

A List of Metrics from the DMOP Ontology (94 Classes)

<http://www.e-lico.org/?q=book/export/html/173>

<http://www.e-lico.org/sites/all/public/trunk/software/ontologies/dmo/DMOP/DMOP.owl>

By: Noureddin Sadawi

12 Feb 2014

1) **AlgorithmCharacteristic**

- A) BiasVarianceProfile:
- B) ClassificationProblemType:
- C) CoordinateSystem:
- D) FeatureEvaluationContext:
- E) FeatureEvaluationTarget:
- F) HandlingOfCategoricalFeatures:
- G) HandlingOfClassificationCosts:
- H) HandlingOfContinuousFeatures:
- I) HandlingOfInstanceWeights:
- J) LearningPolicy:
- K) RandomComponent:
- L) RangeOfNeighborhood:
- M) RuleInductionStrategy:
- N) SampleHandlingForRuleInduction:
- O) ToleranceToClassImbalance:
- P) ToleranceToCorrelatedFeatures:
- Q) ToleranceToHighDimensionality:
- R) ToleranceToIrrelevantFeatures:
- S) ToleranceToMissingValues:
- T) ToleranceToNoise:
- U) ToleranceToWideDataSets:
- V) TransformationFunction:
- W) TreeBranchingFactor:

X) UniquenessOfSolution:

2) **DataCharacteristic**

A) **DataSetCharacteristic**

- I) AverageAbsoluteFeatureCorrelation: METAL characteristic: Average absolute correlation between continuous features.
- II) AverageCategoricalFeaturePairsMutualInformation: METAL characteristic: Average mutual information between pairs of categorical features.
- III) AverageFeatureEntropy: METAL characteristic: Average feature Entropy
- IV) BetweenGroupsSumSquaresCrossProducts: METAL characteristic: A matrix containing the difference between the matrix of total and the matrix of within-groups sums of squares and cross products.
- V) EigenValuesLinearDiscriminantFunctions: METAL characteristic: A vector of eigen values of linear discriminant functions.

VI) **LabeledDataSetCharacteristic**

- 1) AverageSVMFeatureWeight:
- 2) **CategoricalLabeledDataSetCharacteristic**
 - a) AverageMutualInformation: METAL characteristic: Average mutual information
 - b) AverageReliefFeatureWeight:
 - c) CanonicalCorrelationBestLinearCombination: METAL characteristic: Canonical correlation of the best linear combination of features to distinguish between classes.
 - d) ClassAbsoluteFrequencies: METAL characteristic: Absolute class frequencies. Stored in a vector indexed by each class value.
 - e) ClassCovarianceMatrices: METAL characteristic: Class covariance matrices. Stored in a vector indexed by class and each containing a matrix of (features x features)
 - f) ClassEntropy: METAL characteristic: Class entropy.
 - g) ClassRelativeFrequencies: METAL characteristic: Relative class frequencies. Stored in a vector indexed by each class value.
 - h) ErrorRateOf1NNClassifier: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
 - i) ErrorRateOfLinearClassifierLP: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.

- j) FeatureMutualInformationPerClass: METAL characteristic: For each categorical feature, the mutual information between the feature and the class. It is stored in a vector indexed by each categorical feature.
- k) FeatureValueFrequenciesPerClass: METAL characteristic: For each k value of each j categorical feature and each i class, the proportion of cases that have the k value in the j feature and belong to the i class. It is stored in a vector indexed by each categorical feature and containing a flat contingency tables that combine the values of the categorical feature with the class values.
- l) MaximumFeatureEfficiency: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- m) MaximumFishersDiscriminantRatio: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- n) MinimumSumOfErrorDistanceByLP: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- o) NonLinearityOf1NNClassifier: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- p) NonLinearityOfLinearClassifierLP: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- q) NumberOfClasses:
- r) ProportionOfBoundaryPoints: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- s) ProportionPointsWithRetainedAdherence: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- t) RatioOfAverageIntraInterDistances: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- u) VolumeOfOverlapRegion: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- 3) ContinuousLabeledDataSetCharacteristics:
- VII) NoiseSignalRatio: METAL characteristic: Noise signal ratio
- VIII) NumberOfCategoricalFeatures:
- IX) NumberOfContinuousFeatures:

