Artificial Intelligence

Rough Set

Task Report 3

Harun GÜLEÇ

Bialystok,2017

Task 1

At this task our main object is proposing inconsistent data, where the set of objects includes at least ten elements, the set of condition attributes includes at least three elements, and the number of decision classes must be at least three. For this purpose, I create a table that composed of basketball player success criteria's. And then I wrote required expressions.

NO	Training	Nutrition	Team Adaptation	Hight	Success
1	Y	Y	Y	N	T
2	Y	Y	N	N	I
3	Y	N	Y	N	T
4	N	Y	Y	N	TN
5	Y	Y	Y	Y	T
6	N	N	Y	Y	N
7	Y	Y	N	Y	I
8	Y	N	Y	Y	T
9	N	Y	Y	Y	I
10	N	N	N	Y	N

$IND(A) = \{(x,y) \in U \times U = \forall_{k \in A}\, k(x) = k(y)\}$ for $A = \{$Training, Nutrition$\}$

$[x]_{IND(A)} = \{y \in U : (x,y) \in IND(A)\}, \quad x \in U$

$IND_{(B)} = \{(x,y) \in U \times U = \forall_{k \in B}\, k(x) = k(y)\}$ for $B = \{$Training, Team Adaptation$\}$

$[x]_{IND(B)} = \{y \in U = (x,y) \in IND(B)\}, \quad y \in U$

$X_T = \{x \in U : success(x) = Team\}$ $\quad X_I = \{x \in U : success(x) = Individual\}$

$X_N = \{x \in U : success(x) = None\}$

$Low(AS_B, X_T) = \{x \in U : [x]_{IND(B)}) \subseteq X_T\}$

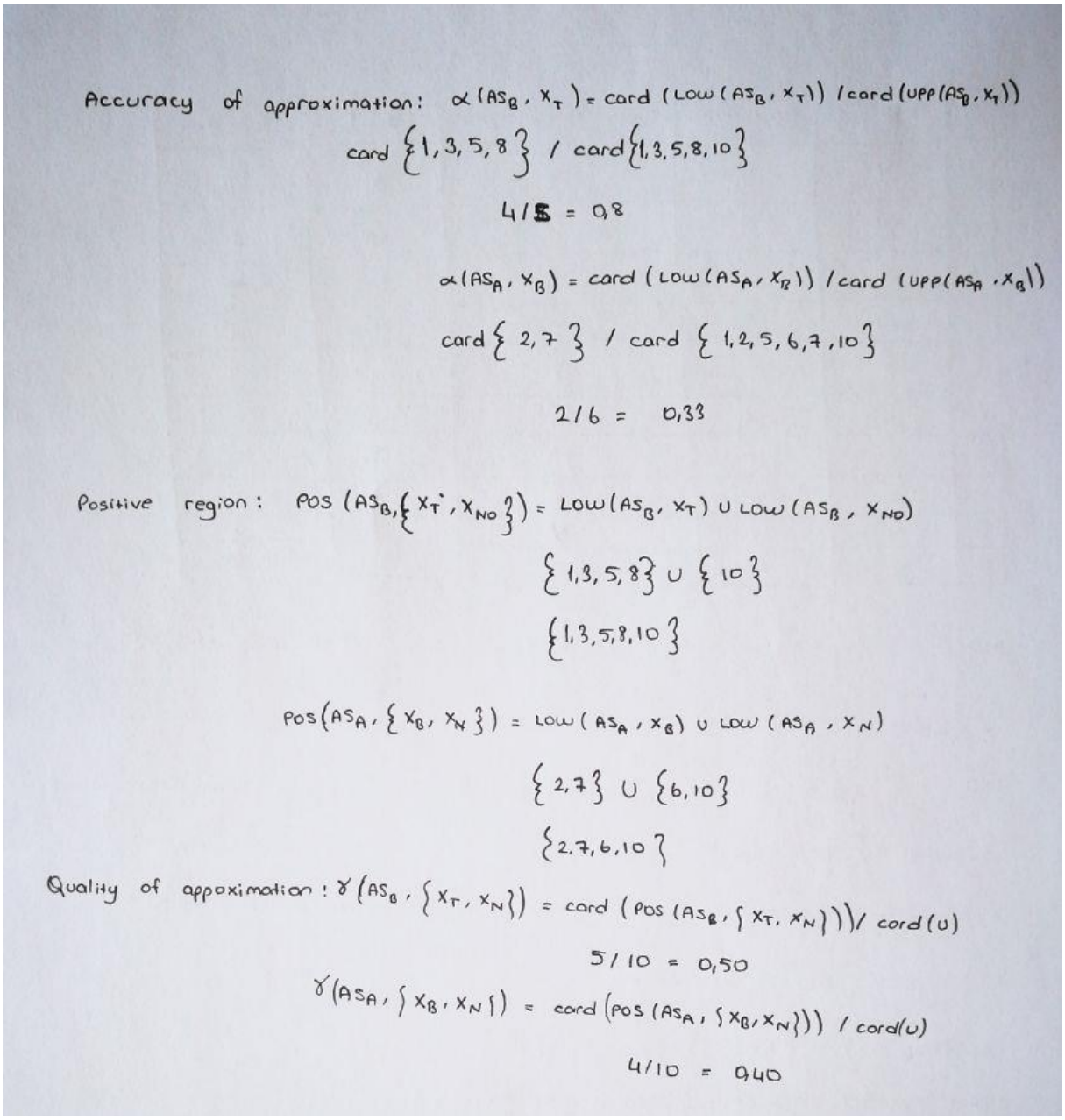
$UPP(AS_B, X_T) = \{x \in U : [x]_{IND(B)} \cap X_T \neq \emptyset\}$

