ФОРМИРОВАНИЕ КЛАСТЕРНОЙ СТРУКТУРЫ ЗОНЫ РИСКА БАНКРОТСТВА ПРЕДПРИЯТИЯ

Герман Ю.О.[1], Герман О.В.[2], Кузнецов М.В.[2]
[1]Факультет компьютерных сетей и систем, Белорусский государственный университет информатики и радиоэлектроники
[2]Факультет информационных технологий и управления, Белорусский государственный университет информатики и радиоэектроники
Минск, Республика Беларусь
E-mail: ovgerman@tut.by

Рассматривается задача формирования кластеров на множестве состояний зоны риска банкротства («серой» области). Предлагается новый оригинальный метод построения кластеров на основе метода решения задачи нечеткого математическогр программирования. Представленный подход позволяет устранить некоторые недостатки известных методов - Fuzzy KMeans и Possibility CMeans.

Введение

Производственная система (финансовая организация) рассматривается как сложная система, состояния которой описываются многомерными векторами. Значения векторов меняются во времени в зависимости от многих факторов по некоторой траектории. Значения векторов формируют определенные кластеры, из которых нас интересуют те, которые попадают в «серую» область. Эту область состояний системы нельзя однозначно идентифицировать как устойчиво позитивную (штатную) или зону банкротства. Мы выделили такие кластеры в серой области: PS (кластер отклоняющихся от нормы состояний, которые идентифицируются как позитивные, более близкие к N) и NG (кластер отклоняющихся от нормы состояний, которые идентифицируются как негативные, более близкие к F). Соответственно различными должны быть управляющие воздействия в этих кластерах. Предлагается новый способ формирования кластерной структуры серой зоны, базирующийся на идеях нечеткого математического программирования.

I. Краткий анализ проблематики

Следует определиться с формальным критерием «серой» области. Сначала сформулируем идею концептуально. Пусть дано множество точек из двух классов. Рассмотрим произвольное разбиение этого множества точек на непересекающиеся подмножества A (точек первого класса), B (точек второго класса), C (нет требований по классам), так что их объединение дает исходное множество точек. Тогда множество C будет серой областью, если выполнены следующие условия:

1. A и B разделимы линейной дискриминаторной функцией;

2. Множества $\{x\} \cup A$ и B не являются линейно разделимыми, равно как и A и $\{y\} \cup B$ не являются линейно разделимыми для любых $x \in A$, $y \in B$, $x, y \in C$.

Допущения 1,2 далее используем как исходные для излагаемого материала.

Как было отмечено выше, состояния (объекты) в «серой» области принадлежат разным кластерам (классам). Известны два базовых подхода к построению нечетких кластеров [1]: алгоритм Fuzzy C-Means (нечетких C-средних) - FCM и Possibilistic C-Means (возможных C-средних)–PCM.

Алгоритм FCM выполняется как итерационный процесс. На каждой итерации объекты перераспределяются по кластерам и для них выполняется пересчет нечетких мер принадлежности и значений центроидов. Итерации повторяются, пока центроиды не стабилизируются. Недостатки FCM усматриваются в следующем [2]:

1. алгоритм плохо оценивает крайние точки или шумы, давая им завышенные значения степеней принадлежности к кластерам;

2. степени принадлежности к кластеру зависят от координат центроидов других кластеров, что в ряде случаев искажает логику определения самих нечетких мер.

Недостатки PCM усматриваются в следующем [2]:

1. высокая чувствительность к начальной инициализации (начальному разбиению на кластеры);

2. кластеры могут «накладываться» друг на друга. Этих недостатков лишен предлагаемый метод.

II. Математическое содержание метода

Нас будет интересовать подход [3] (Zimmermann), который мы развиваем в данной работе. Этот подход изначально ориентирован на решение задач нечеткого математического программирования и используется в материалах данной работы для отыскания нечетких кластеров. Под кластерной структурой будем понимать совокупность кластеров и их состав. Для ясности в качестве исходных данных используется некоторая заданная a priori таблица, строки которой представляют состояния производственной системы в моменты измерений (например, указаны значения критериев пятифакторной модели Альтмана из серой области и результирующая оценка риска банкротства z).Предполагается, что кластерная структура не известна и должна быть найдена. Базируясь на работе [4], составим следующую систему для определения двух кластеров в серой области. Целевая функция $G = \sum_i \lambda_i^2 \to min$.

