Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0499U000811

Особливі позначки: відкрита

Дата реєстрації: 18-09-1999

Статус: Захищена

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Харченко Андрій Вячеславович
2. Kharchenko Andrij Vyacheslavovych

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: кандидат наук

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 01.05.03

Назва наукової спеціальності: Математичне та програмне забезпечення обчислювальних машин і систем

Галузь / галузі знань: Не застосовується

Освітньо-наукова програма зі спеціальності: Не застосовується

Дата захисту: 09-04-1999

Спеціальність за освітою: 22.01

Місце роботи здобувача: Інститут кібернетики імені В. М. Глушкова

Код за ЄДРПОУ: 05417173

Місцезнаходження: 252022 Київ, проспект Академіка Глушкова, 40

Форма власності:

Сфера управління: Національна академія наук України

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): 26.194.02

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Інститут кібернетики імені В. М. Глушкова

Код за ЄДРПОУ: 05417173

Місцезнаходження: 252022 Київ, проспект Академіка Глушкова, 40

Форма власності:

Сфера управління: Національна академія наук України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації:

Коди тематичних рубрик: 50.41.01

Тема дисертації:

1. Інтеграція методів аналізу великих наборів даних у інтелектуальному інтерфейсі користувача

2. Integration of methods for massive data sets analysis in intelligent user interfaces

Реферат/анотація:

1. Робота присвячена методам аналізу великих наборів даних та їх використанню для інтелектуалізації інтерфейсу користувача. Головна увага приділена питанням маніпулювання базою даних, що містить великі набори, операціям над об'єктами аналізу за допомогою спеціалізованої мови програмування та автоматичному моделюванню великих наборів часових даних. Запропоновані теоретичні підходи та алгоритми, які були застосовані у програмних системах Predictor, Forecasting ToolPak та ГеоІнт, що використовуються у виробництві.

2. Multiple data sets (MDS) emerge from the progress of measurement tools, technologies of data storage, and accumulation of data in VLDB and data warehouses. Methods of solving problems of ordinary data sets analysis, e.g. global modeling, discovery of dependencies, exploratory data analysis and manipulation, get less applicable when dealing with massive data owing to huge overheads in time and computer memory to produce models. Even if workable, the models reveal poor performance on the whole MDS. The main approach of adequate solving the problems that this thesis is based on is using artificial intelligence methods for flexible modelling data and

providing the tradeoff between generality of models being created and the size of a MDS, and computation resources. Besides special algorithms implemented, a necessary component of MDS analysis software is means for user's control of the algorithms' run, representation of results, decision making support, control of data input - that is, the user interface. A ccordingly, a part of research in the thesis was aimed at integration of modelling techniques developed in it as well as known ones within the so called intelligent user interface. The principal idea behind this integration is an approach proposed in the thesis - modelling automation which requires that composing training subsets of possible MDS's models, training them, selection of the most adequate one, its timely adjustment, retraining and discarding as the training sets grow or new examples get available be performed without human intervention. Some domains like geophysical interpretation require more complex analysis, e.g. multi-phase building different models of an MDS, its consequent modification by a human expert by the knowledge about models' performance, support of analysis alternatives (models developed and their training data including intermediate ones). Features of this type of analysis are inadequacy of automatic modelling and moreover the need of means for interactive analysis. In the thesis, a specialized programming language with constructs for data analysis operations is proposed. Its features are opeations for manipulations of MDS's and their constituents (parameter vectors) as instances of special data types - collections, sub-collections, and single elements. In the language, various MDS analysis objects (vectors of parameters and their sets - classes, clusters created via different types of analysis) can be represented and manipulated with various domain-specific operations, e.g. (a) forming classes out of parameter vectors or their projections stored in the database and treating them via set theory operations, (b) processing single parameter vectors, e.g. classifying and projecting vectors, (c) performing control of a previousely built-up classification configuration, e.g. automatic classification by universal methods (decision trees, nearest neighbours, fuzzy techniques), retraining, discarding and merging classes. In the thesis, specifics of MDS primarily those resulting in inadequacy of their models were considered and a principle of limited history impact was proposed. It claimes that the joint probability distribution of the unknown function modelled and its parameters drift in time as the effect of the underlying process of an MDS is stable only locally but not globally; so local models only can be trained on the data set. Based on the order of locally stable subsets, an MDS can be further analysed as an equivalent time series. (A substantial fraction of massive datasets are in fact temporal.) An algorithm for local training models of temporal MDS's were presented along with a criterion of model performance. Another algorithm based on the same criterion seeks a boundary between adjasent local subsets as a point of unimprovable model performance. Statistical aspects of training on MDS's were considered. By assuming parameter drift of an underlying process as a series of discrete, binomially distributed "revolutions", estimates of the VC confidence of the expected risk in the classification and regression regain problems for the case of a training set being an MDS. With the probability of "revolution" vanish, the estimates approach to the Vapnik-Chervonenkis ones.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки:

Стратегічний пріоритетний напрям інноваційної діяльності:

Підсумки дослідження:

Публікації:

Наукова (науково-технічна) продукція:

Соціально-економічна спрямованість:

Охоронні документи на ОПІВ:

Впровадження результатів дисертації:

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Перевозчикова Ольга Леонідівна
2. Перевозчикова Ольга Леонідівна

Кваліфікація: д.ф.-м.н., 01.05.03

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Грінченко Тамара Олексіївна
2. Грінченко Тамара Олексіївна

Кваліфікація: д.ф.-м.н., 01.05.01

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Гороховський Семен Самуїлович
2. Гороховський Семен Самуїлович

Кваліфікація: к.ф.-м.н., 01.05.01

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Рецензенти

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради	Сергієнко Іван Васильович
Власне Прізвище Ім'я По-батькові головуючого на засіданні	Сергієнко Іван Васильович
Відповідальний за підготовку облікових документів	
Реєстратор	

Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності

Юрченко Т.А.