$Low(AS_A, X_B) = \{x \in U : [x]_{IND(A)} \subseteq X_B\}$

$UPP(AS_A, X_B) = \{x \in U : [x]_{IND(A)} \cap X_B = \emptyset\}$

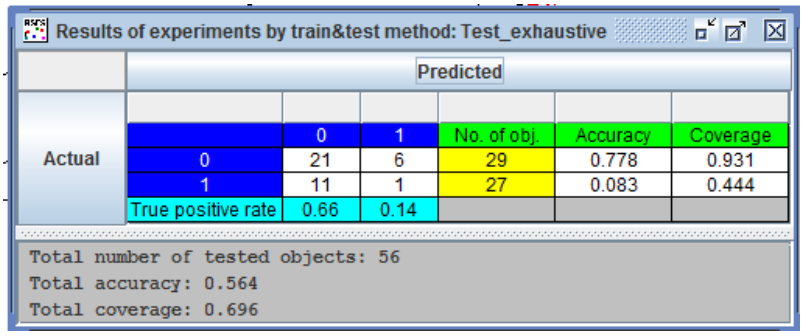
$Low(AS_B, X_N) = \{x \in U : [x]_{IND_B} \subseteq X_N\}$

$Low(AS_A, X_N) = \{x \in U : [x]_{IND_A} \subseteq X_N\}$



Task 2

At this task our main object is classifying data CreditCards.tab using the RSES application. Firstly, I classified data using *exhaustive, genetic, cover, lem2 algorithms* without discretization and the results shown below.



