



Zürich, September 28, 2011

Topic: Clustering of Amino Acids Profiles

Amino acids are chemical units or 'building blocks' of proteins and play central role in metabolism of living organisms. There are 22 essential amino acids and their containment varies between different feed types of domestic animals. Therefore, it is crucial task for farmers to ensure the right balance of different feed types for each species of domestic animals depending on the age, season any many other factors.

The containment of amino acids is determined based on the chemical analyses. This process is time consuming, expensive and, therefore, results on various mistakes. For example, due to wrong setup of chemical analyses the derived containment of some amino acid(s) might be different from the true value. Another example, are misclassified or unclassified feed types. In this case, based on the amino acids it is possible to correctly classify the feed type or even to determine a new one.

In this project we will use clustering analyses techniques on amino acid profiles in order to detect outliers and correctly classify the feed data. In particular we will use DBScan and Optics which are efficient and accurate density based clustering techniques for high dimensional data. The project is divided into following tasks:

1. learning of DBScan and Optics that are described in
 - M. Ester, H.-P. Kriegel et al: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In KDD'96, pp.226-231, Portland, Oregon, Aug. 1996.
 - M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. OPTICS: ordering points to identify the clustering structure. In SIGMOD '99. ACM, New York, NY, USA, 49-60.
2. implementation of the above clustering techniques in any of programming languages



and graphical visualization of the results;

3. experimental evaluation and comparison of the above techniques on amino acids profiles.

The deliverables are:

1. report of 10-20 pages;
2. 10 minutes presentation.

Supervisor:

- Andrej Taliun

Starting date: to be assigned

Ending date: to be assigned

Department of Informatics, University of Zurich

Prof. Dr. Michael Böhlen