- X) NumberOfFeatures:
- XI) NumberOfHOutliers: METAL characteristic: Number of continuous features with outliers.
- XII) NumberOfInstances:
- XIII) NumberOfInstancesPerFeature: From Mitra Basu and Tin Kam Ho. Data Complexity in Pattern Recognition. Springer, 2006.
- XIV) ProportionOfCategoricalFeatures:
- XV) ProportionOfHOutliers: METAL characteristic: Proportion of continuous features with outliers.
- XVI) TotalSumSquaresCrossProducts: METAL characteristic: Matrix of total sums of squares and cross products of features.
- XVII) WithinGroupsSumSquaresCrossProducts: METAL characteristic: matrix of within-groups sums of squares and cross products of features.

B) **FeatureCharacteristic**

I) **CategoricalFeatureCharacteristic**

- 1) FeatureClassMutualInformation: FeatureClassMutualInformation: For each categorical feature, the mutual information between the feature and the class. It is stored in a vector indexed by each categorical feature. [From METAL project].
- 2) FeatureEntropy: METAL characteristic: For each categorical feature, the feature entropy.
- 3) FeatureFrequentValue: A categorical value should be marked as frequent, iff it is 50% more frequent than we would expect under a uniform distribution. (e.g. 75% for a binary value)
- 4) FeatureModalValue:
- 5) FeatureRareValue: A categorical value should be marked as rare, iff it is 50% less frequent than we would expect under a uniform distribution. (e.g. 25% for a binary value)
- 6) FeatureValueFrequencies: METAL characteristic: For each k value of the feature, the value frequency. It is stored in a vector indexed by each feature value.
- 7) FeatureVeryFrequentValue: A categorical value should be marked as very frequent, iff it is 90% more frequent than we would expect under a uniform distribution. (e.g. 95% for a binary value)
- 8) FeatureVeryRareValue: A categorical value should be marked as very rare, iff it is 90% less frequent than we would expect under a uniform distribution. (e.g. 5% for a binary value)

9) NumberOfDistinctValues:

II) **ContinuousFeatureCharacteristic**

1) FeatureCorrelation: METAL characteristic: For each continuous feature, its correlation with other continuous features. Stored in a vector indexed by each continuous features.

2) FeatureHOutlier: METAL characteristic: For each continuous feature, the ratio between the standard deviation and the standard deviation of alpha trimmed mean. If the standard deviation is 0, then the ratio is set 1.

3) FeatureMaximumValue:

4) FeatureMeanValue:

5) FeatureMinimumValue:

6) FeatureStandardDeviation:

III) NumberOfMissingValues:

IV) ProportionOfMissingValues:

3) **HypothesisCharacteristic**

A) Interpretability:

B) **ModelCharacteristic**

I) ModelComplexity:

II) ModelPerformance:

C) PatternSetCharacteristic:



