

Protein Classification By feature extraction

Computer Science, UCL, London, WC1E 6BT, UK

Aim : The goal of this assignment is to provide an automated system that is able to classify proteins (Amino Acid sequences) into four classes each being a subcellular locations : **[Cytosolic, Secreted, Nuclear, Mitochondrial]**

Improvements: In order to improve the results of the classifier, deepening the feature extraction method seems to be the way to go. Another method would be to use neural network techniques.

Contact: ucabsm1@ucl.ac.uk

Currently, There is a growing need for fully automated methods to analyse amino acids sequences. One of the process that need to be automated is **the identification of the protein's subcellular location**. This problem can be splitted into two sub problems :

- Within the last few years the complete sequence has been determined for over 3000 genomes. This has created the need for fully automated methods to analyse the vast amount of sequence data now available. The assignment of a function for a given protein has proved to be difficult where no clear homology to proteins of known function exists. Knowing the subcellular location of a protein (i.e. where in the cell it is found) may give some clue as to its possible function, making an automated method that assigns proteins to a certain subcellular location a useful tool for analysis.

$$\sum x + y = Z \quad (1)$$

Equation (1) Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text. Figure 2 shows that the above
method Text Text Text Text Text Text Text Text Text Text Text Text Text. Bofelli
et al., 2000 might want to know about text text text text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text. Figure 2 shows that the above method Text
Text Text Text Text Text Text Text Text Text Text Text. Bofelli *et al.*, 2000
might want to know about text text text text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text.
Figure 2 shows that the above method Text Text Text Text Text Text Text
Text Text Text Text Text. Bofelli *et al.*, 2000 might want to know about
text text text text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows that the
above method Text Text Text Text Text Text Text Text Text Text Text Text.
Bofelli *et al.*, 2000 might want to know about text text text text

- Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text. Figure 2 shows that the above method
Text Text Text Text Text Text Text Text Text Text Text Text Text. Bofelli *et al.*,
2000 might want to know about text text text text Text Text Text Text Text

head1	head2	head3	head4
row1	row1	row1	row1
row2	row2	row2	row2
row3	row3	row3	row3
row4	row4	row4	row4

FPO

above method Text Text Text Text Text Text Text Text Text Text Text.
Bofelli *et al.*, 2000 might want to know about text text text text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text. Figure 2 shows that the above method Text
Text Text Text Text Text Text Text Text Text Text Text Text. Bofelli *et al.*, 2000
might want to know about text text text text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text.
Figure 2 shows that the above method Text Text Text Text Text Text Text Text
Text Text Text Text Text. Bofelli *et al.*, 2000 might want to know about
text text text text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text. Figure 2 shows that the
above method Text Text Text Text Text Text Text Text Text Text Text Text Text.
Bofelli *et al.*, 2000 might want to know about text text text text

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text. Figure 2 shows that the above method Text
Text Text Text Text Text Text Text Text Text Text Text Text. Bofelli *et al.*, 2000
might want to know about text text text text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text.
Figure 2 shows that the above method Text Text Text Text Text Text Text Text
Text Text Text Text Text. Bofelli *et al.*, 2000 might want to know about
text text text text Text Text Text Text Text Text Text Text Text Text Text.

Bofelli *et al.*, 2000 might want to know about text text text text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text. Figure 2 shows that the above method Text Text Text
Text Text Text Text Text Text Text Text Text Text. Bofelli *et al.*, 2000 might
want to know about text text text text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text.
Figure 2 shows that the above method Text Text Text Text Text Text Text Text
Text Text Text Text Text.

1. this is item, use enumerate
2. this is item, use enumerate
3. this is item, use enumerate

(Table 1) Text Text Text Text Text Text Text Text Text Text Text Text
Text Text Text Text Text Text Text Text Text. Figure 2 shows that the
above method Text Text Text Text Text Text Text Text Text Text Text.

Text Text. Figure 2 shows that the above method Text Text Text Text

Text Text Text Text Text Text Text. Bofelli *et al.*, 2000 might want to know about text text text text

This work has been supported by the... Text Text Text Text.

Bofelli,F., Name2, Name3 (2003) Article title, *Journal Name*, **199**, 133-154.

Bag,M., Name2, Name3 (2001) Article title, *Journal Name*, **99**, 33-54.

Yoo,M.S. *et al.* (2003) Oxidative stress regulated genes in nigral dopaminergic neuron cell: correlation with the known pathology in Parkinson's disease. *Brain Res. Mol. Brain Res.*, **110**(Suppl. 1), 76-84.

Lehmann,E.L. (1986) Chapter title. *Book Title*. Vol. 1, 2nd edn. Springer-Verlag, New York.

Crenshaw, B.,III, and Jones, W.B.,Jr (2003) The future of clinical cancer management: one tumor, one chip. *Bioinformatics*, doi:10.1093/bioinformatics/btn000.

Auhtor,A.B. *et al.* (2000) Chapter title. In Smith, A.C. (ed.), *Book Title*, 2nd edn. Publisher, Location, Vol. 1, pp. ???-???.

Bardet, G. (1920) Sur un syndrome d'obesite infantile avec polydactylie et retinite pigmentaire (contribution a l'etude des formes cliniques de l'obesite hypophysaire). PhD Thesis, name of institution, Paris, France.