

Stacked Denoising Autoencoders in Author Profiling

Hubert Pomorski, Aleksandra Bańkowska
Paulina Winkowska, Marcin Kuta

AGH University of Science and Technology, Kraków

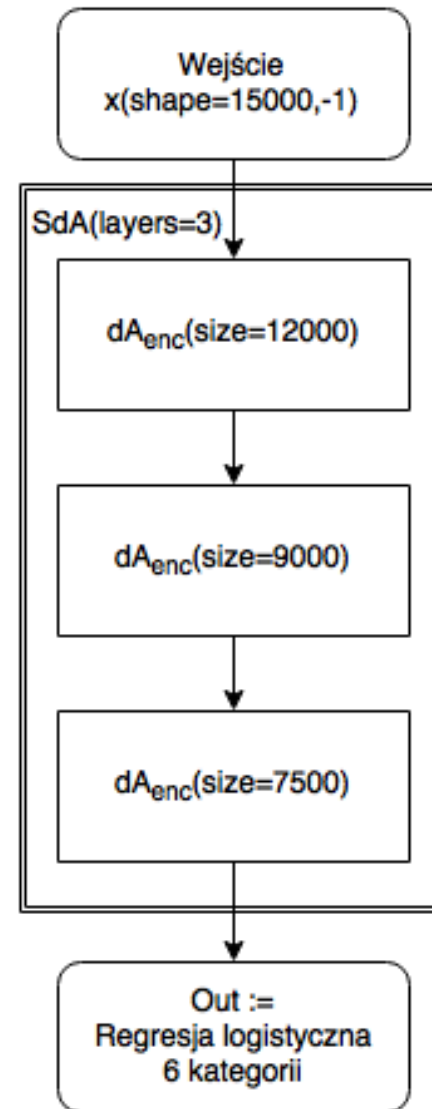
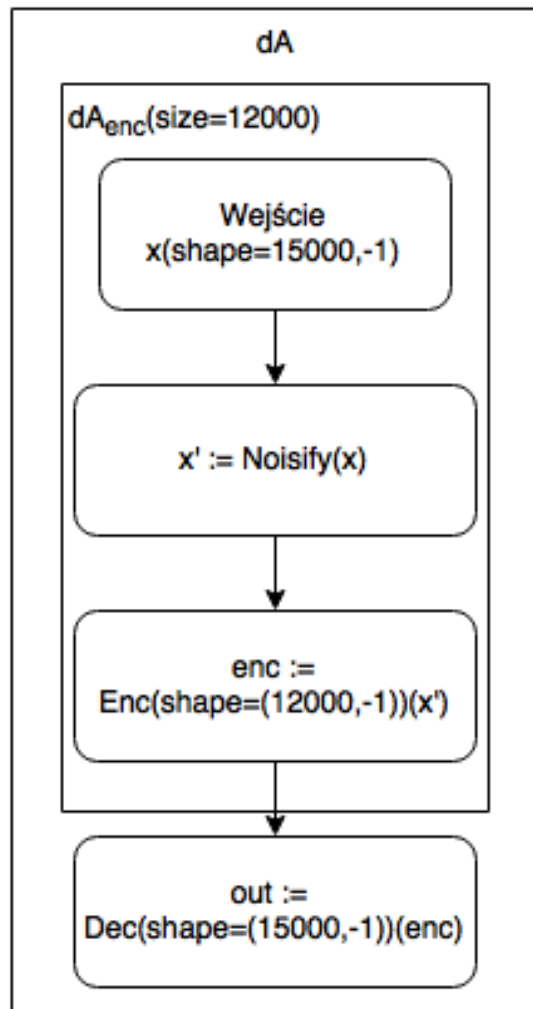
Department of Computer Science

Faculty of Computer Science, Electronics and Telecommunications

Author profiling

- Problem
 - Assign a profile to a text, whose authorship is unknown
- Author profile
 - Sex
 - Age
 - Personality
 - ...

Stacked denoising autoencoder



Feature Engineering

- number of emails, hyperlinks, long citations, dates, currencies, numbers, emotions (7 features)
- number of part of speech tags (45 features)
- number of emotion words (10 types of emotion)
- number of words describing subjectivity (8 levels),
- readability of the text (Flesch Reading Ease norm)
- 5000 most frequent 3-grams in text

Implementation

- Tasks parallelisation
 - Single-thread model (SVM classification)
 - Multi-threaded model (preprocessing)
 - GPGPU (stacked denoising autoencoder)