

Tuyển tập Tóm tắt báo

A new self-adaptive CQ algorithm with an application to the LASSO problem	2
<i>Pham Ky Anh, Nguyen The Vinh, Vu Tien Dung</i>	
A Brief Introduction to Independent Set Reconfiguration and Related Problems	3
<i>Duc A. Hoang</i>	
A New Object Description based on Sparsity over Learning Dictionary	4
<i>Do Thanh Ha</i>	
Using the bilingual phrase table in training to improve the quality of the Vietnamese - English neural machine translation system	5
<i>Nguyen Tien Ha</i>	
An approach based on deep learning solves the problem of network intrusion detection	6
<i>Tran Thi Huong</i>	
A study on Vietnamese medical text documents for UMLS	6
<i>Nguyen Viet Hung</i>	
Identify the interpersonal relationships based on linguistic features in conversation	7
<i>Nguyen Thi Minh Huyen</i>	
Multifactorial optimization	8
<i>Nguyen Duong Kien</i>	
Sử dụng phân bố từ trong phân tích cú pháp thành phần tiếng Việt.....	8
<i>Nguyễn Thị Lương</i>	
A Comparative Study of Neural Network Models for Sentence Classification	9
<i>Le Hong Phuong</i>	
Vietnamese social text processing: Part Of Speech Tagging	10
<i>Ngo The Quyen</i>	
Neural-Fuzzy with Representative Sets for Prediction of Student Performance	11
<i>Le Hoang Son</i>	
Lifetime optimization for wireless underground sensor networks.....	12
<i>Nguyen Thi Tam</i>	
Online picture fuzzy clustering: a new approach for real-time fuzzy clustering on picture fuzzy sets.....	13
<i>Pham Huy Thong</i>	
DC Algorithm for Image Denosing based on Dictionary Learning	14
<i>Nguyen Thi Bich Thuy</i>	
Clustering Functional Data using Projection	14
<i>Pham Huy Tung</i>	
Pseudo-probabilistic encryption algorithms	15
<i>Nguyen Hai Vinh</i>	

Thuật toán CQ tự thích nghi và ứng dụng giải bài toán Lasso

Pham Ky Anh, Nguyen The Vinh, Vu Tien Dung
Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Trong báo cáo này, chúng tôi trình bày thuật toán CQ tự thích nghi mới giải bài toán chấp nhận tách trong không gian Hilbert và chứng minh sự hội tụ của thuật toán dưới một số điều kiện thích hợp. Các kết quả thử nghiệm số thu được khi áp dụng thuật toán giải bài toán Lasso một bài toán thường xuất hiện trong lĩnh vực xử lý tín hiệu thưa đã chứng tỏ tính hiệu quả của thuật toán mới đề xuất.

A new self-adaptive CQ algorithm with an application to the LASSO problem

Pham Ky Anh, Nguyen The Vinh, Vu Tien Dung
Faculty of Mathematics, Mechanics and Informatics
VNU University of Science

Abstract: In this report, we introduce a new self-adaptive CQ algorithm for solving split feasibility problems in real Hilbert spaces. The algorithm is designed such that the stepsizes are directly computed at each iteration. We also consider the corresponding relaxed CQ algorithm for the proposed method. Under certain mild conditions, we establish weak convergence of the proposed algorithm as well as strong convergence of its hybrid-type variant. Finally, numerical examples illustrating the efficiency of our algorithm in solving the LASSO problem are presented.

Giới thiệu sơ lược về bài toán cấu hình lại các tập ổn định trong và các vấn đề liên quan

