

Exploring convolutional neural networks and topic models for user profiling from drug reviews

Elena Tutubalina¹  · Sergey Nikolenko^{1,2}

Received: 26 January 2017 / Revised: 19 August 2017 / Accepted: 20 October 2017
© Springer Science+Business Media, LLC 2017

Abstract Pharmacovigilance, and generally applications of natural language processing models to healthcare, have attracted growing attention over the recent years. In particular, drug reactions can be extracted from user reviews posted on the Web, and automated processing of this information represents a novel and exciting approach to personalized medicine and wide-scale drug tests. In medical applications, demographic information regarding the authors of these reviews such as age and gender is of primary importance; however, existing studies usually either assume that this information is available or overlook the issue entirely. In this work, we propose and compare several approaches to automated mining of demographic information from user-generated texts. We compare modern natural language processing techniques, including extensions of topic models and convolutional neural networks (CNN). We apply single-task and multi-task learning approaches to this problem. Based on a real-world dataset mined from a health-related web site, we conclude that while CNNs perform best in terms of predicting demographic information by jointly learning different user attributes, topic models provide additional information and reflect gender-specific and age-specific symptom profiles that may be of interest for a researcher.

Keywords Text mining · Natural language processing · Topic modeling · Deep learning · Convolutional neural networks · Multi-task learning · Single-task learning · User reviews · Demographic prediction · Demographic attributes · Social media · Mental health

✉ Elena Tutubalina
elvtutubalina@kpfu.ru
Sergey Nikolenko
sergey@logic.pdmi.ras.ru

¹ Kazan (Volga Region) Federal University, Kazan, Russia

² Steklov Institute of Mathematics, St. Petersburg, Russia