A Nonlinear Filter for Markov Chains and its Effect on Diffusion Maps

Stefan Steinerberger
Yale University, stefan.steinerberger@yale.edu

Follow this and additional works at: http://elischolar.library.yale.edu/dayofdata

Part of the Applied Mathematics Commons, Categorical Data Analysis Commons, Discrete Mathematics and Combinatorics Commons, and the Other Statistics and Probability Commons

http://elischolar.library.yale.edu/dayofdata/2015/Posters/5

This Event is brought to you for free and open access by EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Yale Day of Data by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
We are interested in the interplay between Markov chains on a high-dimensional data set \(\{x_i\}_{i=1}^n \subset \mathbb{R}^d\) and the inner workings of spectral methods. There are many different methods, see e.g. the work of Belkin & Niyogi, Coifman & Lafon, Coifman & Maggioni, Donoho & G"{u}nther. Usually, these techniques proceed by imposing a Markov chain on the data set and analyzing diffusion on the arising graph. A popular and natural choice for the Markov chain is to declare that the probability \(p_{ij}\) to move from point \(x_j\) to \(x_i\) is

\[
p_{ij} = \frac{\exp(-\frac{1}{\varepsilon} \|x_i - x_j\|_2^2)}{\sum_{i=1}^n \exp(-\frac{1}{\varepsilon} \|x_i - x_j\|_2^2)},
\]

where the value of \(\varepsilon\) needs to be chosen depending on the given data as it induces a natural length scale \(\sim \sqrt{\varepsilon}\) which should match the distance between neighboring points.

This can work very well, see, for example, the wine data set mapped into two dimensions using a diffusion map (colors were added afterwards).

The standard diffusion paradigm proceeds by defining a random walk. However, given this local structure one would certainly believe that \(x_1, x_2, x_3, x_4\) are well connected while \(x_3\) seems to be an outlier. Consider a random walk starting in \(x_1\). A simple computation yields

- probability of being in \(x_1\) \(1\)
- probability of being in \(x_2\) \(1/4\)
- probability of being in \(x_3\) \(1/4\)
- probability of being in \(x_4\) \(1/4\)

The sources automatically detect unlikely outliers. Formally, assume we are given \(\{x_i\}_{i=1}^n \subset \mathbb{R}^d\) and an associated Markov chain described by the matrix \(P = (p_{ij})_{i,j=1}^n\). We propose using another matrix \(Q\) instead: we obtain the matrix \(P^*\) by setting \(p_{ii} = 0\) and rescaling every column of \(P\) so that to we are once again given a transition matrix. \(Q\) is then given by

\[
Q_{ij} = \min\{P_{ij}, \frac{1}{n} \sum_{k=1}^n (P_{ik} P_{kj}) \}, \quad (i,j = 1,2,\ldots,n).
\]

We may then proceed with an analysis of the data set using \(Q\) instead of \(P\). It seems that \(k = 2\) is most effective in practice but there are certainly cases where a larger \(k\) may prove advantageous.

The method can be easily adapt to detect and correct for error. In the following example, we have taken a set of points in the shape of a perfect circle and added random noise in the affinity matrix.

Theoretical results. Let \(G = (V,E)\) be a finite graph with the property that every vertex has at most \(\epsilon N\) at distance at most \(\delta\) with transition matrix \(P\). Construct \(G_\epsilon\) by adding every possible edge with probability \(0<p<1\) and let \(Q\) be the affinity assigned by the filter applied to the random walk on \(G_\epsilon\).

Theorem. The number \(X\) of vertices \((x,y)\in V \times V\) that are incorrectly thought of as present by the filter

\[
(P)_{xy} = 0 \quad \text{and} \quad Q_{xy} > 0
\]

satisfies

\[
\mathbb{E}X \leq c n + c n^3 p^2 + \frac{1}{2 np}.
\]

This implies that the filter can successfully detect \(|V|\) fake edges while only making \(O(1)\) mistakes on average. The trickiest part of the (not very complicated) proof uses the reproducing property of the binomial distribution

\[
\mathbb{P}(X(n,p,q) = \mathbb{E}(X))
\]

and combines it with a classical theorem of Wald.

References