Chapter 33
Particle Swarm Optimization Algorithm as a Tool for Profiling from Predictive Data Mining Models

Goran Klepac
Raiffeisenbank Austria Zagreb, Croatia

ABSTRACT

This chapter introduces the methodology of particle swarm optimization algorithm usage as a tool for finding customer profiles based on a previously developed predictive model that predicts events like selection of some products or services with some probabilities. Particle swarm optimization algorithm is used as a tool that finds optimal values of input variables within developed predictive models as referent values for maximization value of probability that customers select/buy a product or service. Recognized results are used as a base for finding similar profiles between customers. The presented methodology has practical value for decision support in business, where information about customer profiles are valuable information for campaign planning and customer portfolio management.

INTRODUCTION

This chapter will present novel methodology of particle swarm optimization algorithm usage as a tool for finding customer profiles based on previously developed predictive models which predicts events like selection of some products or services with some probabilities.

It means that particle swarm optimization algorithm will be used as a tool which should find optimal values of input variables within developed predictive models as referent values for maximization value of probability that customer will select/ buy some product or service.

Basic idea is holistic process which includes development of multinomial predictive models (predictive models with more than two states which represents probability for selecting/buying product /services), and usage of this model developed on historical data sample for finding typical customer/buyer based on evaluated values of input variables by using particle swarm optimization algorithm.

DOI: 10.4018/978-1-5225-0788-8.ch033

This problem is not so expressed in case of predictive models with binomial outputs, which represents probability of buying some product or service.

Even in situation when those predictive models are not solved with linear models (like neural networks, SVM, Bayesian networks), and linear dependencies are not so obvious, regarding binominal output, profiles could easily been recognized, thanks to attribute relevance analysis which is relatively simple and unambiguous.

Task about profiling customers based on selecting/buying product /services in situation when predictive model are used for probability of N possible states calculation, became hard task which demands lot of manual work with doubtful results.

Contrary to situation with binomial states of output variable in predictive model, attribute relevance analysis for multinomial output, regarding overlapping in impact zones, and combinatory expansion could not provide clear information for setting input variable values for determination buying preferences represented as values for buying or not, one of the many product/services.

For retailers, web shops, and other industries it is challenging task, which can be solved by using particle swarm optimization algorithm in combination with existing predictive model.

In that case, for each observed product/service, aim variable within predictive model should reach criteria of maximum probability of selecting/buying product /services by using particle swarm optimization algorithm. Final result will give optimal values of input variables for each product /services which could be behavior characteristics, or socio demographic characteristics and from which model are developed based on attribute relevance analysis.

Result of those analytical approach are characteristic points in multidimensional spaces which can be used for case base reasoning, or clustering by distance measure usage as well as for profiling.

Other benefits, which proposed method provides is answers on questions like:

- Does typical buyer of product/service “A” is most similar to buyer of product/service “F” or most similar to buyer of product/service “G”?
- Buyers of which products are similar to each other?
- If buyer “A” bought product “X”, regarding its similarity, which next product should be offered to him?

This methodology could be applied on predictive models based on multinomial logistic regression, Bayesian networks, neural networks, support vector machines, and other types of predictive models.

Main advantage of proposed solution is automatic determination of profiles based on selecting/buying product /services in situation where single model makes prediction for numerous product /services.

Beside advantages of proposed methodology, consumption about similar behavior in neighborhood of recognized points within input variables demands additional checks.

Consumption about similar behavior in neighborhood of recognized points could be, or not be the case. It depends on empirical customer behavior, and it demands additional tests during implementation of proposed methodology. This test can also be interesting for determination of optimal distances in recognized points within input variables.

Practical value of presented method is also visible in situation after predictive model calibration and repeating analytical procedure based on particle swarm optimization, for new characteristic point values calculation in multidimensional spaces.