Results of experiments by train&test method: Test_genetic

Actual \ Predicted	0	1	No. of obj.	Accuracy	Coverage
0	21	6	29	0.778	0.931
1	11	1	27	0.083	0.444
True positive rate	0.66	0.14			

Total number of tested objects: 56
Total accuracy: 0.564
Total coverage: 0.696

Results of experiments by train&test method: Test_cover

Actual \ Predicted	0	1	No. of obj.	Accuracy	Coverage
0	10	2	29	0.833	0.414
1	6	1	27	0.143	0.259
True positive rate	0.62	0.33			

Total number of tested objects: 56
Total accuracy: 0.579
Total coverage: 0.339

Results of experiments by train&test method: Test_lem2

Actual \ Predicted	0	1	No. of obj.	Accuracy	Coverage
0	11	0	29	1	0.379
1	3	0	27	0	0.111
True positive rate	0.79	0			

Total number of tested objects: 56
Total accuracy: 0.786
Total coverage: 0.25

As we can understand clearly that biggest accuracy rate is for lem2 algorithm. It means lem2 algorithm better than others to make exact decision, but coverage is not enough. And then, I test same algorithms with discrete table. The results shown below.

Results of experiments by train&test method: TestD_exhaustive

Actual \ Predicted	0	1	No. of obj.	Accuracy	Coverage
0	21	8	29	0.724	1
1	12	13	27	0.52	0.926
True positive rate	0.64	0.62			

Total number of tested objects: 56
Total accuracy: 0.63
Total coverage: 0.964

Results of experiments by train&test method: TestD_genetic

Actual \ Predicted	0	1	No. of obj.	Accuracy	Coverage
0	21	8	29	0.724	1
1	12	13	27	0.52	0.926
True positive rate	0.64	0.62			

Total number of tested objects: 56
Total accuracy: 0.63
Total coverage: 0.964

Results of experiments by train&test method: TestD_cover

Actual	Predicted 0	Predicted 1	No. of obj.	Accuracy	Coverage
0	14	7	29	0.667	0.724
1	11	9	27	0.45	0.741
True positive rate	0.56	0.56			

Total number of tested objects: 56
Total accuracy: 0.561
Total coverage: 0.732

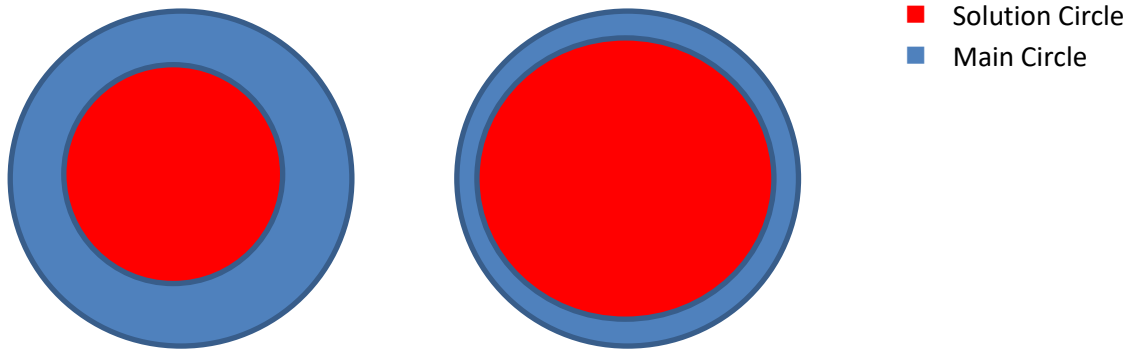
Results of experiments by train&test method: TestD_lem2

Actual	Predicted 0	Predicted 1	No. of obj.	Accuracy	Coverage
0	14	4	29	0.778	0.621
1	9	8	27	0.471	0.63
True positive rate	0.61	0.67			

Total number of tested objects: 56
Total accuracy: 0.629
Total coverage: 0.625

After discretization process, Genetic and Exhaustive algorithms are better than others because accuracy value is higher. So these can make decision more exactly.

Also when we compare the results of after discretization and before, After discretization coverage value bigger than before. It means our circle more closer to main objective circle.



Before Discretization

After Discretization

Conclusion

When we are working on rough data set, we cannot say exact definitions but we can be in a different approximity. There are 2 approximation lower and upper approximations. In lower approximation we only take into account the situations that we are sure about that. In upper approximation we don't take into account the situations that we are sure about it is not in our circle. In addition, in both approximation when we discrete the table we can achieve more clear results.