



## Performance Evaluation of Gene Based Ontology Using Attribute Selection Methods

<sup>1</sup> B. Revathi, PG Student, B.V.C College of Engineering, Rajahmundry, India

<sup>2</sup> M.V.R Narasimha Rao, Research Scholar, JNTUK,  
Associate Professor, B.V.C College of Engineering, , Rajahmundry ,India.

### ABSTRACT:

With the mounting quantity of ageing-related data on model organisms obtainable on the web, in specific linked to the heredities of ageing, it is opportune to smear data mining methods to that data, in instruction to attempt to determine novel patterns that may backing ageing research. The foremost new facet of the proposed feature selection methods is that they adventure hierarchical relationships in the set of features Gene Ontology terms in order to progress the prognostic exactness of the Naive Bayes and 1-Nearest Neighbor (1-NN) classifiers, which are castoff to pigeonhole model organisms' genes into pro-longevity or anti-longevity genes. The marks show that our hierarchical feature selection methods, when used organized with Naive Bayes and 1-NN classifiers, get advanced prognostic correctness than the normal deprived of feature selection.

**KEYWORDS:** Ageing, data mining, classification, gene ontology, feature selection, naive Bayes, 1-nearest neighbor.

### INTRODUCTION:

The classification task of data mining where each occurrence thing existence classified contains of a set of topographies and a class mutable. The objective of a classification algorithm is to shape, from a set of training instances the training set a classification model that forecasts the value or label of the class variable for a case, given the values of the features for that instance. The classification model is then rummage-sale to forecast the class values of a dissimilar set of testing instances the testing set. Henceforth, the testing set is secondhand to amount the predictive presentation, or simplification aptitude, of the model constructed from the training set. Ageing is a extremely multifaceted organic course which is still unwell unspoken, and much more investigation is wanted in this area. Due to the countless trouble and moral issues related with conducting ageing experiments in humans, study on

the biology of ageing is usually done by using perfect creatures

### PROBLEM DEFINITION:

In the classification task, when the number of features is large it is shared to smear feature selection methods, beforehand smearing a classification algorithm to the data. Feature selection methods goal to choice a subsection of the most pertinent and non-redundant features out of all input features, in instruction to attempt to advance the prognostic accurateness of a classification algorithm. Note that feature selection is a solid computational problem, since the quantity of candidate solutions is  $2^m - 1$ , where  $m$  is the number of features

### PROPOSED APPROACH:

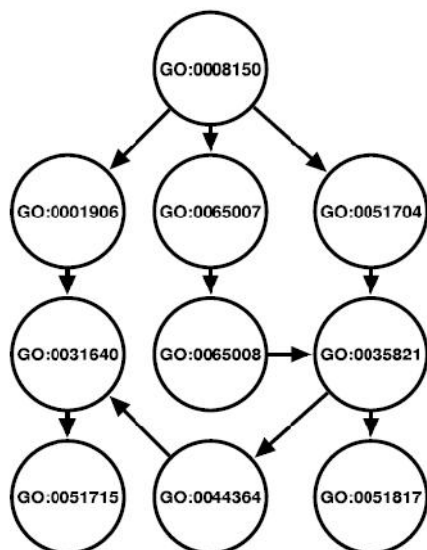
Two new feature selection methods that adventure hierarchical relationships among GO terms, in directive to curtail the redundancy in the selected GO terms. We use the term "hierarchical feature selection" to mention to feature selection methods that manage with hierarchical relationships among features. The proposed hierarchical feature selection methods work with the Naive Bayes and 1-NN (nearest neighbor) classifiers in the context of "lazy learning" where a set of features is designated exactly for each testing case gene. This is in difference to the much more common "eager learning" situation, where the same set of features is designated and used to categorize all testing set examples.

### PROPOSED METHODOLOGY:

#### HIERARCHICAL INFORMATION-PRESERVING:

It concentrations only on lessening the dismissal in the set of selected GO terms, discounting the importance values of individual GO terms. The HIP feature selection method usages only material about the GO term feature values and their graded relationships; the GO terms' bearing values are castoff only by the two other feature selection methods described later

### SYSTEM ARCHITECTURE:



A small part of Gene Ontology DAG.

