



Міністерство освіти і науки України

Полтавський національний технічний університет
імені Юрія Кондратюка

Кафедра комп'ютерних та інформаційних технологій та систем

data mining

smarter websites

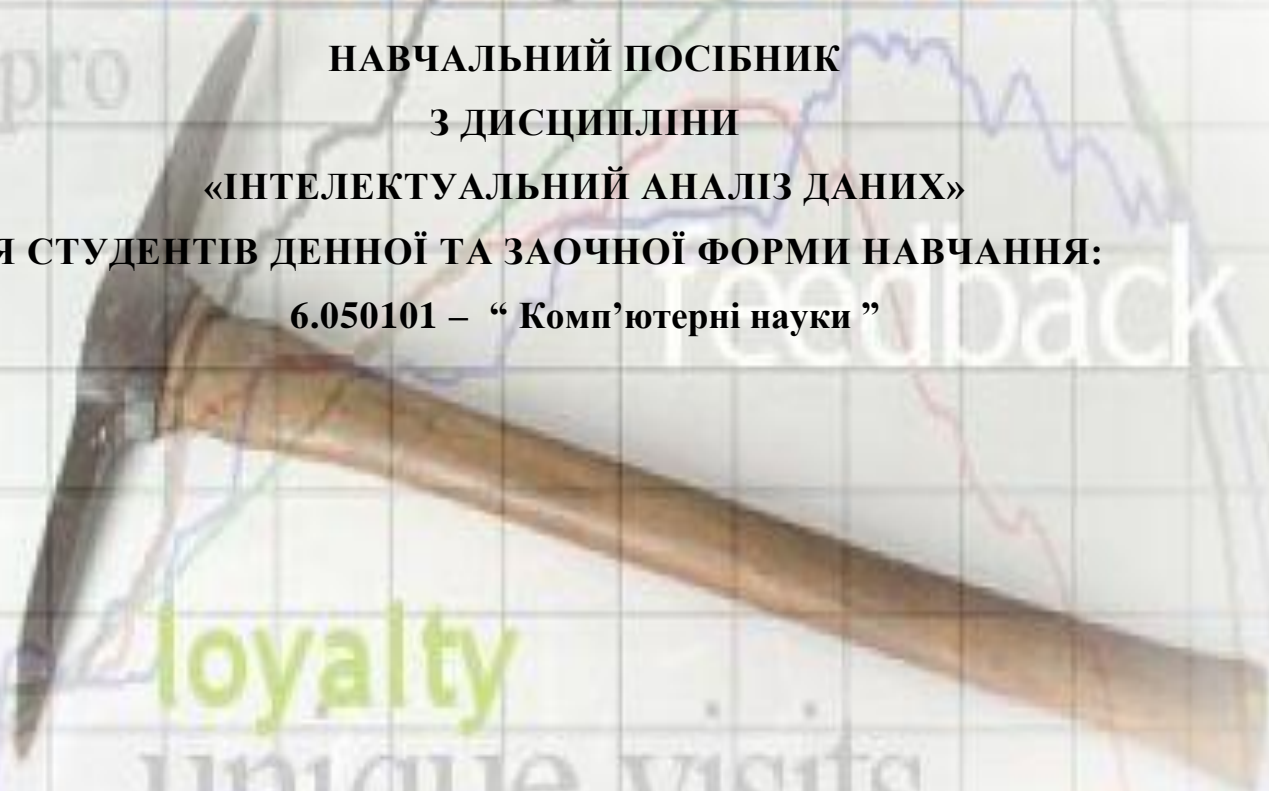
НАВЧАЛЬНИЙ ПОСІБНИК

З ДИСЦИПЛІНИ

«ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ»

ДЛЯ СТУДЕНТІВ ДЕННОЇ ТА ЗАОЧНОЇ ФОРМИ НАВЧАННЯ:

6.050101 – “ Комп'ютерні науки ”



Полтава 2014



Навчальний посібник з дисципліни «Інтелектуальний аналіз даних» для студентів денної та заочної форми навчання: 6.050101 – «Комп'ютерні науки» – Полтава: ПолтНТУ, 2014. – 187 с.

Укладачі: кандидат техн. наук., доцент, С.П. Альошин, асистент кафедри О.О. Бородіна.

Відповідальний за випуск: завідувач кафедри комп'ютерних та інформаційних технологій і систем О.Л. Ляхов, доктор техн. наук, професор

Рецензент: к.е.н., доцент кафедри КІТіС А.М. Гафіяк

Затверджено науково-
методичною радою університету
Протокол № 6 від 29.12.2014 р.