Figure 1

7

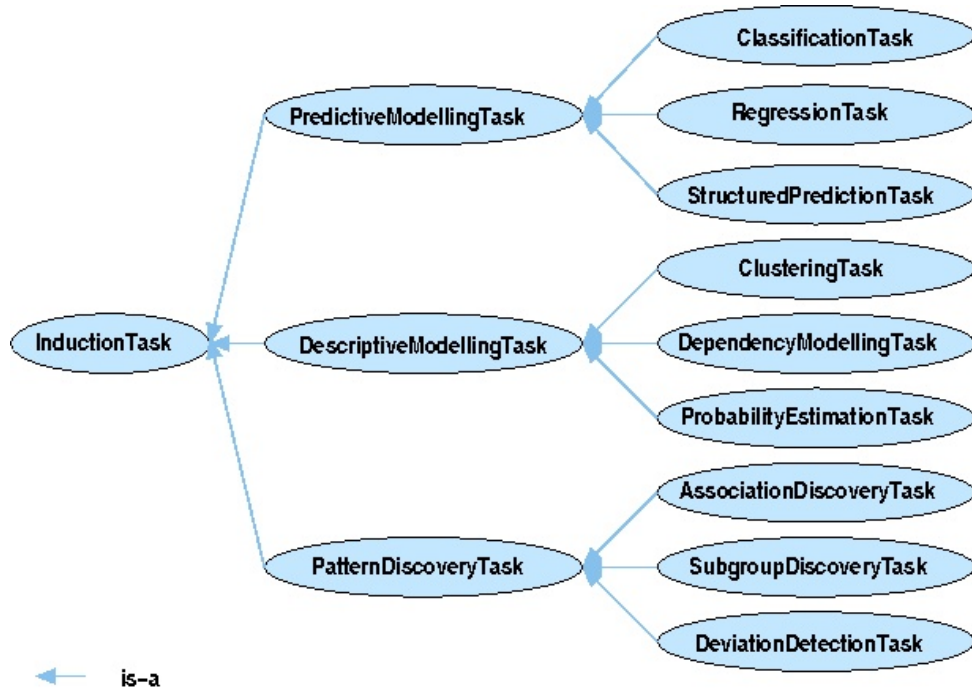


Figure 3: dmop-task-hierarchy

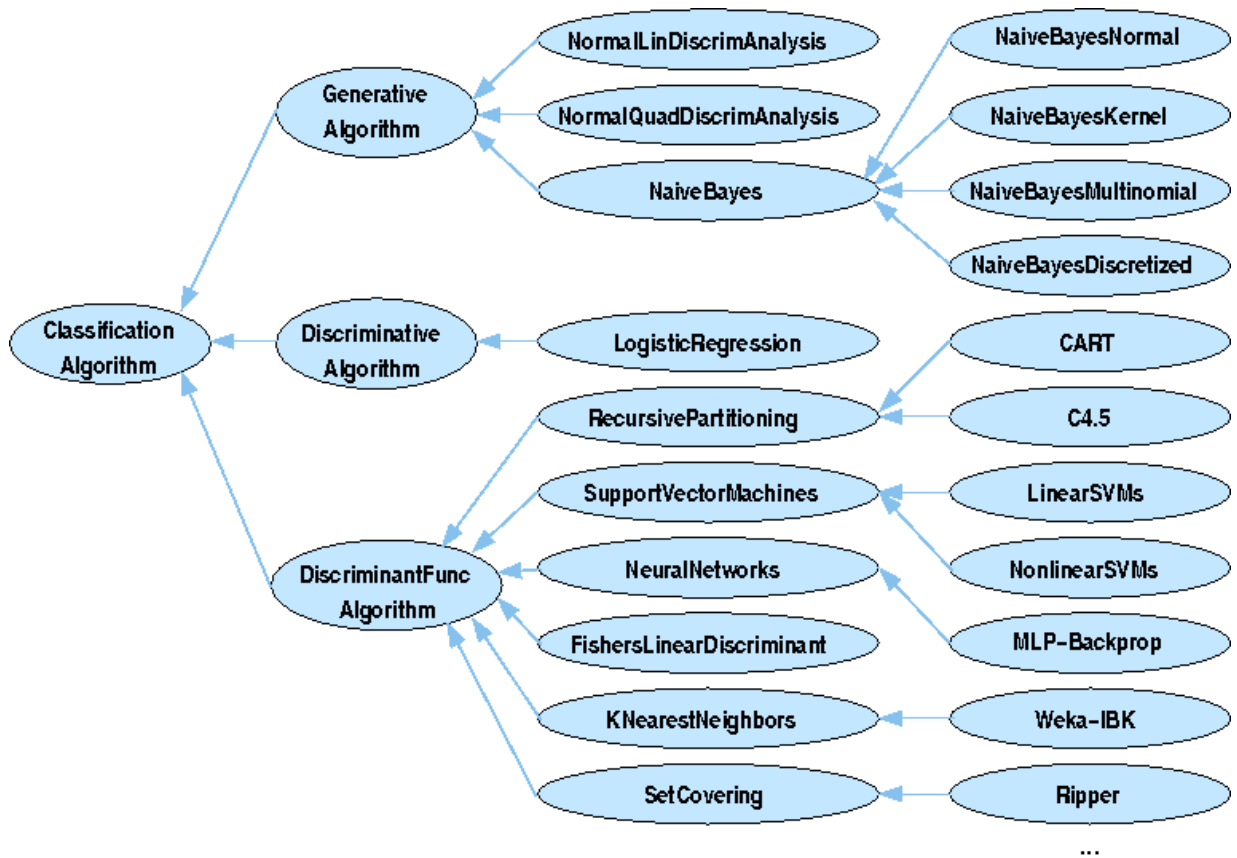


Figure 4: dmop-classif-algo-tree

