

MSc Barbara Probierz

Institute of Computer Science
Faculty of Computer Science and Materials Science
University of Silesia in Katowice, Poland

Title of the dissertation:

„Automatic categorization of e-mail messages to folders with the use of Social Networks and Ant Colony Optimization algorithms”

Abstract

The dissertation deals with methods that allow the use of Ant Colony Optimization algorithms and Social Networks to solve the problem of automatic categorization of e-mails to folders. The main aim of this work is to create an algorithm that would allow one to improve the classification of emails into folders along with the ability to suggest the creation of new folders.

During the implementation of the objectives, the Enron E-mail data set was thoroughly analyzed, cleaned up, adapted to the analyzed problem and transformed to the appropriate structure. Next, a social network was created based on the contacts between the senders and recipients of e-mail messages, as well as on the basis of conducted analysis and observation of the social network, groups of users with a similar social structure were identified. Mailboxes of users belonging to selected groups have been transformed into decision tables. Based on them, a classifier was built using Ant Colony Optimization algorithms, thanks to which it is possible to search a larger space of solutions and find alternative methods of solutions.

For the study used classic classifiers, ensembles of classifiers, Ant Colony Decision Tree and Ant Colony Decision Forest algorithms. Analysis of the results obtained contributed to the design of an original algorithm for automatically categorizing e-mails to folders. The proposed algorithm was also used to develop a mechanism for suggesting the creation of new folders, based on the structure of the folders of other users in a given group.

All the aims in the dissertation were fully achieved. The theoretical analysis and experiments carried out completely confirmed the thesis.