Semi-supervised eigenvectors for large-scale locally-biased learning

In many applications, one has side information, e.g., labels that are provided in a semi-supervised manner, about a specific target region of a large data set, and one wants to perform machine learning and data analysis tasks nearby that prespecified target region. For example, one might be interested in the clustering structure of a data graph near a prespecified seed set of nodes, or one might be interested in finding partitions in an image that are near a prespecified ground truth set of pixels. Locally-biased problems of this sort are particularly challenging for popular eigenvector-based machine learning and data analysis tools. At root, the reason is that eigenvectors are inherently global quantities, thus limiting the applicability of eigenvector-based methods in situations where one is interested in very local properties of the data.

In this paper, we address this issue by providing a methodology to construct semi-supervised eigenvectors of a graph Laplacian, and we illustrate how these locally-biased eigenvectors can be used to perform locally-biased machine learning. These semi-supervised eigenvectors capture successively-orthogonalized directions of maximum variance, conditioned on being well-correlated with an input seed set of nodes that is assumed to be provided in a semi-supervised manner. We show that these semi-supervised eigenvectors can be computed quickly as the solution to a system of linear equations; and we also describe several variants of our basic method that have improved scaling properties. We provide several empirical examples demonstrating how these semi-supervised eigenvectors can be used to perform locally-biased learning; and we discuss the relationship between our results and recent machine learning algorithms that use global eigenvectors of the graph Laplacian.

General information
State: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems, University of California
Contributors: Hansen, T. J., Mahoney, M. W.
Pages: 3691-3734
Publication date: 2014
Peer-reviewed: Yes

Publication information
Journal: Journal of Machine Learning Research
Volume: 15
Issue number: 1
ISSN (Print): 1533-7928
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 4.56 SJR 1.271 SNIP 2.905
Web of Science (2017): Impact factor 2.281
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 3.09 SJR 1.29 SNIP 2.143
Web of Science (2016): Impact factor 5
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 2.64 SJR 1.431 SNIP 2.032
Web of Science (2015): Impact factor 2.45
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 3.41 SJR 1.156 SNIP 2.224
Web of Science (2014): Impact factor 2.473
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 3.5 SJR 1.235 SNIP 2.309
Web of Science (2013): Impact factor 2.853
ISI indexed (2013): ISI indexed no
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): CiteScore 3.31 SJR 1.055 SNIP 2.635