Hoàng Anh Đức
Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Trong thời gian gần đây, các bài toán *cấu hình lại* (*reconfiguration problems*) đã và đang được quan tâm khá nhiều từ góc độ lý thuyết khoa học máy tính. Ý tưởng của việc nghiên cứu các bài toán cấu hình lại bắt nguồn từ việc nghiên cứu các tình huống

khi ta cần tìm một dãy các phép biến đổi từ lời giải này sang lời giải khác của một bài toán theo một quy tắc cho trước sao cho tất cả các kết quả trung gian đều là một lời giải của bài toán. Một ví dụ của các bài toán dạng này là bài toán 15-puzzle—một bài toán đã được nghiên cứu từ những năm 1879. Một cấu hình của 15-puzzle là một cách đặt 15 ô được đánh số 1, 2, . . . , 15 trên một bảng cỡ 4×4 . Quy tắc cấu hình lại (reconfiguration rule) được cho trước như sau: một ô (được đánh số) được phép di chuyển đến ô trống khi nó nằm ở bên trái/phải/trên/dưới của ô trống. Cho một mô tả của các cấu hình và một mô tả của các quy tắc cấu hình lại, câu hỏi đặt ra trong bài toán 15-puzzle là liệu có tồn tại một dãy các cấu hình từ một cấu hình này sang cấu hình khác sao cho mỗi kết quả trung gian nhận được từ việc áp dụng quy tắc cấu hình lại cho trước cũng là một cấu hình của bài toán. Một trong số các bài toán NP-complete quan trọng là bài toán Independent Set. Trong nhiều năm qua, vô số các bài toán liên quan đến Independent Set đã được nghiên cứu. Trong báo cáo này, chúng tôi giới thiệu khái quát về bài toán cấu hình lại các tập ổn định trong (Independent Set Reconfiguration) và các bài toán liên quan.

A Brief Introduction to Independent Set Reconfiguration and Related Problems

Duc A. Hoang

*Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: Recently, the *reconfiguration problems* attracted the attention from the viewpoint of theoretical computer science. The problem arises when we like to find a step-by-step transformation between two feasible solutions of a problem such that all intermediate results are also feasible and each step abides by a fixed reconfiguration rule, that is, an adjacency relation defined on feasible solutions of the original problem. A classic example of such problems is the so-called 15-puzzle—a research topic since 1879. A configuration of 15-puzzle consists of a placement of 15 tiles numbered 1, 2, . . . , 15 on a 4×4 board, leaving one empty square. The allowed (reconfiguration) rule is that a tile can move to the empty square if it is on the left/right/top/bottom of that square. Given a collection of configurations and the reconfiguration rule, the 15-puzzle problem asks whether there exists a sequence of configurations that transforms one configuration into another, such that each intermediate member of the sequence is obtained from the previous one by a single tile-move. One of the important NP-complete problems is the Independent Set problem. Over the years, countless problems involving Independent Set have been studying. In this report, we give a brief introduction to Independent Set Reconfiguration and related problems in this area.

Xây dựng đặc tả cho đối tượng sử dụng biểu diễn thưa trên từ điển học

Đỗ Thanh Hà

Khoa Toán - Cơ - Tin học

Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Bài báo trình bày cách tiếp cận mới trong nhận dạng đối tượng sử dụng biểu diễn thưa. Đóng góp chính của nghiên cứu là đề xuất cách xây dựng đặc tả cho đối tượng sử dụng tính thưa của các đặc tả địa phương trên từ điển học. Các đặc tả mới này được sử dụng để nhận dạng các đối tượng khác nhau. Các kết quả thực nghiệm trên bộ dữ liệu gồm 512 ảnh dấu giầy đã chỉ ra phương pháp đề xuất cho hiệu quả tốt hơn các phương pháp trước đó.

A New Object Description based on Sparsity over Learning Dictionary

Do Thanh Hà

Faculty of Mathematics, Mechanics and Informatics

VNU University of Science

Abstract: This paper presents a new approach for object recognition based on sparse representation. The main idea is to create a object descriptor using the information of sparsity over the learning dictionary. We use these descriptors as a point of comparison between images. The experiments done on a database including 512 footprints images shows that the proposed method is competitive related to the state-of-the-art methods.

