**Хмільовий Сергій Володимирович. Інформаційна технологія видобування знань для прогнозування часових рядів на прикладі завантаженості обладнання зв'язку. : Дис... канд. наук: 05.13.06 - 2009.**

|  |  |
| --- | --- |
| |  | | --- | | **Хмільовий С.В. Інформаційна технологія видобування знань для прогнозування часових рядів на прикладі завантаженості обладнання зв’язку.** - Рукопис.  Дисертація на здобування вченої степені кандидату технічних наук за спеціальністю 05.13.06 – інформаційні технології. – ДонНУ, Донецьк, 2008.  В дисертації розглядається важлива науково-практична проблема видобування знань при прогнозуванні часових рядів.  На етапі попередньої обробки даних найважливішим для даної задачі признано відбір значущих факторів. За умов стохастичності оцінки набору факторів (що отримується при прогнозуванні часових рядів за допомогою нейронних мереж) відомі методи мало використовувані та потребують модифікації. Запропоновано компактний генетичний алгоритм, модифіковано величину зміни його імовірнісного вектору. Модифікація змінює величину імовірнісного вектору КГА в залежності від достовірності порівняння підмножин атрибутів.  Задача видобування знань виконувалась шляхом автоматичної побудови бази нечітких правил (бази знань). Автоматична побудова правил виконується еволюційним алгоритмом.  Для поліпшення точності правил бази знань, що створюється, модифікована фітнес-функція еволюційного алгоритму. З базової функції зберігаються частини, що відповідають за несхожість правил та за мале покриття помилкових точок. Головною «точністною» частиною є J-міра.  Модифікована лінгвістична база даних (ЛБД) для поліпшення точності отримуваної бази знань. Модифікація системи нечіткого висновку шляхом заміни алгоритму Мамдані на спрощений алгоритм нечіткого висновку дозволяє використовувати любий вид функції належності. Оптимізація параметрів ЛБД на основі довільного виду функції належності за допомогою (1+1) – еволюційної стратегії дозволило практично досягнути точність прогнозу, що отримується НМ.  Вперше для створення бази знань на основі нечітких правил розроблено паралельний еволюційний алгоритм. Він дозволяє збільшувати швидкість створення бази знань практично лінійно при малій кількості клієнтів, що генерують правила.  Модифіковано етап постпроцесінгу, де запропоновано засоби як для поліпшення точності, так і для інтерпретуємості правил: мультисимпліфікація, зважування правил, тюнинг. Доведено ефективність введення 1+1 – еволюційної стратегії в процедуру тюнингу.  Розроблені алгоритми програмного забезпечення інформаційної технології, спроектована ієрархія об’єктів для об’єктно-орієнтованої програмної реалізації. Створене програмне забезпечення апробоване на тестових наборах (benchmarks), доведена успішність запропонованих методів та модифікацій. Розроблено технічне і організаційне забезпечення технології.  Проведена апробація пакету на ОАО «Промтелеком» для задачі прогнозування завантаженостіобладнання і прогнозування для цього кількості підключень до АТС підприємства. | |
| |  | | --- | | У дисертаційній роботі вирішено актуальну наукову задачу створення інформаційної технології для прогнозування часових рядів на прикладі прогнозування кількості підключень до АТС, що використовується для визначення завантаженості обладнання зв’язку. Розроблене математичне забезпечення інформаційної технології дозволило підвищити точність та інтерпретованість одержуваного прогнозу за рахунок створення бази знань на основі нечітких правил, що настроюють за допомогою еволюційного алгоритму. Дана база знань допомагає підвищити обґрунтованість прийняття рішень про подальший розвиток АТС у тому числі особам, які не володіють спеціальними математичними знаннями, що підтверджується актом про впровадження.  У процесі проведення досліджень були отримані такі основні результати:   1. Розроблено послідовність створення інформаційної технології для виділення знань (одержання бази знань на основі нечітких правил) у задачах прогнозування часових рядів, що регламентує процес її створення. Послідовність містить 7 основних етапів: попередня обробка даних, структурний та параметричний синтез системи нечіткого висновку, постпроцесинг створених правил (перші 4 етапи становлять математичне забезпечення розробленої технології), створення програмного забезпечення, розробка технічного й організаційного забезпечення і впровадження інформаційної технології. Для параметричної ідентифікації й постпроцесингу впроваджується еволюційний алгоритм створення й настроювання правил бази знань. 