conferenceseries.com SciTechnol

J Comput Eng Inf Technol 2016, 5:4(Suppl)
http://dx.doi.org/10.4172/2324-9307.C1.006

Mining coherent patterns in multidimensional big data based on spectral decomposition

Hong Yan
City University of Hong Kong, Hong Kong

Although many multidimensional datasets can be very large, they may contain coherent patterns of much smaller sizes. For example, in gene expression data, we may be interested in a subset of genes that co-express under a subset of conditions during a subset of time intervals. It is a challenge to extract these coherent patterns from a large multidimensional data array. In this presentation, I will discuss a robust coherent pattern detection technique that our research group has developed recently based on spectral decomposition. In our method, we analyzed the data in singular vector spaces and detect hyperplanes that are related to subsets of data features or samples, which correspond to coherent data points. This procedure effectively suppressed the noise and removed irrelevant elements in the data. We have found several useful applications of the coherent pattern extraction algorithm to image and biomedical data analysis. Based on these patterns, we were able to recognize human facial expressions from face images and locate the points that play important roles in characterizing different expressions. With DNA microarray gene expression data, we can identify co-expressed genes for different types and subtypes of cancers. We have also used the coherent patterns to analyze the molecular mechanisms of non-small cell lung cancer drug resistance and predict the drug resistance levels for different protein mutations.

h.yan@cityu.edu.hk

Text mining as a new discipline in searching and data mining

Tanzila Saba
Prince Sultan University, Saudi Arabia

Indeed, today, a lot of available information and e-business data are captured in text files that are normally unstructured such as customer's feedback on product/service using different social medias, bibliographic databases including authors details/customers detail, journal articles and so forth. This information is multiplying day by day and mostly is in unstructured form. Consequently, such information cannot be used positively unless and until are converted into structured format. Currently researchers focused on the knowledge discovery from huge databases, warehouses to transform unstructured text in to meaningful information. The discovery of knowledge from such database sources containing free text is called 'text mining' rather more specific Text Data Mining. Text mining is either the discovery of texts or the exploration of texts in search of valuable, yet hidden information. Formally, text mining is a set of linguistic, statistical, and machine learning techniques that model and structure the information content of textual sources for business intelligence, exploratory data analysis, research or investigation. Text mining applications are: Document visualization- it refers to documents clustering based on keywords such as research publications on similar issues and text analysis and understanding- it refers to natural language processing techniques such as text categorization, information extraction and summarization. Text mining challenges includes: 1) Forms semantic analysis- documents semantic analysis is relatively a new trend in text data mining. To come out with refine knowledge, it is mandatory to apply semantic analysis techniques to derive a sufficiently rich representation to capture the relationship between the objects or concepts reflected in text documents; 2) multilingual text refining- multilingual data mining is basically language independent. Hence, to cluster information being extracted from multilingual sources is a fresh research area with a lot of scope. Therefore, text refining algorithms development is the need of the day, that could process multilingual text documents and produce language-independent intermediate forms; and Personalized autonomous mining- current text mining tools are not simple and need training , expertise. Future text mining tools, as part of the knowledge management systems, need simple personalization features for end users without having enough technical skills such as business executives.

tanzilasaba@yahoo.com