**Группировка обучающих образов и ранжирование групп.**

**Дано:**

Обучающие образы 1 и 2:

Обучающие образы 1 находятся в матрице:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| s0 | s1 | s2 | s3 | s4 | s5 | s6 | s7 | s8 |
| 5 | 5,5 | 5,3 | 4,9 | 5,2 | 5,1 | 5,4 | 5,3 | 5 |
| 6 | 6 | 6,5 | 5,5 | 6,3 | 6 | 6,2 | 6,2 | 6,1 |
| 1 | 1,2 | 0,9 | 1,4 | 1,3 | 1,1 | 1 | 1,2 | 1,1 |

Обучающие образы 2 находятся в матрице:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| s0 | s1 | s2 | s3 | s4 | s5 | s6 | s7 | s8 | s9 | s10 |
| 1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| -1 | -1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 1 | -1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | -1 | 1 | -1 | 1 | 1 | 1 |
| 1 | -1 | -1 | -1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 |
| -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | -1 | -1 | -1 |

**Задание:**

Разработать алгоритм группировки обучающих образов и ранжирование групп.

**Решение:**

Был разработан и реализован на встроенном языке Matlab алгоритм, который осуществляет группировку обучающих образов и ранжирование групп.

Описание метода.

Суть метода заключается в том, чтобы на каждом шаге из оставшейся выборки элементов выделять наибольшую группу размером не менее 20% (это можно задавать и менять) от оставшейся выборки таких элементов, которые имеют модуль относительной погрешности с соответствующим эталонным элементом не более текущей допустимой.

Эталонным элементом является поочередно каждый элемент обучающей выборки. Каждый элемент, включая тот, который стал эталоном, сравнивается с эталонным следующим образом:

* Вычисляется относительная погрешность (её модуль) значений контролируемых признаков к соответствующим признакам эталонного элемента.
* Далее берется среднее значение относительных погрешностей всех контролируемых признаков по текущему элементу, который сравнивается с эталоном.
* Если среднее значение относительных погрешностей менее или равно значению текущей допустимой относительной погрешности, элемент включается в текущую группу по текущему эталону.

Изначально текущая, допустимая относительная погрешность задается равной 0. Если в оставшейся выборке не находится групп элементов размером 20%, которые соответствуют вышесказанным условиям, следует увеличивать значение текущей допустимой погрешности до тех пор, пока не появится такая группа или группы.

Наибольшая группа изымается из оставшейся выборки записывается в конец результирующего массива с образами.

Процесс с оставшейся выборкой, множество образов которой не пересекается с множеством образов, которое составляет результирующий массив, продолжается до тех пор, пока все элементы из оставшейся выборки не попадут в результирующий массив.

Пример1.

Обучающая выборка находится в матрице:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| s0 | s1 | s2 | s3 | s4 | s5 | s6 | s7 | s8 | s9 | s10 |
| 1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| -1 | -1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 1 | -1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | -1 | 1 | -1 | 1 | 1 | 1 |
| 1 | -1 | -1 | -1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 |
| -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | -1 | -1 | -1 |

Возьмем за эталон элемент образ s0

Рассчитаем относительные погрешности значений контролируемых элементов образа s1 и эталонного образа s0.

|1+1/1| = 2

|-1+1/1| = 0

|1-1/1| = 0

|1-1/1| = 0

|1+1/1| = 2

|-1+1/1| = 0

Среднее значение относительной погрешности = 4/6 =0,666

Числитель – сумма значений погрешностей контролируемых признаков.

Знаменатель – количество контролируемых признаков.

Составляется таблица взаимных средних относительных погрешностей для всех элементов обучающей выборки.

