University of Zurich[UZH]

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

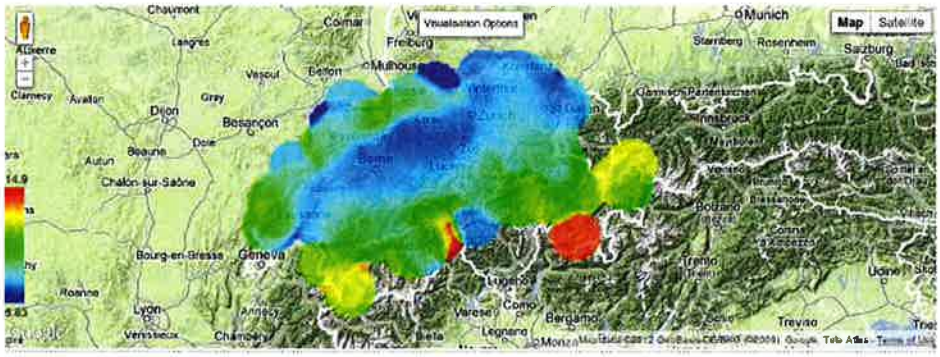
Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, November 7, 2012

Master Basismodul in Informatik
Datenbanktechnologie
Topic: Dynamic Data Summary Structures and Computation of the Color Plots for the Temporal Analyses of Nutritive Containment

For the regional analyses of the nutritive containment the on-line web application of the Swiss Feed Database offers color plots. Consider the below Figure. The user aims to compare the containment of nutrient 'calcium' in hey across different regions of Switzerland. In total, there are 2000 measurements of this nutrient from 555 different locations. The colored image on the top of the map employs hot and cold colors to represent high and low concentration of calcium. It can be easily observed that mountain hey has much higher concentration of calcium comparing to the hey in the central part of Switzerland.

Computation of the color plots is done with a help of Kernel Density Estimation and Kernel Regression. These methods take into account all measurements of the query result and provide the expected nutritive value even in the areas that are between origins of the feed samples. On a technical level, the efficient computation of the color plots requires to maintain a data summary structure. Specifically, we compute the density on grid points of a sparse regular grid and, then, interpolate the density between the grid points. Such an approach is efficient for the analyses of the static data, however, does not scale for the temporal analyses of the nutritive containment: while computing multiple color plots for the history of nutrient measurements, we must recompute the entire summary structure for each distinct time stamp in the database.



LIMS-Nr.	Date	Canton	PLZ	Futtermittel	Ca[g/kg TS]
1751 xxx01	Feb 1, 2006	Thurgau	8564	Heu / Emd gemischt	6.530
1752 xxx01	Feb 1, 2006	Thurgau	8508	Heu / Emd gemischt	6.187
1753 xxx02	Feb 1, 2006	Thurgau	8506	Heu / Emd gemischt	6.387
1754 xxx47	Feb 1, 2006	Thurgau	8264	Heu / Emd gemischt	7.284
1755 xxx01	Feb 1, 2007	Thurgau	8264	Heu / Emd gemischt	6.310
1756 xxx47	Feb 1, 2006	Thurgau	8272	Heu / Emd gemischt	6.718
1757 xxx99	Feb 1, 2010	Thurgau	8273	Dürrfutter nicht spezifiziert	7.478
1758 xxx47	Feb 1, 2006	Thurgau	8268	Heu / Emd gemischt	7.674
1759 xxx98	Feb 1, 2010	Thurgau	8268	Dürrfutter nicht spezifiziert	6.649
1760 xxx01	Dec 1, 2011	Thurgau	8268	Heu / Emd gemischt	6.732
1761 xxx01	Feb 1, 2007	Thurgau	8252	Heu / Emd gemischt	6.910
1762 xxx04	Feb 1, 2006	St. Gallen		Heu / Emd gemischt	7.599

The goal of this project is to investigate the Kernel Density Estimation and Kernel Regression for the computation of the color plots in the Swiss Feed Database. The student will report about the efficiency of different Kernel functions, impact of the optimal bandwidth parameter and practically investigate the kernel density estimation on the feed data. In his studies the student will experimentally evaluate how the granularity of the grid and the size of the data affects the performance and estimation quality. The deliverables of the project are:

1. Implementation of the kernel density estimation on a regular grid. The student will use Epanechnikov and Gaussian Kernel Functions, and the data of the Swiss Database as an input.
2. Experiments evaluating the impact of the bandwidth parameter, the size of grid and the size of the input data.
3. Report of 5-15 pages.
4. Oral exam (30 minutes).

The list of the literature is:

1. Scott, D.W. (1992). Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley&Sons, New York.
2. Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Supervisor:

- Andrej Taliun

Starting date: 19.10.2012

Ending date: 26.02.2013

Department of Informatics, University of Zurich

Prof. Dr. Michael Böhlen