Nâng cao chất lượng hệ thống dịch máy nơ ron Việt – Anh bằng việc sử dụng bảng cụm từ song ngữ trong huấn luyện hệ thống

Nguyễn Tiến Hà

Khoa Toán - Cơ - Tin học

Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Tiếp cận dịch máy dựa vào mạng nơ ron đã cho thấy những ưu thế vượt trội so với các tiếp cận dịch máy khác. Để có một hệ thống dịch máy nơ ron tốt cần phải có một kho ngữ liệu song ngữ có chất lượng và đủ lớn dùng trong huấn luyện, đặc biệt là vấn đề cải tiến phương pháp, kỹ thuật dùng trong hệ thống dịch. Hệ thống dịch máy nơ ron hiện nay xử lý sinh câu dịch dựa trên từ mà chưa có khả năng xử lý sinh câu dịch dựa trên cụm từ để có thể thu được bản dịch tốt hơn. Trong bài báo này, chúng tôi đề xuất một kỹ thuật có thể hướng hệ thống dịch máy nơ ron xử lý sinh câu dịch dựa trên cụm từ nhằm cải tiến chất lượng hệ thống dịch máy nơ ron Việt - Anh. Thực nghiệm trên kho ngữ liệu

song ngữ Việt – Anh mà chúng tôi thu thập được cho thấy phương pháp mà chúng tôi đề xuất đã cải tiến được trên 0,4 điểm BLEU.

Using the bilingual phrase table in training to improve the quality of the Vietnamese - English neural machine translation system

Nguyen Tien Ha

*Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: The approach to machine translation depending on neural networking has got more advantages than other approaches. It is necessary to have the quality and big size bilingual corpora, especially for improvement of methods and techniques using in neural machine translation system so that it can work better. At present, the neural machine translation system generates target sentences word by word, but not phrases which can make the translation better. In this paper, we propose a new technique which creates the neural machine translation system generates target sentences in phrases tendency in order to improve the quality of the machine translation system. Experiments on Vietnamese – English bilingual corpora which we have collected show that our proposal gets 0,4 BLEU more than that in the baseline neural machine translation system.

Một cách tiếp cận dựa vào phương pháp học sâu cho bài toán phát hiện xâm nhập mạng

Trần Thị Hương

*Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG HN*

Tóm tắt: Những năm gần đây, phát hiện xâm nhập mạng đang trở thành vấn đề cấp bách trong an ninh mạng. Sự gia tăng về số lượng, độ phức tạp, tinh vi của các cuộc tấn công đang là thách thức đối với các hệ thống phát hiện xâm nhập đã có như hệ thống phát hiện dựa vào luật (*Rule-based Intrusion Detection System*), hệ thống phát hiện dựa vào thống kê, . . . Một cách tiếp cận mới dựa vào trí tuệ nhân tạo (*Artificial Intelligence*) đang thu hút sự quan tâm của các nhà khoa học bởi khả năng phát hiện các cuộc tấn công chưa biết một cách linh hoạt và hiệu quả. Trong nghiên cứu này, tác giả sẽ trình bày và xây dựng mô hình phát hiện xâm nhập dựa vào phương pháp học sâu *Selt-Taught Learning*. Đồng thời đánh giá và so sánh với các phương pháp học máy khác như neural network, support vector machine, . . . Các thực nghiệm được tiến hành trên các bộ dữ liệu chuẩn KDD CUP và bộ dữ liệu NSL-KDD.

An approach based on deep learning solves the problem of network intrusion detection

Tran Thi Huong

*Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: In recent years, Network intrusion detection has become an urgent issue in cybersecurity. The increase in the number, complexity, and sophistication of attacks is challenging the existing intrusion detection systems such as the Rule-based Intrusion Detection System, Statistics-based detection system. A new approach based on artificial intelligence (AI) is attracting the attention of scientists by the ability to detect unknown attacks in a flexible and effective way. In this study, the author will present and build an intrusion detection model based on the Self-Taught Learning approach. Simultaneous evaluation and comparison with other machine learning methods such as neural network, support vector machine, . . . Experiments were conducted on KDD CUP and NSL-KDD data sets.