Результат – симметричная матрица.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | s0 | s1 | s2 | s3 | s4 | s5 | s6 | s7 | s8 | s9 | s10 |
| s0 | 0,0000 | 0,6667 | 1,3333 | 0,3333 | 0,6667 | 1,0000 | 0,0000 | 0,3333 | 0,0000 | 0,0000 | 0,0000 |
| s1 | 0,6667 | 0,0000 | 1,3333 | 0,3333 | 1,3333 | 1,0000 | 0,6667 | 1,0000 | 0,6667 | 0,6667 | 0,6667 |
| s2 | 1,3333 | 1,3333 | 0,0000 | 1,0000 | 0,6667 | 1,0000 | 1,3333 | 1,6667 | 1,3333 | 1,3333 | 1,3333 |
| s3 | 0,3333 | 0,3333 | 1,0000 | 0,0000 | 1,0000 | 0,6667 | 0,3333 | 0,6667 | 0,3333 | 0,3333 | 0,3333 |
| s4 | 0,6667 | 1,3333 | 0,6667 | 1,0000 | 0,0000 | 1,0000 | 0,6667 | 1,0000 | 0,6667 | 0,6667 | 0,6667 |
| s5 | 1,0000 | 1,0000 | 1,0000 | 0,6667 | 1,0000 | 0,0000 | 1,0000 | 0,6667 | 1,0000 | 1,0000 | 1,0000 |
| s6 | 0,0000 | 0,6667 | 1,3333 | 0,3333 | 0,6667 | 1,0000 | 0,0000 | 0,3333 | 0,0000 | 0,0000 | 0,0000 |
| s7 | 0,3333 | 1,0000 | 1,6667 | 0,6667 | 1,0000 | 0,6667 | 0,3333 | 0,0000 | 0,3333 | 0,3333 | 0,3333 |
| s8 | 0,0000 | 0,6667 | 1,3333 | 0,3333 | 0,6667 | 1,0000 | 0,0000 | 0,3333 | 0,0000 | 0,0000 | 0,0000 |
| s9 | 0,0000 | 0,6667 | 1,3333 | 0,3333 | 0,6667 | 1,0000 | 0,0000 | 0,3333 | 0,0000 | 0,0000 | 0,0000 |
| s10 | 0,0000 | 0,6667 | 1,3333 | 0,3333 | 0,6667 | 1,0000 | 0,0000 | 0,3333 | 0,0000 | 0,0000 | 0,0000 |

Проводим группировку и получаем массив индексов. Этот массив говорит о том, в каком порядке должны быть использованы образы обучающей выборки при расчете изображения технического состояния i исследуемой системы.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 6 | 8 | 9 | 10 | 1 | 3 | 5 | 7 | 2 | 4 |

Результаты без группировки:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | E10 |
| 1 | -1 | 0 | 0,33 | 0,5 | 0,6 | 0,67 | 0,71 | 0,75 | 0,78 | 0,8 |
| -1 | -1 | 0 | -0,3 | -0,5 | -0,6 | -0,7 | -0,7 | -0,8 | -0,8 | -0,8 |
| 1 | 1 | 0 | 0,33 | 0 | 0,2 | 0,33 | 0,43 | 0,5 | 0,56 | 0,6 |
| 1 | 1 | 1 | 1 | 1 | 0,6 | 0,67 | 0,43 | 0,5 | 0,56 | 0,6 |
| 1 | -1 | -1 | -1 | -0,5 | -0,6 | -0,3 | -0,1 | 0 | 0,11 | 0,2 |
| -1 | -1 | 0 | -0,3 | 0 | 0,2 | 0 | -0,1 | -0,3 | -0,3 | -0,4 |

Резултаты после группировки

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | E10 |
| 1 | 1 | 1 | 1 | 1 | 0,6 | 0,67 | 0,71 | 0,75 | 0,78 | 0,8 |
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -0,8 | -0,8 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0,78 | 0,6 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0,71 | 0,5 | 0,56 | 0,6 |
| 1 | 1 | 1 | 1 | 1 | 0,6 | 0,33 | 0,14 | 0,25 | 0,11 | 0,2 |
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | -0,7 | -0,8 | -0,6 | -0,4 |

Пример 2.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 |
| 5 | 5,5 | 5,3 | 4,9 | 5,2 | 5,1 | 5,4 | 5,3 | 5 |
| 6 | 6 | 6,5 | 5,5 | 6,3 | 6 | 6,2 | 6,2 | 6,1 |
| 1 | 1,2 | 0,9 | 1,4 | 1,3 | 1,1 | 1 | 1,2 | 1,1 |

Составляется таблица взаимных средних относительных погрешностей для всех элементов обучающей выборки.

