Cumulative Index

Ab initio band structure of quasi-metallic carbon nanotubes for terahertz applications

P N D'yachkov, I A Bochkov

Two integers (n_1, n_2) determine the band structure of single-walled carbon nanotubes (CNTs). According to π-electron zone-folding model, an energy gap between the valence and conduction bands disappears if a difference $n_1 - n_2 = 3q$ is divisible by three. Such CNTs are called the quasi-metallic tubules. An account of surface curvature of tubules predicts that a small gap opens in such CNTs and they are the narrow-gap semiconductors really. Available experimental and theoretical information on the gap energies is very limited. In this paper, the band structures of the 50 CNTs (n_1, n_2) with $4 \leq n_1 \leq 18$ and $n_2 = n_1 - 3q$ are calculated using a linearized augmented cylindrical waves method. The quasi-metallic CNTs with optical gaps falling within the terahertz range (1 - 40 meV) are identified, which can be used to design the high-frequency devices like the terahertz emitters, detectors, multipliers, antennas, polarizers, and transistors.

K-Medoids algorithm used for english sentiment classification in a distributed system

Vo Ngoc Phu, Vo Thi Ngoc Tran

Sentiment classification is significant in everyday life, such as in political activities, commodity production, and commercial activities. Finding a fast, highly accurate solution to classify emotion has been a challenge for scientists. In this research, we have proposed a new model for Big Data sentiment classification in the parallel network environment – a Cloudera system with Hadoop Map (M) and Hadoop Reduce (R). Our new model has used a K-Medoids Algorithm (PAM) with multi-dimensional vector and 2,000,000 English documents of our English training data set for English document-level sentiment classification. Our new model can classify sentiment of millions of English documents based on many English documents in the parallel network environment. However, we tested our new model on our testing data set (including 1,000,000 English reviews, 500,000 positive and 500,000 negative) and achieved 85.98% accuracy.

Image steganography algorithm based on edge region detection and hybrid coding

Kumar Gaurav, Umesh Ghanekar

In this paper, a novel steganography algorithm based on local reference edge detection technique and exclusive disjunction (XOR) property is proposed. Human eyes are less sensitive towards intensity changes in the sharp edge region compared to the uniform region of the image. Because of this, the secret message bits have been embedded in the sharp regions by local reference pixels that are located in the edge blocks. The predefined sets of pixels are easily identified with less computational complexity in the stego image. The embedding algorithm improved in terms of security and capacity using bit plane dependent XOR

coding technique that makes least possible alterations in LSB bits of edge pixels. The existing edge-based steganography techniques provide better imperceptibility but relatively limits the embedding capacity. The proposed method efficiently improves the embedding capacity with an acceptable range of imperceptibility and robustness. The simulation results evaluated using full reference image quality assessment method, it exhibits better embedding capacity (bpp) compared to existing steganography techniques retaining the values of PSNR and structural similarity (SSIM).

Key words: Steganography, Edge detection, Colour image, XOR

English sentiment classification using a Fager & MacGowan coefficient and a genetic algorithm with a rank selection in a parallel network environment

Vo Ngoc Phu, Vo Thi Ngoc Tran

Pages 57-112

We have already studied a data mining field and a natural language processing field for many years. There are many significant relationships between the data mining and the natural language processing. Sentiment classification has had many crucial contributions to many different fields in everyday life, such as in political Activities, commodity production, and commercial Activities. A new model using a Fager & MacGowan Coefficient (FMC) and a Genetic Algorithm (GA) with a fitness function (FF) which is a Rank Selection (RS) has been proposed for the sentiment classification. This can be applied to a big data. The GA can process many bit arrays. Thus, it saves a lot of storage spaces. We do not need lots of storage spaces to store a big data. Firstly, we create many sentiment lexicons of our basis English sentiment dictionary (bESD) by using the FMC through a Google search engine with AND operator and OR operator. Next, According to the sentiment lexicons of the bESD, we encode 7,000,000 sentences of our training data set including the 3,500,000 negative and the 3,500,000 positive in English successfully into the bit arrays in a small storage space. We also encrypt all sentences of 7,500,000 documents of our testing data set comprising the 3,750,000 positive and the 3,750,000 negative in English successfully into the bit arrays in the small storage space. We use the GA with the RS to cluster one bit array (corresponding to one sentence) of one document of the testing data set into either the bit arrays of the negative sentences or the bit arrays of the positive sentences of the training data set. The sentiment classification of one document is based on the results of the sentiment classification of the sentences of this document of the testing data set. We tested the proposed model in both a sequential environment and a distributed network system. We achieved 88.21% accuracy of the testing data set. The execution time of the model in the parallel network environment is faster than the execution time of the model in the sequential system. The results of this work can be widely used in applications and research of the English sentiment classification.

Key words: computer and information technologies, natural and engineering sciences, operation research and decision making, mathematical and computer modelling

English sentiment classification using a BIRCH algorithm and the sentiment lexicons-based one-dimensional vectors in a parallel network environment

Vo Ngoc Phu, Vo Thi Ngoc Tran

Pages 113-156

Sentiment classification is significant in everyday life, such as in political activities, commodity production, and commercial activities. In this survey, we have proposed a new model for Big Data sentiment classification. We use a Balanced Inerative Reducing and Clustering using Hierarchies algorithm (BIRCH) and many one-dimensional vectors based on many sentiment lexicons of our basis English sentiment dictionary (bESD) to cluster one document of our testing data set based on our training data set in English. We calculate the sentiment scores of English terms (verbs, nouns, adjectives, adverbs, etc.) by using a BARONI-URBANI &

BUSER-II coefficient (BUBC) through a Google search engine with AND operator and OR operator. We do not use any multi-dimensional vector. We also do not use any one-dimensional vector based on a vector space modelling (VSM). We do not use any similarity coefficient of a data mining field. The BIRCH is used in clustering one sentence of one document of the testing data set into either the positive or the negative of the training data set. We tested the proposed model in both a sequential environment and a distributed network system. We achieved 87.76% accuracy of the testing data set. The execution time of the model in the parallel network environment is faster than the execution time of the model in the sequential system. The results of this work can be widely used in applications and research of the English sentiment classification.

Key words: computer and information technologies, natural and engineering sciences, operation research and decision making, mathematical and computer modelling