

Dec 2013

Qlucore announces aid to better visualisations of large data sets

New traffic-light 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 visualisations of large amounts of data, of whether the patterns they are seeing are statistically valid, or random. Qlucore Projection Score is a unique functionality that will be available in the new version of [Qlucore Omics Explorer 3.0](#). Projection Score will provide the user with information on how accurately the visual representation is actually portraying data.

The patent-pending 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 visualisation 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 visualisation 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 3.0 the projection score is coloured 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 colours, 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 Sonesson in the prestigious scientific journal BMC Bioinformatics in 2011.

END

About Magnus Fontes

Fontes is PhD and Professor of Mathematics at Lund University, Chairman of the Centre for Mathematical Sciences at Lund University, President of ECMI and Vice chairman of the Swedish National Committee for Mathematics at the Royal Swedish Academy of Science. 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 visualises the data in 3D and will aid the user in identifying hidden structures and patterns. Over the last two years the major efforts have been to optimise the early ideas and to develop a core software engine that is extremely fast, allowing the user to interactively and in real time instantly explore and analyse high-dimensional data sets with the use of a normal PC.

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, Version 1.1, represents a major step forward with the advanced statistics support. 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.

One of the key methods used by Qlucore Gene Expression Explorer to visualise data is dynamic principal component analysis (PCA), an innovative way of combining PCA analysis with immediate user interaction. Dynamic PCA is PCA analysis combined with instant user response, a combination which provides an optimal way for users to visualise and analyse a large dataset. By presenting a comprehensive view of the data set at the same time, the user is given full freedom to explore all possible versions of the presented view.

PCA analysis works by projecting high dimensional data down to lower dimensions. The specific projections of the high-dimensional data are chosen in order to maintain as much variance as possible in the projected data set. With Qlucore Gene Expression Explorer, data is projected and plotted on the two dimensional computer screen and then rotated manually or automatically and examined by the naked eye.

Additional information is available at www.qlucore.com

Press Contact:

Chaz Brooks, CBC

Tel: +44 (0)1483 537 890

Email: chazb@chazb.com

Company Contact:

Carl-Johan Ivarsson

Phone +46 46 286 3110

Email: carl-johan.ivarsson@qlucore.com

Web: <http://www.qlucore.com>