Khảo sát dữ liệu văn bản y tế tiếng Việt cho UMLS

Nguyễn Việt Hùng

*Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG HN*

Tóm tắt: UMLS (Unified Medical Language System) là một hệ thống tích hợp và chuẩn hoá các nguồn tài nguyên y tế. UMLS gồm ba thành phần: siêu từ điển (metathesaurus), mạng ngữ nghĩa (semantic network), các công cụ xử lý ngôn ngữ tự nhiên cho văn bản y tế. Siêu từ điển của UMLS hiện tại được tổng hợp từ hơn 200 từ điển. Đa số các từ điển này là tiếng Anh, phần còn lại là tiếng Đức, Pháp, . . . Mục tiêu của nghiên cứu là khảo sát các nguồn dữ liệu tiếng Việt cho hệ thống UMLS.

A study on Vietnamese medical text documents for UMLS

Nguyen Viet Hung

*Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: UMLS (Unified Medical Language System) integrates medical resources and build standards for medical system. UMLS has three tools: metathesaurus, semantic network and SPECIALIST Lexicon and Lexical Tools. Metathesaurus linkings concepts and terms from over 200 different vocabularies. Almost are in English and others are in French, This study aims to survey on Vietnamese medical text documents for UMLS.

Xác định quan hệ cá nhân dựa vào đặc điểm ngôn ngữ trong hội thoại

Nguyễn Thị Minh Huyền

Khoa Toán - Cơ - Tin học

Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Một trong những bài toán thú vị trong xử lý ngôn ngữ tự nhiên được quan tâm gần đây là làm thế nào dựa trên đặc điểm ngôn ngữ của hội thoại để có thể tự động xác định quan hệ và sự tiến triển quan hệ giữa hai người tham gia hội thoại. Dữ liệu phục vụ cho nghiên cứu giải quyết vấn đề này có thể là các kho ngữ liệu hội thoại trên mạng xã hội hay các kho ngữ liệu hội thoại với khách hàng của các trung tâm chăm sóc khách hàng. Báo cáo này sẽ khảo sát các đặc trưng ngôn ngữ và các độ đo liên quan nhằm áp dụng các thuật toán học máy cho bài toán nêu trên.

Identify the interpersonal relationships based on linguistic features in conversation

Nguyen Thi Minh Huyen

Faculty of Mathematics, Mechanics and Informatics

VNU University of Science

Abstract: One of the interesting problems recently emerged in natural language processing is how to identify the interpersonal relationship and its dynamics based on linguistic features in a conversation dyad. Two data sources can be exploited for this work, one is the corpora collected from social interaction channels, the other is conversational datasets from customer-relationship services from industry. In this research, we will investigate the linguistic features and the related metrics for applying machine learning techniques to solve the above problem.

Mô hình tiến hóa đa nhân tố

Nguyễn Dương Kiên

Khoa Toán - Cơ - Tin học

Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Mô hình tiến hóa đa nhân tố là mô hình tiến hóa tổng hợp khai thác tính song song ngầm bằng việc tìm kiếm dựa trên quần thể để giải quyết nhiều vấn đề tối ưu hóa đồng thời sử dụng duy nhất một quần thể các cá thể. Mỗi bài toán tối ưu được coi như một nhân tố ảnh hưởng đến quá trình tiến hóa. Có một vài ưu điểm của đa nhiệm thông qua đa nhân tố. Thứ nhất, một nhiệm vụ tối ưu hóa đơn giản có thể nằm trong một nhiệm vụ tối ưu phức tạp hơn trong môi trường đa nhiệm. Việc thực hiện đa nhiệm thông qua đa nhân tố có thể đẩy nhanh quá trình tối ưu hóa cho các nhiệm vụ tối ưu phức tạp. Ưu điểm thứ hai là chúng ta có thể chuyển vật liệu di truyền từ các bài toán tối ưu đơn giản đến các bài toán tối ưu phức tạp.

Multifactorial optimization

Nguyen Duong Kien

*Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: Multifactorial optimization is a composite evolutionary paradigm that harnesses the implicit parallelism that of population-based search to solve multiple, diverse optimization problems concurrently using a single population of evolving individuals. Each optimization problem is viewed as an additional factor influencing the evolutionary process. Multitasking via multifactorial optimization has several advantages. Firstly, a simple optimization task may complement some other, more complex task in a multitasking environment. This lead to accelerating the optimization process for complex functions. Secondly, transfer of genetic material across outwardly unrelated complex optimization problems.

