On Rough Set Based Non Metric Model

Yasunori Endo[1,4], Ayako Heki[2], and Yukihiro Hamasuna[3]

[1] Faculty of Eng., Info. and Sys., University of Tsukuba
Tennodai 1-1-1, Tsukuba, 305-8573 Ibaraki, Japan
endo@risk.tsukuba.ac.jp
[2] Graduate School of Sys. and Info. Eng., University of Tsukuba
Tennodai 1-1-1, Tsukuba, 305-8573 Ibaraki, Japan
s1120638@u.tsukuba.ac.jp
[3] Department of Informatics, Kinki University
3-4-1, Kowakae, Higashiosaka, 577-8502 Osaka, Japan
yhama@info.kindai.ac.jp
[4] International Institute for Applied Systems Analysis (IIASA)
Schlossplatz 1, A-2361 Laxenburg, Austria

Abstract. Non metric model is a kind of clustering method in which belongingness or the membership grade of each object to each cluster is calculated directly from dissimilarities between objects and cluster centers are not used.

By the way, the concept of rough set is recently focused. Conventional clustering algorithms classify a set of objects into some clusters with clear boundaries, that is, one object must belong to one cluster. However, many objects belong to more than one cluster in real world, since the boundaries of clusters overlap with each other. Fuzzy set representation of clusters makes it possible for each object to belong to more than one cluster. On the other hand, the fuzzy degree sometimes may be too descriptive for interpreting clustering results. Rough set representation could handle such cases. Clustering based on rough set representation could provide a solution that is less restrictive than conventional clustering and less descriptive than fuzzy clustering.

This paper shows two type of Rough set based Non Metric model (RNM). One algorithm is Rough set based Hard Non Metric model (RHNM) and the other is Rough set based Fuzzy Non Metric model (RFNM). In the both algorithms, clusters are represented by rough sets and each cluster consists of lower and upper approximation. Second, the proposed methods are kernelized by introducing kernel functions which are a powerful tool to analize clusters with nonlinear boundaries.

1 Introduction

Computer system data has become large-scale and complicated in recent years due to progress in hardware technology, and the importance of data analysis techniques has been increasing accordingly. Clustering, which means a data classification method without any external criterion, has attracted many researchers as a significant data analysis technique.

V. Torra et al. (Eds.): MDAI 2012, LNAI 7647, pp. 394–407, 2012.

Bezdek et al. proposed Fuzzy c-Means (FCM) [1, 2], which can be regarded as the fuzzification of Hard c-Means (HCM). In FCM, cluster centers are introduced to obtain belongingness or the membership grade of each object to each cluster. Roubens proposed a clustering algorithm without cluster centers [3] in which membership grades are calculated with only dissimilarities between objects. The algorithm is then called a Fuzzy Non Metric model (FNM) which means fuzzified Hard Non Metric model (HNM). The FNM has a property enabling membership grades to be calculated directly from dissimilarities, and thereby object space need not necessarily be Euclidean space. After that, Bezdek et al. discussed relations between FCM and FNM [4, 5]. One of us proposed Entropy based FNM (EFNM) by the different fuzzification of HNM from Roubens, and discussed some applications [6]. In these algorithms, fuzzy set concept plays very important role.

On the other hand, it is pointed out that the fuzzy set representation sometimes may be too descriptive for interpreting clustering results [7]. In such cases, rough set representation becomes a useful and powerful tool [8, 9]. The basic concept of the rough representation is based on two definitions of lower and upper approximations of a set. The lower approximation means that "an object surely belongs to the set" and the upper one means that "an object possibly belongs to the set". Clustering based on rough set representation could provide a solution that is less restrictive than conventional clustering and less descriptive than fuzzy clustering [10, 7], and therefore the rough set based clustering has attracted increasing interest of researchers [11–16, 7].

In this paper, we will mainly develop theoretical discussion for a new non metric model based on rough set representation. We will first construct two type of Rough set based Non Metric model (RNM). One algorithm is Rough set based Hard Non Metric model (RHNM) and the other is Rough set based Fuzzy Non Metric model (RFNM). In RNM, clusters are represented by rough sets and each cluster consists of lower and upper approximation.

RFNM is constructed by fuzzifying RHNM with the entropy regularized term. The purpose of fuzzification of RHNM is to overcome a problem of RHNM that if an object does not belong to any lower approximation it belongs to only two upper approximation, not three or more. To fuzzify RHNM, we introduce an entropy regularized term which was considered by Miyamoto et al. as another way to fuzzify HCM [17, 18] from by Bezdek. Bezdek introduced a fuzzification parameter on membership grades in the objective function of HCM, while Miyamoto et al. introduced an entropy regularization term into the function, instead of the fuzzification parameter. In comparison with FCM, EFCM has, among other properties, one in which the classification function converges to 0 or 1 as the dissimilarity between a cluster center and an object goes to infinity.

RHNM is more simple than RFNM so that it is expected that the calculation cost of RHNM is less than RFNM. On the other hand, RFNM is more flexible than RFNM in the meaning that each object can belong to two or more upper approximation.

Second, we will kernelize the proposed methods by introducing kernel functions. Vapnik proposed the Support Vector Machine (SVM) [19, 20]. The SVM