Epitope Prediction Based on Peptide Array Data

Mitja Luštrek[1], Peter Lorenz[2], Felix Steinbeck[2],
Georg Füllen[1], Hans-Jürgen Thiesen[2]

*[1]Institute for Biostatistics and Informatics in Medicine and Ageing Research, [2]Institute of Immunology,
University of Rostock*
mitja.lustrek@uni-rostock.de

Background. Determining antibody binding sites (epitopes) in antigen sequences is important for vaccine design and diagnostics. In-silico epitope prediction is a cost-effective way of doing this. Epitopes may consist of multiple discontinuous amino-acid sequences or of one continuous sequence. We used machine learning methods to predict continuous epitopes.

Data and methods. We incubated 15-mers placed on peptide arrays with IVIg, commercially available mixture of antibodies from thousands of healthy donors [Lo09]. A dataset of 6,841 peptides that bind antibodies and 20,437 that do not was obtained. It was split in half for training and testing.

Eight attribute sets were designed to represent the data, among them the counts of each amino acid in a peptide, the counts of classes of amino acids (e.g., aromatic), the counts of pairs of amino acids at a certain distance, and physico-chemical properties of amino acids.

Almost 50 machine learning algorithms were evaluated on the training set. The peptides were represented by all eight attribute sets and the best machine learning algorithm was selected to train a classifier on each attribute set. Stacking was used to train the final classifier that used the outputs of the eight classifiers as attributes.

Results. The test set was classified with the stacked classifier. Its accuracy was 83.7% and the area under ROC (AUC) 0.883. For comparison, a state of the art published method [EDH08] achieved the accuracy 82.0% and AUC 0.868 when trained and tested on our dataset.

References

[EDH08] Yasser EL-Manzalawy, Drena Dobbs and Vasant Honavar. Predicting linear B-cell epitopes using string kernels. *Journal of Molecular Recognition*, 21(4): 243–255, 2008.

[Lo09] Peter Lorenz, Michael Kreutzer, Johannes Zerweck, Mike Schutkowski and Hans-Jürgen Thiesen. Probing the epitope signatures of IgG antibodies in human serum from patients with autoimmune disease. *Methods in Molecular Biology*, 524(2): 247–258, 2009.