

Matching and alignment in protein bioinformatics, using Bayesian hierarchical models

Peter J. Green*¹ and Kanti.V. Mardia ²

¹ School of Mathematics, University of Bristol

² Department of Statistics, University of Leeds

Among the challenges for statistics posed by proteomics are various alignment and matching problems. Here we consider matching protein gels in 2 dimensions, and aligning active sites of proteins in 3 dimensions. In the latter case, we also want to use information related to the grouping of the amino acids. We introduce hierarchical models for matching configurations of points in space, where the points are either unlabelled, or have at most a partial labelling constraining the matching, and in which some points may only appear in one of the configurations. We derive procedures for simultaneous inference about the matching and the transformation, using a Bayesian approach. Our model is based on a Poisson process for hidden true point locations; this leads to considerable mathematical simplification and efficiency of implementation.

We find a novel use for classic distributions from directional statistics in a conditionally conjugate specification for the case where the geometrical transformation includes an unknown rotation. Throughout, we focus on the case of affine or rigid motion transformations. Under a broad parametric family of loss functions, an optimal Bayesian point estimate of the matching matrix can be constructed, that depends only on a single parameter of the family. Implementation and performance of these methods on the proteomic tasks is described, and we discuss some open problems and suggest directions for future work.

The full details are available on the following site:

<http://www.stats.bris.ac.uk/~peter/Research.html/#Complex>