Ограничения однотипны и имеют вид линейных алгебраических неравенств:

$\sum_i a_i x_{i,k} = d - \lambda_k$,

d – вещественное число, например $d = 0$, x_k – вектор состояния системы на шаге k.

Решение системы позволяет найти два кластера. Векторы состояний x_k попадают в кластера сообразно тому, с каким значением λ_k эти векторы (состояния) ассоциированы в решении: $\lambda_k \geq d$ - первый кластер, $\lambda_k < d$ - второй кластер.

Обоснование использованной модели состоит в следующем. Принято, что между двумя кластерами можно провести гиперплоскость, если

$\sum_i a_i x_{i,k} \geq (<)\ d$,

с неравенством типа $\geq$ для одного кластера и жестким неравенством $<$ для другого (d – вещественное число, например 0). В исходной системе ограничений гиперплоскость, разделяющая два кластера, строится таким образом, чтобы векторы (точки в многомерном пространстве) каждого кластера группировались как можно плотнее друг к другу в силу целевого функционала G. Функционал такого вида «подтягивает» удаленные точки кластера к «основной» массе, которая «группируется» рядом с разделяющей гиперплоскостью. Если точки лежат на гиперплоскости, то вопрос об отнесении их к какому-либо кластеру требует дополнительного исследования (либо решается произвольным образом – случайно). Из этого наблюдения можно считать, что чем дальше точка лежит от разделяющей гиперплоскости, тем меньше степень ее принадлежности к кластеру. Базируясь на этом замечании, можно предложить метод вычисления нечетких мер принадлежности. Константа d выбрана произвольно (как можно доказать, выбор константы d не влияет на результат определения кластерной структуры). Проблема связана с тем, что описанная концепция ведет к большому числу неравенств с малыми отклоняющимися от 0 значениями λ_i. Таким образом, схема должна быть адаптирована так, чтобы получить как можно меньшее число неравенств, для которых значение λ_i не нулевое.

III. Заключение

Таким образом, нами установлен способ получения кластеров в «серой» области на базе общего подхода к решению задач нечеткого математического программирования [1]. Ясно, что произвольное множество точек можно разделить на два кластера (класса) не единственным способом. Предложенный способ, минимизирует суммарное расстояние между точками каждого кластера за счет использования целевого функционала G. Этот функционал «подтягивает» точки кластера к гиперповерхности, отделяющей его от другого кластера и позволяет с этих соображений обосновать предложенный способ вычисления нечеткой меры принадлежности к кластеру. Достоинством предложенного подхода является то, что он не является итерационным и позволяет сразу получать необходимые оценки. Кроме того, он не связан с недостатками, указанными для методов Fuzzy KMeans и Possibilistic C-Means. Данный способ можно использовать для получения не только двух, но и трех и более кластеров, однако это требует специальной техники.Например, для разбиения на три кластера описанную технику модернизируем так: разбиваем исходное множество на два кластера: A и B. Затем кластер B разбиваем на два кластера B_1 и B_2. Разбиваем объединения $A \cup B_1$ и $A \cup B_2$ каждое на два множества, то есть имеем подмножества X, Y, U, W. Находим два из этих подмножеств, объединение Z которых в наибольшей степени представлено в A и полагаем, что первый кластер определен и им является Z. Удаляем его из исходного множества и определяем в оставшемся множестве векторов два других кластера.

IV. Список литературы

1. Grover, N. A study of various fuzzy clustering algorithms. / N. Grover// International Journal of Engineering Volume No.3, Issue No.3, pp : 177–181.
2. Binsy, T. Comparative Analysis Of Fuzzy Clustering Algorithms In Data Mining / T. Binsy,Madhu Nashipudimath. // International Journal of Advanced Engineering Research. Volume No.3, Issue No.3, pp.177–181.
3. Fuller, R. Fuzzy reasoning for solving fuzzy mathematical programming problems /R. Fuller,H-J. Zimmerman. //Fuzzy sets and systems, 60, 1993, p.p.121-133.
4. Герман, О. В. Подход к выбору управления в системе кластеров / О. В. Герман, Ю. О. Герман, М. В. Кузнецов // Труды БГТУ. Сер. 3, Физико-математические науки и информатика. - Минск: БГТУ, 2020. - № 1 (230). - С. 63-68