

THE PERSONALITY OF THE SNAKE:

Personality Recognition using Convolutional Neural Networks

AUTHOR PROLING (AP)

Given a text, AP seeks to classify authors with respect to their demographics traits.

Demographics:

- Age.
- Gender.
- **Personality traits.**
- Political preferences.
- ...

Texts from:

- Essays.
- Blog posts.
- **Social media.**
- Source code.
- ...



METHODOLOGY

- Convolutional Neural Networks trained with word embeddings for Author Proling
- Deep learning.
- Explore a task-independent text representation.
- Global Vectors for Word Representation. (Pennington et al., 2014)

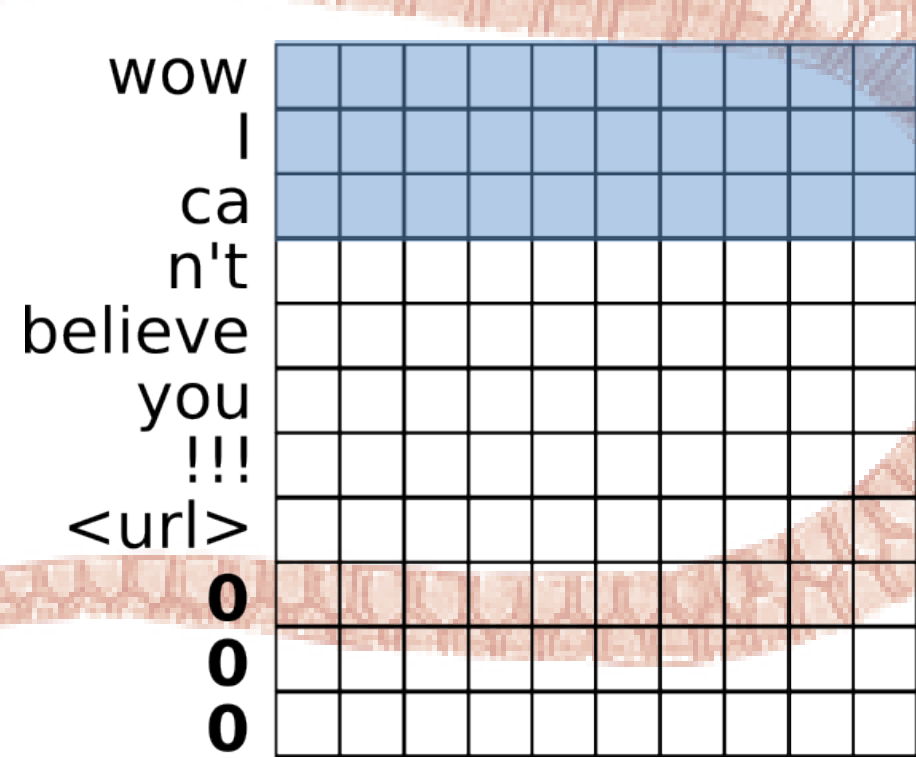


Figure 2: Input matrix padded to feed a CNN. Shaded rows represent the area covered by a filter of height 3.

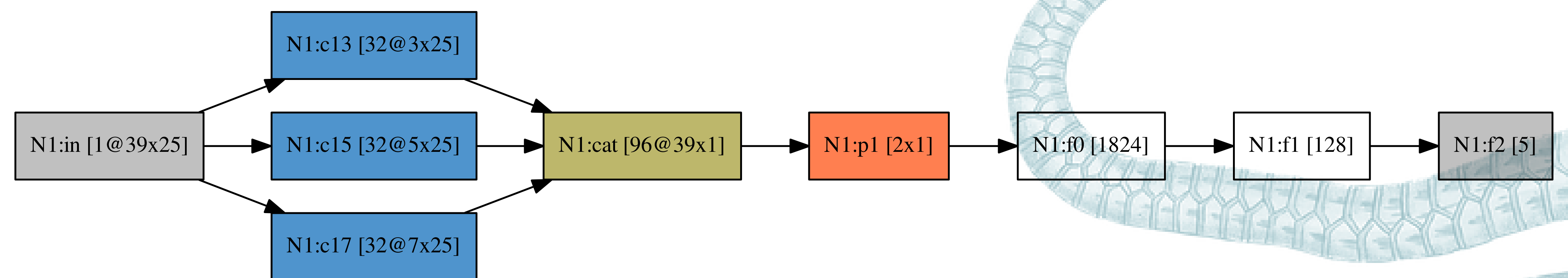


Figure 3: CNN architecture of our best performing system.

RESULTS

CNN Architecture	WE	RMSEt	RMSE
F=3, NF=32, HU=1, BN, MP	25	0.2253	0.1650
F=7, NF=32, HU=1, BN, MP	25	0.2090	0.1647
F=357, NF=32, HU=1, BN, MP	25	0.2213	0.1625
F=357, NF=32, HU=1, BN	50	0.2204	0.1633
F=3, NF=64, HU=2, DR	50	0.1995	0.1692

Table 1: Evaluation results of our best CNN models. RMSE: metric used in the competition. RMSEt: RMSE for each tweet. WE: Dimension of the word embedding. MP: If a max pooling layer was employed. F: Height of the filters used. NF: Number of filters used. HU: the number of hidden units. DR: If dropout was applied. BN: If batch normalization was applied.

Big Five or Factor Model which include five traits (Boyle et al., 2008):

- Agreeable** Agreeableness relates to a focus on maintaining positive social relations.
- Conscientious** Conscientiousness measures preference for an organized approach to life.
- Extroverted** Extroversion measures a tendency to seek stimulation in the external world.
- Open** Openness is related to imagination, creativity, curiosity, tolerance, political liberalism, and appreciation for culture.
- Stable** Emotional Stability measures the tendency to experience mood swings.

AUTHOR PROLING TASK AT PAN 2015

- PAN-AP-2015 dataset, used in the 3rd Author Proling shared task (Rangel et al., 2015).
- Twitter.
- English.

CONCLUSIONS AND FUTURE WORK

- We have presented a shallow CNN model for predicting Personality Traits from Social Media texts.
- We have used pre-trained word embeddings.
- Without handcrafted features or any external resource, we have trained a model able to achieve a good performance compared against state-of-the-art models for Author Proling.
- Learn word-embeddings in an unsupervised fashion from the task at hand.
- Explore deeper CNN architectures.
- Validate this model in datasets from different domains and different sources.

TOOLKITS



ABOUT ME

Maite Giménez

PhD student in Natural Language Processing,
Deep Learning and Social Media

hi@maigimenez.es

Github: @maigimenez