

A new approach to data classification and its applications to adverse drug reaction problems

Musa Mammadov
m.mammadov@ballarat.edu.au
School of ITMS
University of Ballarat
University Drive, Mt. Helen, PO Box 663
Ballarat, Victoria 3353
AUSTRALIA

Abstract

We present a new approach to text classification problems developed at the University of Ballarat. This approach is based on global optimization and targets to study *Shorter Featured Multi Label* (SFML) datasets. The significance of research on SFML datasets, from different areas, and their characteristics are addressed. In this talk, we mainly concentrate on applications to some important Adverse Drug Reaction (ADR) problems. The talk consists of three parts.

In the first part, we briefly describe a recently developed method for global optimization problems. The first such algorithm (AGOP) has already been applied to many difficult practical problems from different areas, where the objective function is non-smooth (the robust stabilization problem and others), or even is discontinuous (location problems in Telecommunication and inter-market influences in Finance). In the applications to text classification, we have smooth but large scale global optimization problems.

In the second part we describe our approach, developed for ADR problems, where the problem of prediction of reactions (classes) for a given set of drugs (words) is formulated as a text classification problem. The key point in this approach is the drug-reaction relationships described by corresponding weights. We formulate a global optimization problem to find these weights. It is a large scale optimization problem having a large number of local optimal solutions. We present a method that allows us to find good solutions to this problem. This method consists of 3 steps. The first step uses a heuristic method to

Data Mining and Mathematical Programming
October 10-13, 2006

find a good initial point. The next two steps involve global optimization problems formulated using this initial point. We apply AGOP to solve these global optimization problems. This approach, developed for ADR problems, can be used for classification on SFML datasets.

In the third part, we consider some applications of the proposed approach. It can be used as a starting point for solving many ADR problems. We consider here only two of them:

- 1) identification of suspected drugs; that is, given a set of drugs and a set of reactions occurred, to determine the drugs that are the most likely cause these reactions;
- 2) drug-drug interactions.

The approach is applied to different classes of reactions from the Australian Adverse Drug Reaction Advisory Committee (ADRAC) database.