Коректор

40.40.01.01



ЗМІСТ

ВСТУП	6
1. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ.....	7
1.1. Вихідні визначення та поняття	7
1.2. Завдання, методи та засоби інтелектуального аналізу даних	10
1.3. Варіативне моделювання.....	11
2. ІНТЕЛЕКТУАЛЬНІ ТЕХНОЛОГІЇ	13
2.1. Від набору даних до нових знань. Логіка еволюції.....	13
2.2. Технології та методи виявлення знань.....	14
3. МОДЕЛЮВАННЯ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ДАНИХ	19
3.1. Навчання як процедура реалізації штучного інтелекту	19
3.2. Машинне розпізнання як основа інтелектуального аналізу даних.....	21
3.3. Перцептрон – інструмент інтелектуальної обробки даних	23
3.3.1. Навчання перцептрона	26
3.3.2. Лінійна роздільність як механізм розпізнавання образів ...	28
3.3.3. Алгоритми навчання перцептрону	30
4. РОЗПІЗНАВАННЯ ОБРАЗІВ ЯК ЗАВДАННЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.....	33
4.1. Поняття про образ, алфавіт класів, словник ознак, опис класів на мові ознак.....	33
4.2. Модель статистичного класифікатора	39
4.3. Априорна та апостеріорна інформація про об'єкт. Етапи побудови системи розпізнавання класів.....	42
4.4. Етапи побудови системи розпізнавання образу	45



5. ПРЕДОБРОБЛЕННЯ ДАНИХ ТА ОЦІНКА ЯКІСТІ МОДЕЛЕЙ КЛАСИФІКАЦІЇ В STATISTICA NEURAL NETWORKS (SNN).....	46
5.1. Предоброблення даних	46
5.2. Оцінка якості моделей класифікації.....	49
5.2.1. Методика вибору підмножин спостережень	49
5.2.2. Вибір порогів довірчої класифікації	51
6. РЕГРЕСІЙНИЙ АНАЛІЗ ДАНИХ	58
6.1. Загальні відомості про регресійний аналіз даних (РАД)	58
6.2. Технологія регресійного аналізу	62
7. ЗНАЧИМОСТЬ ДАНИХ В РЕГРЕСІЙНОМУ АНАЛІЗІ	71
7.1. Проблема значущості даних	71
7.2. Технологія оцінки значущості даних	75
7.2.1. Методика побудови моделі	77
8. АНАЛІЗ ЧАСОВИХ РЯДІВ.....	80
8.1. Аналіз часових рядів (ЧР) як задача прогнозування	80
8.2. Інтелектуальні технології аналізу часових рядів (ЧР)	84
9. ЯКІСТЬ ПРОГНОЗІВ НА ОСНОВІ ЧАСОВИХ РЯДІВ	92
9.1. Продуктивність та помилки аналізу часових рядів. Забезпечення якості обробки даних у середовищі нейроемулятор.....	92
9.2. Достовірність проєкції часових рядів	96
10. КЛАСТЕРНИЙ АНАЛІЗ ДАНИХ У СЕРОДОВИЩЕ ШТУЧНОЇ НЕЙРОМЕРЕЖІ (ШНМ).....	99



10.1. Особливості кластерного аналізу в середовищі нейроемулятор.....	99	
10.2. Алгомеративний метод ієрархічної класифікації або ієрархічний кластерний аналіз	106	
11. КЛАСТЕРНИЙ АНАЛІЗ В СЕРЕДОВИЩЕ		
СОМОРГАНІЗУЮЧИХ КАРТ КОХОНЕНА (СОКК)	110	
11.1. Загальний огляд самоорганізуючі карти Кохонена (СОКК)	110	
11.2. Методика побудови моделей аналізу в середовищі самоорганізуючих карт Кохонена (СОКК)	115	
12. КОМПЛЕКСНИЙ АНАЛІЗ ДАНИХ МОДЕЛЯМИ		
СОМОРГАНІЗУЮЧИХ КАРТ КОХОНЕНА (СОКК)	119	
12.1. Узагальнена постановка завдання аналізу структури даних	119	
12.2. Аналіз даних по перетинах інформаційної матриці	121	
12.3. Метод К-середніх	124	
13. ОСОБЛИВОСТІ АНАЛІЗУ НЕОДНОРІДНИХ ДАНИХ		
МОДЕЛЯМА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ	126	
13.1. Неоднорідність вихідних даних в практичних завданнях	126	
13.2. Нейромережеві моделі прогнозу в умовах неоднорідності даних.....	129	
14. ФОРМУВАННЯ ВХІДНИХ ВИБІРОК.....		134
14.1. Формування вибірок для навчання	134	
14.2. Формування вибірок для пониження розмірності.....	138	
15. НЕЙРОМЕРЕЖЕВА ОЦІНКА РИЗИКІВ ПРИЙНЯТТЯ		
РІШЕНЬ.....	145	