### MOST RELEVANT GO TERMS:

It completes feature selection seeing both the relevance value of individual GO terms and the joblessness mid hierarchically-related GO terms. Like the HIP method, for every GO term  $t$  in the current occurrence being confidential, MR first recognizes the sets of GO terms whose values are indirect by the worth of  $t$  in that occurrence. The MR method has the benefit that in over-all it selects GO terms with advanced significance values than the GO terms designated by the HIP method which disregards GO term relevance standards.

### NAIVE BAYES:

The extrapolative precision of Naive Bayes is searching to the extrapolative power of individual features; and its correctness can also be damaged by the practice of much terminated features. Though, immaterial or jobless features can be detached by a feature selection method in a preprocessing step, Naive Bayes is a general classifier due to its effortlessness, comparatively influential prognostic aptitude and its decent interpretability

### ALGORITHM:

```

LoadDAGbyrinkingallGOTennsinOS:2:LoadDS<r''''''>
;

```

```

3:LoadDS<T•R>:

```

```

4:foreachGO;in DAGdo

```

```

5:LoadA+(G0;1)fromDAG;

```

```

6:LoadD(GOIJ)fromDAG;

```

```

7:Load S(GO;)+---Selected":

```

```

8:CalculateR(C01)inDS,7_ ;

```

```

9:endifor

```

```

10:foreach/<•>eDS<7,,,do

```

```

11:foreachCO;eDAG do

```

```

12:ifV(C01•)=Ithen

```

```

13:foreachParlz,fromC0 torootinDAGdo

```

```

14 end for

```

```

15:else

```

```

16:foreachDtreDIGOJdo

```

```

17:S{Dtt }+•O.ltttt1•;

```

```

18:endifor

```

```

19:endif

```

```

20:endifor

```

```

21.,+•{G0:S(G0>= S../«ttt;

```

```

22 SVMCDS,_,/);

```

```

23 Reigo\GO,S!GO,)+•SNct,J•;

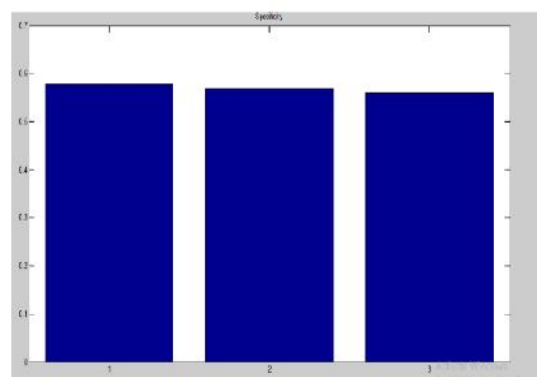
```

```

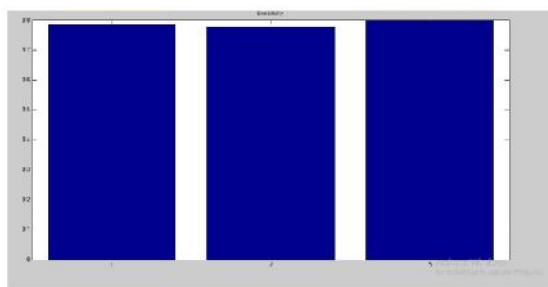
24 endfor

```

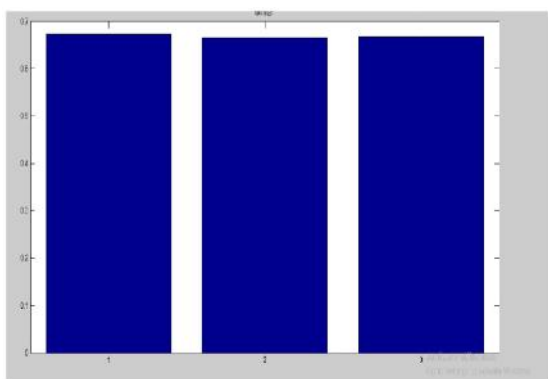
### RESULTS:



Graph of specificity for the gene



Graph of Sensitivity for the gene



Graph of  $G_{mean}$  for the gene

## CONCLUSION:

Graded article collection methods were considered to adventure material in ordered relationships between Gene Ontology terms castoff as features in direction to diminish severance in the set of selected features. The use of the lazy teaching approach tolerable us to select a subset of GO terms definitely for each testing instance (gene) existence classified by Naive Bayes and 1-NN. The current feature selection methods have two part as a relevance measure and a feature hierarchy- processing technique that uses the feature hierarchy to resolve which features should be detached.