Sử dụng phân bố từ trong phân tích cú pháp thành phần tiếng Việt

Nguyễn Thị Lương

*Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG HN*

Tóm tắt: Phân tích cú pháp thành phần, tức là xác định cấu trúc các thành phần ngữ trong câu, là một bài toán có nhiều ứng dụng trong các hệ thống xử lý ngôn ngữ như các hệ thống hỏi đáp, trích chọn thông tin, tóm tắt văn bản, dịch máy... Đã có rất nhiều nghiên cứu cho bài toán phân tích cú pháp thành phần và đạt được kết quả tốt cho khá nhiều ngôn ngữ như tiếng Anh, tiếng Trung, tiếng Pháp, tiếng Ả Rập, Đối với tiếng Việt, một số nhóm nghiên cứu đã xây dựng kho ngữ liệu có gán nhãn cú pháp thành phần. Chất lượng các hệ thống phân tích cú pháp huấn luyện trên những tập dữ liệu này còn thấp. Trong báo cáo này, chúng tôi đề xuất bổ sung đặc trưng biểu diễn phân bố từ cho phân tích cú pháp thành phần tiếng Việt. Sử dụng công cụ Shift-Reduce của nhóm NLP thuộc Đại học Stanford, việc thực nghiệm trên kho ngữ liệu Viettreebank đã chỉ ra rằng đặc trưng này cho phép cải thiện kết quả phân tích.

Từ khóa— Cú pháp thành phần, phân tích cú pháp thành phần, thuật toán Shift-reduce, biểu diễn phân bố từ vựng.

Nghiên cứu đối sánh về các mô hình mạng nơ-ron nhân tạo trong vấn đề phân loại câu

Lê Hồng Phương

Khoa Toán - Cơ - Tin học

Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Báo cáo này trình bày một nghiên cứu đối sánh toàn diện về bốn mô hình mạng nơ-ron, gồm mạng nơ-ron tiến, mạng nơ-ron nhân chập, mạng nơ-ron truy hồi và mạng bộ nhớ ngắn hạn trên hai bộ dữ liệu văn bản tiếng Anh và tiếng Việt. Trên bộ dữ liệu tiếng Anh, mạng nơ-ron nhân chập cho kết quả tốt hơn các mô hình phân loại mạnh khác mà không cần sử dụng các kỹ thuật trích chọn đặc trưng thủ công. Kết quả thực nghiệm cũng cho thấy các véc-tơ từ GloVe cho kết quả tốt hơn các véc-tơ từ Skip-gram. Mô hình mạng nơ-ron nhân chập cũng cho kết quả tốt hơn các mô hình phân loại cơ sở mạnh khác trên bộ dữ liệu báo chí tiếng Việt. Dựa trên những kết quả thực nghiệm này, chúng tôi đưa ra một số khuyến nghị trong việc ứng dụng các mô hình mạng nơ-ron nhân tạo trong vấn đề phân loại câu.

A Comparative Study of Neural Network Models for Sentence Classification

Le Hong Phuong

Faculty of Mathematics, Mechanics and Informatics

VNU University of Science

Abstract: This paper presents an extensive comparative study of four neural network models, including feed-forward networks, convolutional networks, recurrent networks and long short-term memory networks, on two sentence classification datasets of English and Vietnamese text. We show that on the English dataset, the convolutional network models without any feature engineering outperform some competitive sentence classifiers with rich hand-crafted linguistic features. We demonstrate that the GloVe word embeddings are consistently better than both Skip-gram word embeddings and word count vectors. We also show the superiority of convolutional neural network models on a Vietnamese newspaper sentence dataset over strong baseline models. Our experimental results suggest some good practices for applying neural network models in sentence classification.

