

February 2017

Qlucore Projection Score aids better visualization of large data sets

Qlucore Projection Score indicates the usefulness of a Principal Component Analysis (PCA) representation

Historically, scientists and researchers have been faced with a problem when looking at visualizations of large amounts of data, of whether the patterns they are seeing are statistically valid, or random. Qlucore Projection Score is a unique functionality within [Qlucore](#) Omics Explorer that provides scientists and researchers with information on how accurately the visual representation is actually portraying data.

The Qlucore Projection Score technique is the brain child of [Qlucore](#) co-founder Magnus Fontes. It allows detailed comparison of representations obtained by PCA corresponding to different variable subsets, e.g., those obtained by variance filtering of a large data set. The goal of exploratory visualization is to find a representation from which interpretable and potentially interesting information can be extracted, that is, one that contains structures and patterns that are likely to be non-random. By following the evolution of the projection score in real time during variance filtering, the user can easily find the variable subset (and thus implicitly the variance cut-off) giving the most informative representation.

[Magnus Fontes](#), the co-founder of Qlucore and developer of the Projection Score concept comments:

“Qlucore is proud to be at the forefront of visualisation technology for scientific research. The Projection Score technique is one which I have been working on for a considerable time and it will be very valuable in aiding research scientists to validate their data visualization work. The technique has been welcomed by my peers and I am delighted that it is now available on a commercial basis.”

To compute the projection score for a given data set, the user must start by computing the fraction of the total variance that is captured by the first three principal components. Then, an estimate is taken for the expected value of the same entity for completely random data. The projection score is defined as the difference between the square root of the observed quantity and the square root of the expected value for random data. Hence, a large value of the projection score means that the PCA representation of the observed representation contains much more information (variance) than the corresponding representation of a random data set of the same size, which suggests that there are non-random, potentially interesting structures present in the representation.

In contrast, a projection score close to zero indicates that the representation is not more informative than one of a random data set and that there are no broad, consistent patterns to be found by the PCA.

By monitoring the evolution of the projection score during variance filtering, the optimal variable subset can be found. In Qlucore Omics Explorer the projection score is colored according to the displayed value. Red indicates a low projection score, yellow indicates a medium-high score and green corresponds to a high projection score. In practice, almost all real data sets contain some non-random structure, and therefore it is very uncommon to get a projection score close to zero. The colors, and thus the boundaries between what is considered to be a "good" or a "bad" projection score, are based on our experience from applying the projection score to many different data sets, and should be interpreted mainly as rough guidelines suggesting the quality of the representations. The projection score is a widely versatile technique that is applicable for a broad family of different statistical analyses. The statistical and technical details have been published by Magnus Fontes and Charlotte Soneson in the prestigious scientific journal BMC Bioinformatics in 2011.

END

About Magnus Fontes.

Fontes is PhD and Professor of Mathematics at Lund University and Director of International Relations at the Centre for Bioinformatics, Biostatistics and Integrative Biology at Institut Pasteur in Paris. He is a co-founder of Qlucore.

About Qlucore

Qlucore started as a collaborative research project at Lund University, Sweden, supported by researchers at the Departments of Mathematics and Clinical Genetics, in order to address the vast amount of high-dimensional data generated with microarray gene expression analysis. As a result, it was recognised that an interactive scientific software tool was needed to conceptualise the ideas evolving from the research collaboration.

The basic concept behind the software is to provide a tool that can take full advantage of the most powerful pattern recogniser that exists - the human brain. The result is a core software engine that lets the user handle and filter data and the same time instantly visualise it in 3D. This will aid the user in identifying hidden structures and patterns. Over the last four years major efforts have been made to optimise the early ideas and to develop a core software engine that is extremely fast, allowing the user to explore and analyse high-dimensional data sets with the use of a normal PC, interactively and in real time.

Qlucore was founded in early 2007 and the first product released was the "Qlucore Gene Expression Explorer 1.0". The latest version of this software, now called Qlucore Omics Explorer, represents a major step forward with advanced statistics support, streamlined workflows for multiple data types, and a wide selection of presentation methods to aid the user. The presentation methods range from an innovative use of principal component analysis (PCA) to interactive heat maps and flexible scatter

plots. All user action is at most two mouse clicks away. The company's early customers are mainly from the Life-science and Biotech industries, but solutions for other industries are currently under development.

www.qlucore.com