## REFERENCES:

- [1] E. J. Masoro, "Overview of caloric restriction and ageing," *Mech. Ageing Develop.*, vol. 126, no. 9, pp. 913–922, Sep. 2005.
- [2] C. J. Kenyon, "The genetics of ageing," *Nature*, vol. 464, no. 7288, pp. 504–512, Mar. 2010.
- [3] A. A. Freitas and J. P. de Magalhães, "A review and appraisal of the DNA damage theory of ageing,"

*Mutation Res.*, vol. 728, nos. 1/2, pp. 12–22, Jul.–Oct. 2011.

[4] T. B. L. Kirkwood and S. N. Austad, "Why do we age?" *Nature*, vol. 408, no. 6809, pp. 233–238, Nov. 2000.

[5] J. P. de Magalhães, "How ageing processes influence cancer," *Nat. Rev. Cancer*, vol. 13, no. 5, pp. 357–365, May 2013.

[6] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2011.

[7] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Norwell, MA, USA: Kluwer, 1998.

[8] R. Tacutu, T. Craig, A. Budovsky, D. Wuttke, G. Lehmann, D. Taranukha, J. Costa, V. E. Fraifeld, and J. P. de Magalhães, "Human ageing genomic resources: Integrated databases and tools for the biology and genetics of ageing," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D1027–D1033, Jan. 2013.

[9] The GO Consortium, "Gene Ontology: Tool for the unification of biology," *Nat. Genetics*, vol. 25, no. 1, pp. 25–29, May 2000.

[10] D. W. Aha, *Lazy Learning*. Norwell, MA, USA: Kluwer, 1997.

[11] R. B. Pereira, A. Plastino, B. Zadrozny, L. H. de C. Merschmann, and A. A. Freitas, "Lazy attribute selection: Choosing attributes at classification time," *Intell. Data Anal.*, vol. 15, no. 5, pp. 715–732, Aug. 2011.

[12] C. Wan and A. A. Freitas, "Prediction of the pro-longevity or anti longevity effect of *Caenorhabditis Elegans* genes based on Bayesian classification methods," in *Proc. IEEE Int. Conf. Bioinformat. Biomed.*, Shanghai, China, Dec. 2013, pp. 373–380.

[13] A. A. Freitas, O. Vasieva, and J. P. de Magalhães, "A data mining approach for classifying DNA repair genes into ageing-related or non-ageing-related," *BMC Genomics*, vol. 12, no. 27, pp. 1–11, Jan. 2011.

[14] Y. Fang, X. Wang, E. K. Michaelis, and J. Fang, "Classifying aging genes into DNA repair or non-DNA repair-related categories," in *Proc. 9th Int. Conf. Intell. Comput. Theories Technol.*, 2013, pp. 20–29.

[15] Y. H. Li, M. Q. Dong, and Z. Guo, "Systematic analysis and prediction of longevity genes in *Caenorhabditis elegans*," *Mech. Ageing Develop.*, vol. 131, nos. 11/12, pp. 700–709, Nov./Dec. 2010.



B.Revathi is a student of B.V.C College of Engineering, Rajahmundry. Presently she is pursuing her M.Tech [computer science engineering] from this college and she received her B.Tech from pragati engineering

college, affiliated to JNT University, Kakinada in the year 2011. Her area of interest includes Data mining, computer networks, all current trends and techniques in Computer Science



M.V.R Narasimharao pursuing Ph.D in Dept of CS, JNTUK. His research area is Data Mining. He has published papers in different international journals. He also working as Associate Professor in

Dept of CSE and has 15 years of experience, B.V.C College of Engineering, Rajahmundry.