Результат – симметричная матрица.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 | Y8 | Y9 |
| Y1 | 0,0000 | 0,1000 | 0,0144 | 0,0989 | 0,1300 | 0,0400 | 0,0378 | 0,0978 | 0,0389 |
| Y2 | 0,0859 | 0,0000 | 0,0677 | 0,0086 | 0,0263 | 0,0520 | 0,0505 | 0,0010 | 0,0525 |
| Y3 | 0,0075 | 0,0980 | 0,0000 | 0,1087 | 0,1316 | 0,0359 | 0,0279 | 0,0957 | 0,0347 |
| Y4 | 0,0581 | 0,0235 | 0,0312 | 0,0000 | 0,0451 | 0,0275 | 0,0188 | 0,0220 | 0,0283 |
| Y5 | 0,1056 | 0,0223 | 0,0856 | 0,0359 | 0,0000 | 0,0736 | 0,0694 | 0,0245 | 0,0747 |
| Y6 | 0,0368 | 0,0564 | 0,0198 | 0,0501 | 0,0838 | 0,0000 | 0,0004 | 0,0545 | 0,0010 |
| Y7 | 0,0354 | 0,0621 | 0,0234 | 0,0648 | 0,0930 | 0,0041 | 0,0000 | 0,0605 | 0,0033 |
| Y8 | 0,0852 | 0,0018 | 0,0672 | 0,0072 | 0,0269 | 0,0511 | 0,0493 | 0,0000 | 0,0520 |
| Y9 | 0,0358 | 0,0582 | 0,0187 | 0,0515 | 0,0849 | 0,0012 | 0,0018 | 0,0558 | 0,0000 |

Проводим группировку и получаем массив индексов. Этот массив говорит о том, в каком порядке должны быть использованы образы обучающей выборки при расчете изображения технического состояния i исследуемой системы.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 6 | 7 | 9 | 2 | 4 | 8 | 1 | 3 | 5 |

Результаты вычислений изображения без группировки:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 |
| -0,9589 | -0,9589 | -0,8322 | -0,8322 | -0,8698 | -0,8725 | -0,8814 | -0,8659 | -0,8617 |
| 0,9602 | 0,9602 | 0,9602 | 0,9656 | 0,9014 | 0,9211 | 0,9276 | 0,9375 | 0,9448 |
| 0,9093 | 0,9093 | 0,7924 | 0,8529 | 0,7234 | 0,6818 | 0,7029 | 0,7324 | 0,7253 |

Покажем таблицу абсолютных погрешностей.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| j | |eij(2)-eij(1)| | |eij(3)-eij(2)| | |eij(4)-eij(3)| | |eij(5)-eij(4)| | |eij(6)-eij(5)| | |eij(7)-eij(6)| | |eij(8)-eij(7)| |
| 1 | 0,12669197 | 0,00 | 0,03755215 | 0,0027317 | 0,00888114 | 0,01552064 | 0,00420261 |
| 2 | 0 | 0,00547245 | 0,06424324 | 0,01969183 | 0,00651316 | 0,00984823 | 0,00738617 |
| 3 | 0,11691712 | 0,06048911 | 0,12947032 | 0,04157955 | 0,02111281 | 0,02948072 | 0,00711874 |

Резултаты вычислений изображения после группировки

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 |
| -0,9258 | -0,9258 | -0,8493 | -0,8858 | -0,8408 | -0,8691 | -0,863 | -0,8767 | -0,8711 | -0,8725 |
| 0,96017 | 0,96017 | 0,97836 | 0,97999 | 0,97504 | 0,92176 | 0,93423 | 0,93793 | 0,94277 | 0,94911 |
| 0,8085 | 0,8085 | 0,8589 | 0,8421 | 0,80044 | 0,70735 | 0,70203 | 0,73164 | 0,76192 | 0,73454 |

Таблицу абсолютных погрешностей.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| j | |eij(2)-eij(1)| | |eij(3)-eij(2)| | |eij(4)-eij(3)| | |eij(5)-eij(4)| | |eij(6)-eij(5)| | |eij(7)-eij(6)| | |eij(8)-eij(7)| | |eij(9)-eij(8)| |
| 1 | 0,0765251 | 0,04 | 0,0450735 | 0,0283383 | 0,0061386 | 0,0137091 | 0,0055503 | 0,0013706 |
| 2 | 0,0181859 | 0,0016374 | 0,0049558 | 0,0532736 | 0,012463 | 0,0037062 | 0,0048318 | 0,0063437 |
| 3 | 0,0504005 | 0,0168002 | 0,0416584 | 0,09309 | 0,0053142 | 0,029609 | 0,0302756 | 0,0273797 |

