Automated determination of the type of genre and stylistic coloring of Russian texts

Vladimir Barakhnin¹,²*, Olga Kozhemyakina² and Ilya Pastushkov¹

¹ Institute of Computational Technologies of SB RAS, Lavrentiev av., 6, 630090, Novosibirsk, Russia
² Novosibirsk State University, Pirogova st., 1, 630090, Novosibirsk, Russia
Abstract

In this paper we propose the algorithm of automated definition of the genre type and semantic characteristics of poetic texts in Russian. We formulated the approaches to the construction of a joint ("two-dimensional") classifier of genre types and stylistic colouring of poetic texts, based on the definition of interdependence of the type of genre and stylistic colouring of the text. On the basis of these approaches the principles of formation of the training samples for the algorithms for the definition of styles and genre types were analyzed. The computational experiments with a corpus of texts of the Lyceum lyrics of A.S.Pushkin were implemented, which showed good results in determining the stylistic colouring of Russian poetic texts, significantly facilitating the work of the expert in determining their styles and genres by providing appropriate recommendations.

© The Authors, published by EDP Sciences, 2017

This is an Open Access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Automated determination of the type of genre and stylistic coloring of Russian texts. ITM Web of Conferences, 10, 02001. doi: 10.1051/itmconf/20171002001. Python Web Development with Django. Machine learning algorithms are mainly focused on generalization (e.g. predicting unseen data in a classification problem). However a new perspective on their efficiency would be the capability of explaining the predictions. By the process of binary encoding the features, it is possible to propitiate an intuitive interpretation, but it could lead to loss of information. This work aims to explore [Show full abstract] this trade-off, in which a discretization method based on the dispersion values of numerical features and its effects on the accuracy of a machine learning method are evaluated. Speaking of the notion of style and stylistic colouring we cannot avoid the problem of the norm and neutrality and stylistic colouring in contrast to it. 1.5. Stylistic neutrality and stylistic colouring. Most scholars abroad and in this country giving definitions of style come to the conclusion that style may be defined as deviation from the lingual norm. Only ABC-books and texts for foreigners would be considered «normal». Everything that has style, everything that demonstrates peculiarities of whatever kind would be considered abnormal, including works by Dickens, Twain, O'Henry, Galsworthy and so on (47, pp. 21-22). The stylistic colouring is nothing but the knowledge where, in what particular type of communication, the unit in question is current. Notes on Style • Syntactical stylistic devices: 1. Repetition; 2. Syntactical parallelism or a parallel structure – repetition of the same syntactical structure; 3. Antithesis or contrast – opposition of two ideas or features thus heightening the effect of the utterance; • Lexical stylistic devices: 1. An epithet – an attributive word of phrase expressing some quality of a person, thing of phenomenon; “the cynical confidence” 2. A.