Xử lý văn bản tiếng Việt trên mạng xã hội: Bài toán gán nhãn từ loại

Ngô Thế Quyền

Khoa Toán - Cơ - Tin học

Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Xử lý văn bản là một bài toán trong lĩnh vực xử lý ngôn ngữ tự nhiên. Đối với văn bản truyền thống (văn bản báo chí) đã có những nghiên cứu cũng như các công cụ xử lý văn bản cho hiệu quả tốt, tuy nhiên các công cụ này tỏ ra kém hiệu quả trên văn bản có nhiều nhiễu như văn bản trên mạng xã hội. Nghiên cứu tập trung vào một tác vụ trong xử lý văn bản là bài toán gán nhãn từ loại. Từ những đặc thù của văn bản trên mạng xã hội như: câu không đủ cấu trúc chủ vị, từ viết tắt, sai chính tả, . . . các bước tiền xử lý được đưa ra nhằm khử nhiễu trên văn bản đầu vào, nhằm tăng hiệu quả trong việc gán nhãn từ loại. Các bước tiền xử lý bao gồm: xác định và sửa các lỗi chính tả, chú giải các từ đặc biệt như từ viết tắt, số, tên riêng, url, . . . sau đó văn bản sẽ được tách từ và gán nhãn từ loại

Vietnamese social text processing: Part Of Speech Tagging

Ngo The Quyen

Faculty of Mathematics, Mechanics and Informatics

VNU University of Science

Abstract: Text processing is a problem in the field of natural language processing. For traditional text (newspaper texts), there are studies as well as text processing tools for good efficiency, but these tools are less effective on text with more noise such as social text. This research focused on a task in text processing is the problem of part of speech (POS) tagging (POS). From the specifics of the text on social networks such as short sentence, missing syntax component, abbreviations, misspellings, etc., preprocessing steps are introduced to reduce noise on input text, to increase the efficiency of POS tagging task. Preprocessing steps include: identifying and correcting misspelling errors, capturing special words such as acronyms, numbers, proper names, urls, and so on, then the text will be word segmented and POS tagging.

Mạng nơ-ron mờ với tập đại diện trong dự báo kết quả học tập của sinh viên

Lê Hoàng Sơn

Khoa Toán - Cơ - Tin học

Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Dự báo kết quả học tập của sinh viên là một vấn đề quan trọng trong giáo dục. Mục tiêu của bài toán này là dự báo kết quả học tập của sinh viên sau khi nhập học và

làm cơ sở để xác định những sinh viên có kết quả học tập tốt. Trong báo cáo này, chúng tôi đề xuất một mạng nơron mờ mới cho bài toán dự báo kết quả học tập của sinh viên với ý tưởng về sử dụng đa tham số và thuật toán học đặc biệt. Mô hình mới sử dụng đa tham số để tăng độ chính xác của dự báo trong sự kết hợp với các thuật toán học địa phương và toàn cục. Phương pháp đề xuất được kiểm chứng thực nghiệm so sánh với các thuật toán khác như ANFIS, MANFIS, OneR và Random Tree trên các bộ dữ liệu chuẩn UCI và dữ liệu thực tế.

Neural-Fuzzy with Representative Sets for Prediction of Student Performance

Le Hoang Son

*Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: The student performance prediction (SAPP) is an important matter in education. It predicts future performance of a student after being enrolled into a university; thus, determining who would do well and who would receive poor scores. Those results would with help making admission decisions more efficiently and improving the quality of academic services. In this report, we will present a new method for handling the Multi-Input Multi-Output Student Academic Performance Prediction (MIMO SAPP) problem. It uses multiple parameter sets and a special learning strategy to resolve those weaknesses. Specifically, the idea of multiple parameter sets is to approximate the model with many meaningful parameters to ensure the performance of system. This is regarded as the representative problem, which is mathematically formulated and theoretically validated. The idea for the special learning strategy is to use global and local training. In global training, a random parameter set is trained from the first to the last record of the database. Each time of training results in a set of parameters. Global training will rectify and achieve a meaningful subset of parameters by the last training process. In local training, there are 2 types of parameters namely premise and consequent that are trained by the gradient descent and Particle Swarm Optimization in a hybrid way. Lastly, for a new record in the testing set, Fuzzy K-Nearest Neighbor is used to find which group it belongs to. The proposed model is experimentally validated against ANFIS, MANFIS, OneR and Random Tree in a benchmark student performance dataset from UCI, a real student performance dataset from VNU University of Science, Vietnam, and 3 educational datasets taken from KDD Cup. The experiments demonstrated the superiority of the proposed MANFIS-S over the related algorithms in term of accuracy.