2. Для вирішення задачі відбору факторів використано компактний генетичний алгоритм, що відрізняється простотою реалізації при конгруентних із класичними ГА результатами. Для умови стохастичності оцінки набору факторів модифіковано крок зміни імовірнісного вектора КГА. Модифікація основана на визначенні залежності величини зміни імовірнісного вектора від ступеня вірогідності порівняння міри оцінки підмножин атрибутів, і дозволила істотно скоротити час пошуку (на різних задачах до 90.6%). 3. Набув подальшого розвитку метод побудови фітнес-функцій еволюційних алгоритмів. Як основну точністну складову фітнес-функції нечіткого еволюційного алгоритму вперше використано J-міру. Це дало можливість збільшити точність прогнозування для задач, що вимагають побудови правил вільної семантики, на 7.9-21.7%. 4. Набув подальшого розвитку метод побудови паралельних еволюційних алгоритмів. Побудовано розподілений еволюційний алгоритм для паралельного обчислення нечітких правил на основі моделі «робітник-хазяїн». На відміну від відомих методів, процес-сервер робить корекцію навчальної вибірки для процесів-клієнтів, що дає кращу настроєність створюваних правил на малоописані області простору пошуку. Експериментально отримано лінійне зростання швидкості створення бази знань при обмеженій кількості клієнтів (4-5). 5. Модифікація системи нечіткого висновку шляхом заміни алгоритму Мамдані на спрощений алгоритм нечіткого висновку дозволяє збільшити швидкість нечіткого висновку (у порівнянні з алгоритмом Mамдані на 41.8% для тестового набору). Це дозволило застосовувати довільну зі стандартних функцій належності, застосування яких раніше було обчислювально складнішим у зв'язку з неможливістю аналітичного й складністю чисельного інтегрування. 6. Вперше для нечітких продукційних систем використано еволюційну стратегію для оптимізації параметрів лінгвістичної бази даних. Оптимізація цих параметрів для настроювання методу побудови нечіткої бази знань на основі довільної функції належності дала можливість на порядок (у 12-16 разів) поліпшити точність прогнозування й досягти точності прогнозування, одержуваної за допомогою НМ, на довільно складних задачах. 7. Модифіковано фітнес-функцію на етапі постпроцесингу при використанні процедури спільної мультисимпліфікації й зважування. Модифікація дозволила одержувати більш короткі, більш інтерпретовані бази знань за рахунок погіршення точності. Для величини важливості короткості бази знань *wl* = 0.3 відбулося зростання СКВ від 2,8 до 4,1% при зростанні похибки прогнозування максимум на 0.37% і зменшенні кількості правил на 36.7% для задачі прогнозування кількості підключень до АТС. 8. За розробленими алгоритмами і спроектованою ієрархією об'єктів створено програмне забезпечення інформаційної технології (інструментальні засоби) для автоматизованого одержання баз знань, що містять набори інтерпретованих нечітких правил. За допомогою таких баз знань можуть вирішуватися задачі одержання інтерпретованого прогнозу довільної задачі. Точність отриманого результату може варіюватися залежно від необхідного ступеня інтерпретованості результату. Реалізація програмного забезпечення зроблена з використанням мови Object Pascal. 9. Розроблено організаційне забезпечення технології: набір технологічних інструкцій, що регламентують виконання технологічних операцій та їх застосування в конкретних умовах. Розроблено технічне забезпечення технології – комплекс апаратних засобів, що дає можливість безперешкодної роботи технології. 10. Зроблено впровадження створеної системи для вирішення задачі прогнозування. Результати роботи використані для розв`язання задачі прогнозування кількості підключень до АТС ВАТ «Промтелеком». Це необхідно для розрахунку окупності капіталовкладень в АТС. Отримана прогнозна модель інтерпретована, що дає можливість керівному персоналу, який не володіє спеціальними математичними знаннями, обґрунтовувати прийняття рішень при розширенні підприємства, що є необхідним при використанні позикових коштів. 11. Результати роботи використано у держбюджетних темах Н-3-07 (застосовані методи попередньої обробки даних, у тому числі для задач прогнозування); Д-11-04 (застосовано розроблені методи для створення еволюційного алгоритму видобування знань для задач класифікації); Д-11-07 (використано розроблені методи видобування знань у задачах прогнозування); а також у навчальному процесі кафедри «Автоматизовані системи управління» Донецького національного технічного університету. | |