**Волченко Олена Володимировна. Комплексний метод побудови вирішуючих правил імовірнісних систем автоматичного розпізнавання, що навчаються : Дис... канд. наук: 05.13.23 – 2008**

|  |  |
| --- | --- |
| |  | | --- | | **Волченко О.В. Комплексний метод побудови вирішуючих правил імовірнісних систем автоматичного розпізнавання, що навчаються. – Рукопис.**  Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – Системи та засоби штучного інтелекту. – Державний університет інформатики та штучного інтелекту, Донецьк, 2008.  Дисертація присвячена рішенню актуальної задачі розробки ефективних методів побудови вирішуючих правил імовірнісних систем автоматичного розпізнавання відкритого типу, що навчаються. Запропонований і обґрунтований новий метод побудови скороченої зваженої навчальної вибірки мета-об'єктів, вага яких розраховується до побудови вирішуючих правил класифікації, що дозволило істотно скоротити навчальну вибірку. Запропонований новий метод коректного додавання даних у вибірку мета-об'єктів, необхідний при побудові систем розпізнавання відкритого типу, що навчаються. Запропонований метод побудови потенційних функцій по зважених навчальних мета-вибірках, що дозволило використовувати при побудові вирішуючих правил значення ознак розпізнаваних об'єктів і знання про розташування об'єктів навчальної вибірки у просторі ознак. Розроблений генетичний алгоритм побудови вирішуючих правил класифікації, в якому правило представлене у вигляді псевдобулевої функції, запропонований спосіб імовірнісного формування початкової популяції хромосом, що враховує вагу об'єктів навчальної вибірки, запропонований метод обчислення подвійної фітнес-функції, визначений і теоретично обґрунтований набір генетичних операторів. Запропонований новий метод побудови вирішуючих правил імовірнісних систем автоматичного розпізнавання відкритого типу, що навчаються, що комплексно включає спосіб визначення відособленості класів, метод формування зваженої навчальної вибірки мета-об'єктів, генетичний алгоритм побудови вирішуючого правила, що дозволив підвищити якість розпізнавання, знизити часові витрати на побудову вирішуючого правила і виконання класифікації, скоротити об'єм даних, що зберігаються. | |
| |  | | --- | | У дисертаційній роботі запропоновано нове рішення актуальної задачі розробки ефективних методів побудови вирішуючих правил імовірнісних систем автоматичного розпізнавання відкритого типу, що навчаються. Запропоноване рішення полягає в застосуванні комплексного підходу, що включає як обробку вхідних даних, так і побудову вирішуючих правил з використанням генетичних алгоритмів. Отримані результати мають важливе наукове і прикладне значення для створення ефективних систем автоматичного розпізнавання, що навчаються, особливо систем відкритого типу. Проведені дослідження дозволили отримати наступні результати.  1. В результаті аналізу сучасного стану завдання побудови систем автоматичного розпізнавання, методів побудови вирішуючих правил у системах розпізнавання, що навчаються, і методів скорочення навчальних вибірок було показано, що існуючі методи не дозволяють будувати ефективні за якістю і часом розпізнавання системи для прикладних завдань, що характеризуються значним об'ємом початкових даних, і можливістю їх поповнення в процесі роботи.  2. На підставі аналізу можливих варіантів розташування об'єктів навчальної вибірки у просторі ознак був запропонований спосіб визначення відособленості класів шляхом побудови кардинальної гіперплощини, що дозволяє одержувати вирішуюче правило для відособлених класів мінімальної довжини.  3. Вперше був запропонований і обґрунтований метод скорочення початкової навчальної вибірки шляхом побудови вибірки мета-об'єктів для двокласових і багатокласових систем розпізнавання, що дозволяє по безлічі розташованих поруч об'єктів деякого класу будувати мета-об'єкти, що описуються окрім значень ознак розпізнавання масою. Виконаний аналіз способів визначення початкової і конкуруючої точок побудови мета-об'єктів, оцінена достовірність розробленого методу по статистичних характеристиках початкової і одержаної навчальних вибірок. Для систем розпізнавання відкритого типа запропонований новий метод додавання нових об'єктів в навчальну мета-вибірку, запропонований новий метод коректного додавання даних у вибірку мета-об'єктів, що є необхідною умовою створення систем розпізнавання відкритого типу, що навчаються.  4. Одержав подальший розвиток метод потенційних функцій шляхом розповсюдження його на зважені навчальні мета-вибірки, що дозволило використовувати при побудові вирішуючих правил не тільки значення ознак розпізнаваних об'єктів, але знання про розташування об'єктів навчальної вибірки у просторі ознак.  5. Розроблений генетичний алгоритм побудови вирішуючого правила по мета-вибірці, у якому вперше правило представлене у вигляді псевдобулевої функції, запропонований спосіб імовірнісного формування початкової популяції хромосом, що враховує вагу об'єктів навчальної вибірки, запропонований метод обчислення подвійної фітнес-функції, визначений і обґрунтований набір генетичних операторів, серед яких виділяються знижуюча та підвищуюча мутації, і оператор циклічного зрушення.  5. Запропонований новий метод побудови вирішуючих правил імовірнісних систем автоматичного розпізнавання, що навчаються, який комплексно включає спосіб визначення відособленості класів, метод формування зваженої навчальної вибірки мета-об'єктів, генетичний алгоритм побудови вирішуючого правила, що дозволив підвищити якість розпізнавання, знизити часові витрати на побудову вирішуючого правила і виконання класифікації, скоротити об'єм даних, що зберігаються.  6. У експериментах на тестових даних запропонований комплексний метод показав збільшення здатності розпізнавання одержуваними вирішуючими правилами, у порівнянні з вирішуючими правилами, побудованими класичними методами, в середньому на 11% і скорочення часу класифікації розпізнаваних об'єктів на 36%. У експериментах на відомих тестових прикладах комплексний метод дозволив побудувати вирішуючі правила, що забезпечують в середньому 96,88%% правильних класифікацій, що на 3,5%% більше результатів, одержуваних найбільш ефективними на сьогодні методами побудови вирішуючих правил класифікації, зокрема, методами на основі генетичних алгоритмів.  7. Застосування розробленого комплексного методу для вирішення завдання побудови спам-фільтрів при фільтрації вхідної електронної кореспонденції дозволило підвищити точність класифікації листівок у середньому на 37%.  8. Результати, одержані в дисертаційній роботі, використані у науково-дослідних роботах відділу розпізнавання мовних образів Донецького інституту проблем штучного інтелекту МОН і НАН України, у розробках ТОВ «Бі-Тек», а також в учбовому процесі кафедри програмного забезпечення інтелектуальних систем Державного університету інформатики і штучного інтелекту. | |