Tối ưu tuổi thọ của mạng cảm biến không dây ngầm

Nguyễn Thị Tâm

Khoa Toán - Cơ - Tin học

Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Mạng cảm biến không dây ngầm là mạng bao gồm các nút cảm biến được triển khai dưới lòng đất. Có ba loại kênh truyền thông chính trong mạng: kênh truyền trên mặt đất, kênh truyền dưới mặt đất và kênh truyền thông giữa hai môi trường. Một vấn đề quan trọng trong mạng là tối ưu năng lượng của cảm biến để tối đa tuổi thọ của mạng. Một trong những cách tiếp cận là sử dụng các nút chuyển tiếp để chuyển tiếp lưu lượng truy cập giữa các nút cảm biến và các trạm cơ sở. Cách tiếp cận này cũng gặp một vài thách thức về cân bằng tải giữa các nút chuyển tiếp, tổn thất truyền thông khi truyền trong đa môi trường. Chúng tôi sẽ đề xuất một thuật toán heuristic dựa trên việc phân cụm các cảm biến cho bài toán gán các nút cảm biến vào các nút chuyển tiếp và thuật toán trong thời gian đa thức cho bài toán cân bằng tải của các nút chuyển tiếp. Sau đó, chúng tôi cũng đề xuất một giải thuật tiến hóa để cải thiện kết quả của thuật toán heuristic trên.

Lifetime optimization for wireless underground sensor networks

Nguyen Thi Tam

Faculty of Mathematics, Mechanics and Informatics

VNU University of Science

Abstract: Wireless Underground Sensor Networks (WUSNs) are networks of sensor nodes which are deployed below ground. There are three different communication channels in WUSNs including underground to underground channel (UG - UG), underground to aboveground channel (UG-AG), aboveground (AG-AG) to aboveground channel. An important problem in WUSNs is the sensor energy optimization which determines the network lifetime and energy expenditure. Using relay nodes (RNs) to relay the traffic has several challenges including load balancing among relay nodes, transmission loss because of transmission in multi-environment. We will propose a new clusteringbased heuristic for load unrestricted relay node selection problem and a polynomial time exact algorithm for load balanced sensor node assignment problem. We also propose an evolutionary algorithm to improve and stabilize the results of our new heuristic.

Phân cụm mờ ảnh trực tuyến: một cách tiếp cận mới cho phân cụm mờ thời gian thực trên các tập mờ dữ liệu ảnh

Phạm Huy Thông
Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Phân cụm online đóng vai trò quan trọng trong phân cụm thời gian thực. Báo cáo trình bày một cải tiến cho thuật toán việc phân cụm mờ viễn cảnh online. Trước hết, các tham số cho mỗi thành phần được tính toán và đưa vào các luồng khởi tạo nhằm tiết kiệm thời gian tính toán cho việc khởi tạo lại các tham số cho toàn bộ dữ liệu. Cuối cùng, các bước lặp của thuật toán phân cụm mờ viễn cảnh được sử dụng cho toàn bộ dữ liệu cho việc phân cụm. Thực nghiệm đã chỉ ra thuật toán đề xuất có thời gian chạy nhanh hơn với chất lượng cụm giảm không đáng kể.

Online picture fuzzy clustering: a new approach for real-time fuzzy clustering on picture fuzzy sets

Pham Huy Thong
Faculty of Mathematics, Mechanics and Informatics
VNU University of Science

Abstract: Online clustering plays an important roles in real-time clustering. In this paper, an innovative algorithm to online picture fuzzy clustering is proposed. Firstly, we calculate parameter for each element added for each steam for initialization. This step saves a lot time because of re-randomizing the initial of all data and calculating it through steps. Finally, we employ the iteration of picture fuzzy clustering for all data to partition clusters until getting ended condition. Experiments point out that the proposed algorithm has better run-time with trivial quality reduction.