Во втором варианте мы видим, что для нахождения изображения вида технического состояния i, потребовались все элементы обучающей выборки. И даже со всеми элементами условие сходимости не выполняется, что видно из таблицы абсолютных погрешностей. Это скорее всего говорит о том, что самый отличный от других элемент или группа элеменотов, который или которые попадают в выборку последние, имеют очень большое влияние на образ. Следовательно, имеет смысл поставить самые отличимые элементы вначале. Т.е. развернуть вектор с индексами обучающих образов.

Перевернутый массив индексов образов.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | 3 | 1 | 8 | 4 | 2 | 9 | 7 | 6 |

Резултаты вычислений изображения

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 |
| -0,8835 | -0,8835 | -0,8579 | -0,8915 | -0,8767 | -0,8979 | -0,8658 | -0,8791 | -0,8658 | -0,8725 |
| 0,9999 | 0,9999 | 0,9882 | 0,9789 | 0,9833 | 0,9284 | 0,9337 | 0,9408 | 0,9477 | 0,9491 |
| 0,5155 | 0,5155 | 0,7447 | 0,7995 | 0,7685 | 0,6818 | 0,6808 | 0,6990 | 0,7253 | 0,7345 |

Таблица абсолютных погрешностей.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| j | |eij(2)-eij(1)| | |eij(3)-eij(2)| | |eij(4)-eij(3)| | |eij(5)-eij(4)| | |eij(6)-eij(5)| | |eij(7)-eij(6)| | |eij(8)-eij(7)| | |eij(9)-eij(8)| |
| 1 | 0,0255936 | 0,03 | 0,0148203 | 0,0211448 | 0,0320555 | 0,0133009 | 0,0132943 | 0,0066656 |
| 2 | 0,0116355 | 0,0093509 | 0,0044175 | 0,054924 | 0,0053008 | 0,007086 | 0,0069737 | 0,0013827 |
| 3 | 0,2291731 | 0,0548743 | 0,0310214 | 0,0867079 | 0,0010594 | 0,018248 | 0,0262862 | 0,0092447 |

К сожалению желаемый результат не был получен, хотя в отличие от случая первой группировки мы имеем выполнение условия сходимости на последнем шаге.

Тексты программы, реализующей группировку.

function [ otvet ] = myObrazGroup(viborka)

pogrMax=0;

lenInitial=length(viborka(1,:));

for i=1:lenInitial

indexesRest(i)=i;

end

indexesMatch=0;

indexesNotMatch=0;

indexesToFinal=0;

indexesToFinalRest=0;

indexesFinal=0;

zzz=0;

while indexesRest;

zzz=zzz+1;

len = length(indexesRest);

for i=1:len

e=viborka(:,indexesRest(i));

current=indexesRest(i);

for j=1:len

q=viborka(:,indexesRest(j));

if zzz<2

pogr(indexesRest(i),indexesRest(j))=abs(sum((e-q)./e))./length(e);

end

if pogr(indexesRest(i),indexesRest(j))<=pogrMax

if indexesMatch

massLen=length(indexesMatch)+1;

indexesMatch(massLen)=indexesRest(j);

else

indexesMatch=indexesRest(j);

end

else

if indexesNotMatch==0

indexesNotMatch=indexesRest(j);

else

massLen=length(indexesNotMatch)+1;

indexesNotMatch(massLen)=indexesRest(j);

end

end

end

if length(indexesMatch)>=length(indexesToFinal)

indexesToFinal=indexesMatch;

indexesToFinalRest=indexesNotMatch;

end

indexesMatch=0;

indexesNotMatch=0;

end

if length(indexesToFinal)/length(indexesRest)<0.2

pogrMax=pogrMax+0.01;

elseif length(indexesRest) > 1 && length(indexesToFinal)<2

pogrMax=pogrMax+0.01;

else

indexesRest=indexesToFinalRest;

if indexesFinal

indexesFinal = [indexesFinal,indexesToFinal];

else

indexesFinal = indexesToFinal;

end

indexesToFinal=0;

end

end

otvet = [indexesFinal];