MATHEMATICAL AND COMPUTATIONAL LINGUISTICS PROJECT N.2 DIMENSION REDUCTION OF SYNTACTIC PARAMETERS

CS101/MA191 CLASS WINTER 2015, MATILDE MARCOLLI

1. Syntactic parameters and Dependence Problem

One of the main problems in the Principles and Parameters model of linguistics is the existence of dependencies between the syntactic parameters and the difficulty of identifying a set of independent variables. From a computational perspective, this can be seen as an instance of a more general kind of dimensionality reduction problem for a set of data.

2. Dimension Reduction Techniques

Principal Component Analysis is a widely used method for Dimension Reduction, [4]. There are many available implementations of Principal Component Analysis, such as the MLPACK machine learning package for C++, [1], see also the Principal Component Analysis and Self-Organizing Maps applet [6]. Methods for multidimensional data visualization and nonlinear dimension reduction [5] have also been developed in [3], and implemented in the ViDaExpert package [8]. The paper [2] presents a survey of several different methods of dimension reduction in data analysis.

3. Plan of the project

Using various possible dimensional reduction techniques and packages, analyze the syntactic parameter data across language families and within some individual families, and estimate the dimensionality (independent parameters) and possible manifold shapes of the parameter data. A database of syntactic parameters of the world languages is available at [7] .

References

[1] R. Curtin et al. *MLPACK: A Scalable C++ Machine Learning Library*, Journal of Machine Learning Research, Vol. 14 (2013) 801–805.

[2] M.K. Fodor, *A survey of dimension reduction techniques*, `https://computation.llnl.gov/casc/sapphire/pubs/148494.pdf`

[3] A.N. Gorban, A. Pitenko A.Zinovyev *ViDaExpert: user-friendly tool for nonlinear visualization and analysis of multidimensional vectorial data*, ArXiv:1406.5550.

[4] I.T. Jolliffe, *Principal Component Analysis*, Springer Verlag, 1986.

[5] J.A. Lee, M. Verleysen, *Nonlinear Dimensionality Reduction*, Springer, 2007.

[6] E.M. Mirkes, Principal Component Analysis and Self-Organizing Maps: applet. University of Leicester, 2011 `http://www.math.le.ac.uk/people/ag153/homepage/PCA_SOM/PCA_SOM.html`

[7] SSWL Database of Syntactic Parameters: `http://sswl.railsplayground.net/`

[8] ViDaExpert: multidimensional vectorial data visualization `http://bioinfo-out.curie.fr/projects/vidaexpert/`