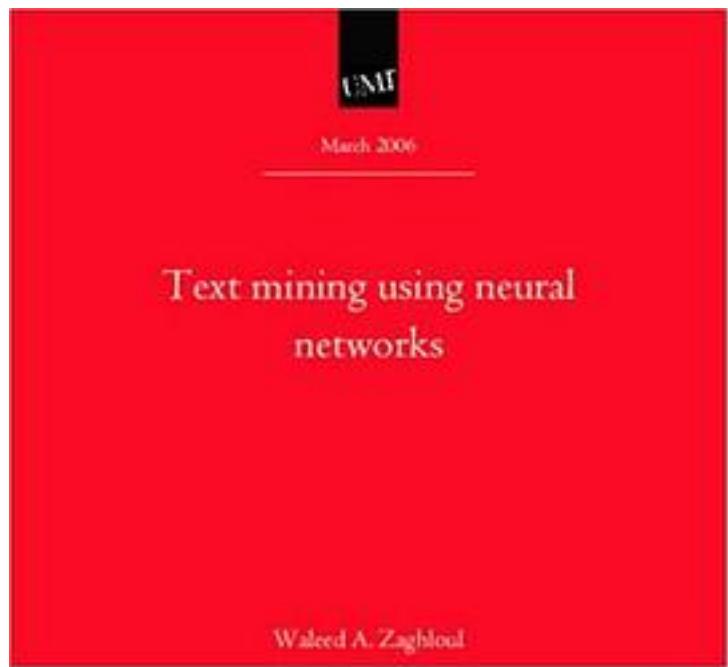# Text mining using neural networks

[Text mining using neural networks_下载链接1_](#)

著者:Waleed A. Zaghloul

出版者:ProQuest / UMI

出版时间:2006-03-18

装帧:Paperback

isbn:9780542321191

The recent advances in information and communication technologies (ICT) have resulted in unprecedented growth in available data and information. Consequently,

intelligent knowledge creation methods are needed. Organizations need efficient intelligent text mining methods for classification, categorization and summarization of information available at their disposal. Neural Networks have successfully been used in a wide variety of classification problems. The purpose of this dissertation is two-fold. First, applying neural networks in text mining. Second, dramatically reducing the document size by using only the summary (abstract) instead of the whole document without affecting performance. To achieve these goals several research questions had to be answered. For example, how can a document be presented in a format suitable to neural networks? Also, how and how much can a document be reduced in size without losing any valuable content? To answer the research questions posed in this study, 729 research papers were collected as data for the study. Those papers were published in MISQ in the period 1977-2004. Only the abstracts of those papers were used to reduce the document size. Those abstracts were further prepared to be used with neural networks. After identifying the most popular 100 terms in the overall population of documents, each document was represented as 100 numbers. The numbers represent the frequency with which the top 100 terms appear within the given document. A neural network processes those numbers and then classifies the document as belonging or not belonging to a certain category. The classification categories used are the MISQ predefined research categories. A separate neural network was used for each category with a total of nine. This specialization improves performance. Each neural network was trained 50 times and their performance averaged out to counter any inherent randomness in their performance. The results obtained are promising with several factors affecting performance being identified. If such factors are controlled it is possible to very efficiently train neural networks to classify documents using only a summary or an abstract. This results in great savings in computing time and cost. This method could easily be adapted to any other population of documents.

作者介绍:

目录:

Text mining using neural networks_下载链接1_

标签

评论

[Text mining using neural networks_下载链接1_](#)

书评

```
------------------------------
```