Scoring Psychological Questionnaires using Geometric Harmonics.

Liberty.E[♮], Almagor.M[†], Keller.Y[‡], Coifman.R.R.[§], Zucker.S.W.[♯]

[♮] Program of Computer Science, Yale University, New Haven, CT 06520, Email: edo.liberty at yale.edu
[†] Department of Psychology, University of Haifa, Mount Carmel Haifa, Israel, Email: malmagor at psy.haifa.ac.il
[‡] Electrical & Computer Engineering Department, Ben-Gurion University, Israel, Email: yosi.keller at yale.edu.
[§] Program of Applied Mathematics, Yale University, New Haven, CT 06520, Email: coifman-ronald at yale.edu
[♯] Program of Applied Mathematics, Yale University, New Haven, CT 06520, Email: steven.zucker at yale.edu

Abstract

Psychological and personality tests provide information relevant to psychiatric diagnosis, job placement, online dating and other social functions. Information is extracted from tests as scoring functions, or weighted sums of responses to specific questions (or pairs of them). Although such scoring functions have been manually refined for years, they remain sensitive to missing responses and do not exploit redundancies among the questions. Attempts to automate scoring function construction and refinement are only partially successful and can be biased by the lack of a uniform or normative sample for the base population [4, 1, 5, 6].

We address these problems by a kernel extension technique that models the redundancy among questions by assuming responses lie on or near a low-dimensional manifold. Scores are then a function on this manifold. Our process does not use external knowledge and does not require a uniform or normative sample. We also show that the scoring function achieved is highly robust to missing answers.

In particular, respondees' tests are viewed as a vectors in an ambient space (dimension = number of questions) sampled from the manifold according to an unknown probability density. Diagonalizing the diffusion kernel on a learning set enables us to approximate the eigenfunctions of the Laplace Beltrami operator on the underlying manifold. The diffusion normalization of the gaussian kernel overcomes possible nonuniformities by which the learning set is distributed on the manifold. Conditions relating the density and the shape of the manifold do exist, so that integrals on the manifold can be approximated with summations [2, 3].

Since the psychological diagnoses (scores) for subjects in the learning set are known and their positions on the manifold are given by the above construction, we obtain the map from position to diagnosis. The scoring function is thus approximated as a linear combination of Laplace Beltrami operator eigenfunctions, yielding an adaptive basis for scoring functions that is not biased by the original (non-uniform) sampling.

New and incomplete cases can then be scored. Using the Nyström extension we extend the kernel eigenfunctions' domain from the training set to the entire space. The extended eigenfunctions are termed *geometric harmonics*. Using our evaluation of the geometric harmonics at a new location and our representation of the scoring function we evaluate the score for the new case. Since the Nyström extension relies only on distance measures a significant stability for missing answers emerges.

We demonstrate our analysis on data from the widely used Minnesota Multiphasic Personality Inventory (MMPI-2); see figure below. The *depression score* is shown color coded on the data points distributed over the manifold. Note how we are able to diagnose *depression* with all the relevant questions (left); with some of the relevant questions (middle) and with none of the relevant questions (right). This shows the appropriateness of geometric harmonics for this problem; it suggests the existence of a low-dimensional manifold for MMPI-2 data; and confirms the non-linearity of MMPI-2 response data set and scales. To our knowledge no other techniques can match the results below.

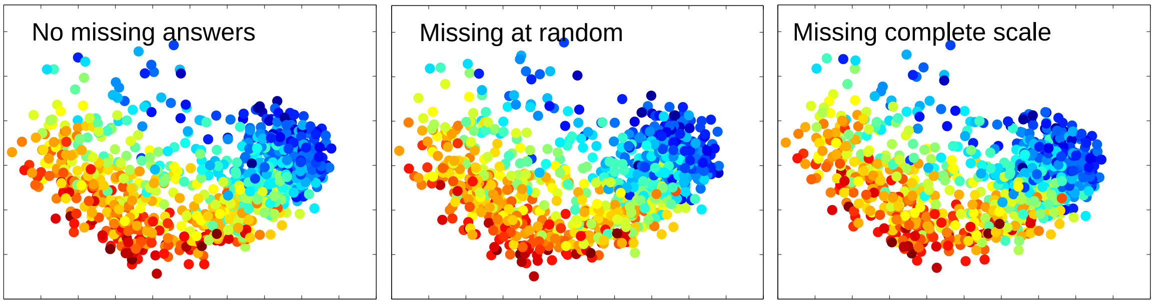

Figure 1: Illustration of MMPI-2 data scoring. The system was trained on 500 data points, and the three images above show the mapping for 1000 new subjects into the first two nontrivial geometric harmonic coordinates. Each dot represents one person and the dot colors represent the level of depression (one often used test score) diagnosed by the test. (**left**) Notice that the scoring function for depression is smooth on the underlying manifold (the colors vary slowly). (**middle**) Calculation of the depression score using only half the responses (randomly) for each person. (**right**) Calculation of the depression score using NONE of the questions that professionals agree are relevant for this diagnosis. It follows that there is substantial redundancy among the questions for this score.

Keywords

Diffusion, Kernel, Geometric Harmonics, Normative sample, Extension, Personality Assessment, MMPI-2, Personality Testing.

References

[1] I. Borg and P. Groenen. *Modern multidimensional scaling*. Springer-Verlag, 1997.

[2] Ronald Raphael Coifman, Stephane Lafon, Ann Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data. part i: Diffusion maps. *Proc. of National Academy of Sciences*, (102):7426–7431, May 2005.

[3] Ronald Raphael Coifman, Stephane Lafon, Ann Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data. part ii: Multiscale methods. *Proc. of National Academy of Sciences*, (102):7432–7438, May 2005.

[4] A. Gifi. *Nonlinear Multivariate Analysis*. John Wiley & Sons, New York, 1990.

[5] R. B. Rubin. *Multiple Imputation for nonresponse in surveys*. Wiley, 1987.

[6] J. Schafer. *Analysis of incomplete multivariate data*. Chapman and Hall, 1997.