



استخراج کلمات کلیدی یک متن با استفاده از شبکه های عصبی

عباس احمدی^۱، طیبه حسینی خواه^۲

^۱استاد یار و عضو هیات علمی دانشکده مهندسی صنایع، دانشگاه صنعتی امیر کبیر؛ abbas.ahmadi@aut.ac.ir

^۲دانشجوی کارشناسی ارشد مهندسی صنایع، دانشگاه صنعتی امیر کبیر؛ hoseinikhah@aut.ac.ir

چکیده:

رشد فزاینده اسناد متنی بر روی وب ضرورت استفاده از ابزارهای پردازش خودکار متون را بیش از پیش آشکار می سازد. طبیعت غیر ساختیافته متون، استفاده از روشهای سنتی استخراج اطلاعات از پایگاه داده های رابطه ای را غیر ممکن می سازد. بنابراین روش ها و الگوریتم های پردازش (و یا پیش پردازش) خاصی برای استخراج الگوهای مفید مورد نیاز است یکی از این روشها استفاده از شبکه های عصبی برای استخراج دانش مورد نیاز از متون است در این پژوهش از شبکه رقابتی LVQ و شبکه عصبی MLP برای استخراج کلمات کلیدی متون داده شده استفاده شده است این الگوریتم ها در واقع کلمات موجود در یک متن را به دو خوشه کلمات کلیدی و غیر کلیدی دسته بندی می کنند. برای این منظور از ۸۰ متن جهت آموزش شبکه استفاده شده است نتایج حاکی از آن است که شبکه عصبی پرسپترون چندلایه (MLP) در معیارهای ارزیابی تعریف شده دارای عملکرد بهتری نسبت به شبکه رقابتی LVQ می باشد.

کلمات کلیدی

استخراج کلمات کلیدی، شبکه عصبی پرسپترون چندلایه، شبکه رقابتی بردار یادگیری.

Keyword Extraction from a text using Neural Network

Abbas Ahmadi, Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran
Tayebe Hoseinikhah, Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran

ABSTRACT

The dramatic increase on the text based documents on web reveals the necessity of using automatic text processing tools. Due to the nature of unstructured texts, using traditional methods such as retrieving information from relational databases is impossible, so it's necessary to use specific processing (or preprocessing) algorithms to extract information. One of these methods is using Neural Network algorithms to extract knowledge from text. We use a competitive network like LVQ and a MLP neural network to extract keywords; these algorithms classify text's words into two clusters: keyword and non-keyword. We use 80 documents as our training set and the results show that Multi-Layer Perceptron has better functionality in comparison with Learning Vector Quantization.

KEYWORDS

Keyword Extraction, Neural Network, Multi-Layer Perceptron, Learning Vector Quantization