/A
Only Mairesse	MLP	57.89	54.66	48.58	55.87	57.62	56.87
LSTM + Mairesse	Logistic Regression	**58.57**	56.95	**56.95**	54.93	58.57	53.55
LSTM + Mairesse	MLP	54.12	58.57	54.25	57.62	62.21	**57.14**
LSTM with Maxpooling + Mairesse	MLP	54.52	59.38	56.01	51.55	56.14	54.06
Hierarchical LSTM + Mairesse	MLP	**58.57**	**60.05**	56.28	56.82	61.54	55.38
Hierarchical CNN + Mairesse	MLP	54.12	55.74	53.85	**57.89**	**63.56**	55.41

MOTHERBOARD

The Data That Turned the World Upside Down

How Cambridge Analytica used your Facebook data to help the Donald Trump campaign in the 2016 election.

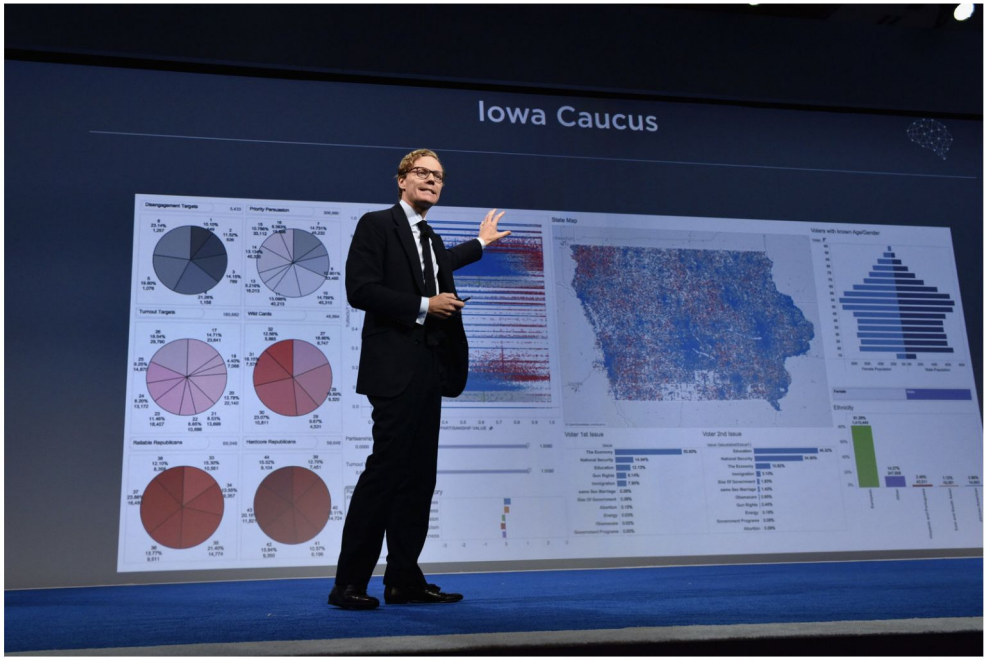
SHARE TWEET

Hannes Grassegger & Mikael Krogerus

Jan 28 2017, 6:15am

The Data That Turned the World Upside Down

MOTHERBOARD Jan 28 2017



Alexander Nix, the currently suspended CEO of Cambridge Analytica, speaks at a 2016 event in New York City. Nix and his firm are accused of misusing the personal data of 50 million people as part of their political consulting work, which included President Trump's 2016 campaign.

Bryan Bedder/Getty Images for Concordia Summit

Alexander Nix

Misuse 50 million Facebook users' personal data

President Trump’s 2016 campaign

68 Facebook "likes" by a user

- Skin color (with 95 percent accuracy)
- Sexual orientation (88 percent accuracy)
- Affiliation to the Democratic or Republican party (85 percent).
- Intelligence, religious affiliation, alcohol, cigarette,drug use, whether someone's parents were divorced…

What’s behind? **Big Five !**

What Did Cambridge Analytica Do During The 2016 Election?

NPR, March 20, 2018

The Data That Turned the World Upside Down

MOTHERBOARD Jan 28 2017

Future work

1.More Data

2.Training (50 epoch for more than 6 hours)

3.Prediction

4.Do **RIGHT** Things

Thank You !