Thuật toán DC khử nhiễu ảnh dựa trên học từ điển

Nguyễn Thị Bích Thủy
Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG HN

Tóm tắt: Khử nhiễu ảnh sử dụng biểu diễn thưa là một hướng nghiên cứu mới trong thời gian gần đây. Trong đó, để tìm dạng biểu diễn thưa của ảnh bằng phương pháp học từ điển, chúng ta sẽ phải giải quyết một bài toán tối ưu với hàm mục tiêu chứa chuẩn ℓ_0 . Do là một hàm không liên tục tại điểm 0 nên bài toán tối ưu này là một bài toán khó (NP-hard). Ngoài các hướng tiếp cận để giải bài toán bằng phương pháp tham lam, xấp xỉ chuẩn bằng một hàm lồi, phương pháp xấp xỉ chuẩn bằng một hàm không lồi đã được chứng minh có thể cho lời giải chính xác hơn.(Vo và Le Thi, 2015) đã trình bày thuật giải DC cho bài toán học từ điển khi sử dụng xấp xỉ chuẩn bằng hàm capped- (Peleg và

Meir, 2008). Trong bài báo này, chúng tôi nghiên cứu cách sử dụng hàm mũ (Bradley, 1998) để xấp xỉ chuẩn và áp dụng thuật giải DC (Hiệu của hai hàm lồi – Difference of Convex functions) cho bài toán mới này cũng như tiến hành thực nghiệm trên bài toán khử nhiễu ảnh. Phương pháp tối ưu DC và giải thuật DCA là một hướng tiếp cận cho các bài toán tối ưu không lồi và đã được áp dụng thành công trên rất nhiều lĩnh vực. Kết quả thực nghiệm cho thấy tính khả thi và hiệu quả của phương pháp đề xuất là rất tốt.

DC Algorithm for Image Denoising based on Dictionary Learning

Nguyen Thi Bich Thuy

*Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: Image denoising is a challenging task and defined as an inverse problem of estimating a signal from the noisy measured version. (Vo and Le Thi, 2015) has presented a DC algorithm for the image denoising problem using sparse representation and dictionary learning when approximated l_0 -norm by capped- l_1 function (Peleg and Meir, 2008). In this report, we will study the Exponention function (Bradley, 1998) to approximate l_0 -norm and propose a DC algorithm for this new approximation problem. The algorithm will be applied in denoising problem. Experimental results on the real data sets are promising and demonstrate the effectiveness of our approach.

Clustering Functional Data using Projection

Pham Huy Tung

*Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: We show that, in the functional data context, by appropriately exploiting the functional nature of the data, it is possible to cluster the observations asymptotically perfectly. We demonstrate that this level of performance can sometimes be achieved by the k-means algorithm as long as the data are projected on 1 dimensional space. In general, the notion of ideal cluster is not clearly defined. We derive our results in the setting where the data come from two populations whose distributions differ at least in terms of means, and where an ideal cluster corresponds to one of these two populations. We propose an iterative algorithm to choose the projection functions in a way that optimises clustering performance. We apply our iterative clustering procedure on simulated and real data, where we show that it works well.

Các thuật toán mã hóa giả xác suất

*Nguyễn Hải Vinh
Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự nhiên, ĐHQG HN*

Tóm tắt: Mô tả một hướng mật mã ứng dụng liên quan tới các bài toán bảo mật thông tin trong tình huống tấn công ép buộc. Đặc trưng của hướng này là nguyên lý không phân biệt được về mặt tính toán so với mã hóa xác suất.

Pseudo-probabilistic encryption algorithms

*Nguyen Hai Vinh
Faculty of Mathematics, Mechanics and Informatics
VNU University of Science*

Abstract: In this report there is described an applied cryptography direction related to information security problems in rubber hose attack situations. The characteristic of this direction is the principle of computational indistinguishability from probabilistic encryption.