

به کارگیری تکنیک های داده کاوی در پزشکی

جمال شهرابی^۱، محمد سعید زائری^۲، محمود پری آذر^۳

چکیده:

پیشرفت های بشر در چند دهه اخیر در زمینه جمع آوری و ذخیره نتایج و داده ها باعث شده است که ابعاد پایگاه داده ها به طور چشم گیری افزایش پیدا کند. صنعت پزشکی جزء معدود صناعی است که با پایگاه داده های بسیار بزرگ همراه با داده های زیاد مواجه است. نکته مهم در این پایگاه داده های بزرگ اطلاعات و دانشی است که از آن استخراج می شود و نیاز به استفاده از روش های هوشمند و ساختارمند وجود دارد. داده کاوی یک رشته علمی جدید در زمینه بازیابی اطلاعات از پایگاه داده ها می باشد. تکنیک های مختلفی به لحاظ خطی و غیر خطی برای کاهش ابعاد و رسیدن به ویژگی های مفید از یک پایگاه داده ها وجود دارد.

در این مقاله سه الگوریتم 'Random Projection، Non-negative Matrix Factorization و Principal Component analysis جهت کاهش ابعاد پایگاه داده ها بر روی سه مجموعه داده با ساختار خطی، غیر خطی و گسسته متنی پیاده سازی شده است. در انتها مشخص می شود که الگوریتم Principal Component Analysis با توجه به سه ماهیت مختلف مجموعه داده ها بهترین نتیجه را نسبت به دو الگوریتم دیگر در کاهش ابعاد و نمایش ویژگی های داده ها ارائه می کند.

کلمات کلیدی:

داده کاوی، استخراج دانش از پایگاه داده ها، پایگاه داده های پزشکی

Using Data Mining Techniques in Medicine

Jamal shahrabi¹, mohammad saeed zaeri², mahmood pariazar³
Amirkabir University of Technology, Industrial Engineering Department

Abstract

Break through of human in recent decades in gathering and storing results and data have culminated in dramatical increase in dimensions of data bases. Medicine industry is amongst those small industries that are facing with very great data bases along with great data. Great point in these big data bases is information and data that would be obtained from it and necessary for utilizing clever and structural methods. Data mining is a new science in data retrieving from data bases. Different techniques, linear or non-linear, exist for reduction of dimensions and getting to favorable characteristics of data bases.

In this article three algorithms random projection, Non - negative Matrix Factorization and Principal component analysis for reduction of dimension of data bases on three complexes of data with linear, non-linear and text structures. In the end, it's knows that Principal Component analysis with regard to three different entities of data, provides the best result in comparison to two other algorithms in dimensions reduction and displaying characteristics of data.

Keywords

Data Mining, Knowledge Discovery in Database, Medicine Database

Shahrabi@dal.ca

M_pariazar@aut.ac.ir و Zaeri@aut.ac.ir

۱- استادیار دانشکده مهندسی صنایع دانشگاه صنعتی امیرکبیر

۲و۳- دانشجوی کارشناسی ارشد مهندسی صنایع دانشگاه امیرکبیر