**Дорошенко Анастасія Володимирівна. Методи та інформаційні технології класифікації для завдань видобування даних : Дис... канд. наук: 05.13.06 – 2008**

|  |  |
| --- | --- |
| |  | | --- | | Дорошенко А.В. Методи та інформаційні технології класифікації для завдань видобування даних. – Рукопис.  Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – Інформаційні технології. – Національний університет ”Львівська політехніка”, Львів, 2008.  В дисертації вирішується актуальна науково-прикладна задача – розроблення методів та інформаційних технологій класифікації для завдань видобування даних на основі моделі геометричних перетворень.  В дисертаційній роботі з метою врахування такої особливості задач видобування даних, як різна вага помилок під час класифікації, розроблено метод та інформаційну технологію класифікації на основі правила „переможець забирає все” (WTA) з використанням матриці штрафів та заохочень, що надає можливість враховувати неоднакову вагу помилок при розпізнаванні об’єктів різних класів. Розроблено метод кускової побудови розділяючих поверхонь на основі дерева поділу на класи, що забезпечує підвищення точності розпізнавання об’єктів, для яких в заданому просторі реалізацій не виконується гіпотеза компактності. Запропоновано, досліджено і реалізовано метод послідовного вирівнювання тренувальної вибірки по кластерах, що формуються у вузлах дерева, який забезпечує покращання точності розв’язання задач класифікації для вибірок, представлених в просторі реалізацій нерівномірно. Вдосконалено метод навчання нейроподібних структур моделі геометричних перетворень доповненням його процедурою оптимізації методом імітації відпалу металу за параметрами головних компонентів, що дало змогу покращити якість класифікації завдяки використанню інших, окрім середньоквадратичного, критеріїв якості. Розроблено та застосовано програмні скрипти мовою Python, які виконують передобробку даних та реалізують розроблені метод вирівнювання вибірки, метод штрафів і заохочень та кусковий метод побудови розділяючих поверхонь. Створено інформаційні технології на основі розроблених методів, які застосовано для задач електронної комерції. Наведено результати проведених експериментів, які свідчать про ефективність застосованих методів. | |
| |  | | --- | | У дисертаційній роботі розв’язано актуальну наукову задачу вдосконалення існуючих та розроблення нових ефективних методів та інформаційних технологій класифікації для задач видобування даних. При цьому отримано такі основні результати:   1. Проведено аналітичний огляд існуючих методів класифікації. Проаналізовано їх ефективність для розв’язання задач видобування даних. Встановлено переваги та перспективність застосування моделі геометричних перетворень для розв’язання задач великої розмірності. 2. Вперше для нейроподібних структур моделі геометричних перетворень розроблено метод класифікації на основі правила „переможець забирає все” з використанням матриці штрафів та заохочень, що надає можливість враховувати неоднакову вагу помилок при розпізнаванні об’єктів різних класів. 3. Вперше для нейроподібних структур моделі геометричних перетворень розроблено метод кускової побудови розділяючих поверхонь на основі дерева поділу на класи, що забезпечує підвищення точності розпізнавання об’єктів, для яких в заданому просторі реалізацій не виконується гіпотеза компактності. 4. Вперше запропоновано, досліджено і реалізовано метод послідовного вирівнювання тренувальної вибірки по кластерах, що формуються у вузлах дерева, який забезпечує покращання точності розв’язання задач класифікації для вибірок, представлених в просторі реалізацій нерівномірно. 5. Вдосконалено метод навчання нейроподібних структур моделі геометричних перетворень доповненням його процедурою оптимізації методом імітації відпалу металу за параметрами головних компонентів, що дало змогу покращити якість класифікації завдяки використанню інших, окрім середньоквадратичного, критеріїв якості. 6. Розроблено та застосовано інформаційні технології на базі програмних скриптів мовою Python, які виконують передобробку даних та реалізують розроблені метод вирівнювання вибірки, метод штрафів і заохочень та кусковий метод побудови розділяючих поверхонь. 7. Розроблені інформаційні технології прийняття рішень для завдань електронної комерції (класифікація клієнтів, виявлення шахраїв, класифікація лотів Інтернет-аукціону eBuy)дозволили підвищити ефективність розв’язання задач видобування даних великих розмірностей (порядку сотні атрибутів), представлених тренувальними вибірками значних обсягів (десятки тисяч векторів) в умовах, коли в заданому просторі реалізації не виконується гіпотеза компактності. 8. Застосування розроблених методів налагодження та оптимізації нейроподібних структур і програмних продуктів на їх основі в задачах електронної комерції забезпечило підвищення точності класифікації в середньому на 2–7 %, а в окремих випадках на 10–20 % (від 85 до 95 % правильно класифікованих даних), скоротило час отримання розв’язків приблизно в 5 разів (з 60–120 до 10–30 секунд), що свідчить про ефективність застосованих методів. | |