15.1. Прийняття рішень в умовах ризику	145
15.2. Імовірність збитку. Матриця втрат	150
16. ОРГАНІЗАЦІЯ ПРОЕКТІВ НА БАЗІ СТАНДАРТНИХ ПАКЕТІВ ТЕХНІЧНОГО АНАЛІЗУ	157
16.1. Загальний підхід до організації самостійних проектів інтелектуального обробки даних (ІОД)	157
16.2. Макроси STATISTICA Visual Basic	162
17. ПРАКТИКА ВИКОНАННЯ ПРОЕКТІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ НА БАЗІ ПАКЕТУ STATISTICA NEURAL NETWORKS (SNN)	167
17.1. Методика організації самостійних проектів інтелектуальної обробки даних (ІОД)	167
17.2. Приклади побудови проектів вирішення завдань	173
18. НЕЙРОКОТРОЛЕРИ	177
18.1. Системи автоматичного управління та ПІД-контролери	177
18.2. Схеми нейроуправління	179
18.3. Алгоритми навчання нейроконтролера	181
ПЕРЕЛІК ПОСИЛАНЬ	186



ВСТУП

Інтелектуальний аналіз даних – процес виявлення в сирих даних раніше невідомих, нетривіальних, практично корисних та доступних інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності, який підрозділяється на задачі класифікації, моделювання та прогнозування, розпізнавання та ін.

Характерною рисою останніх тенденцій комп'ютерного аналізу та інтерпретації даних й прийняття рішень є бурхливий розвиток технологій та засобів “видобутку”, “вилучення” знань з даних, “інтелектуального аналізу даних” (ІАД) . Для початкової стадії розвитку цього напрямку, що природно, характерна наявність невизначених понять даної предметної області.

Навчальний посібник з дисципліни «Інтелектуальний аналіз даних» для студентів денної та заочної форми навчання: 6.050101 – «Комп'ютерні науки» включає матеріал, який відповідає послідовності вивчення інтелектуального аналізу даних, а також спрямований на засвоєння студентами теоретичного та практичного матеріалу в області інтелектуального, регресійного, кластерного аналізу даних. У якості програмної реалізації обрано систему STATISTICA Neural Network, яка надає можливість проведення нейромережевих досліджень.

Мета даної роботи: формування у студентів знань, умінь, й навиків логічного та творчого мислення, необхідних для вирішення теоретичних та практичних завдань, оволодіння основних методів дослідження та розв'язування практичних завдань управління складними об'єктами на основі інтелектуальних інформаційних технологій, необхідних для здобуття професійних знань.

ПЕРЕЛІК ПОСИЛАНЬ

1. Аджиев В. MiniSet – визуальный инструмент аналитика // Открытые системы, 1997. – № 3. – С. 73 – 77.
2. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах. // Открытые системы, 1997. - №4. - С. 41-44.
3. Рузайкин Г.И. Орудие Data Mining: успех в анализе данных // Мир ПК, 1997. – № 1. – С. 102-103.
4. Шапот М. Интеллектуальный анализ данных в системах поддержки принятия решений // Открытые системы, 1998. м № 1. – С. 30 – 35.
5. Шапот М., Рощупкина В. Интеллектуальный анализ данных и управление процессами // Открытые системы, 1998. – № 4,5.
6. Губарев В.В. Информатика в рисунках и таблицах (фрагменты системного путеводителя по концептуальным основам) – Новосибирск: Изд-во НГТУ, 1998. – 152 с.
7. Gubarev V.V. Experimental Data Analysis in the Systems Context // Proceedings The Third Russian-Korean International Symposium on Science and Technology. - Novosibirsk: NSTU, 1999. – Vol.1. – P.241 – 244; // Abstracts. – Vol.1. – P.190.
8. Головкин М. Жизнь в мусорной куче, или нужны ли нам знания? // Computerworld Россия. – 17.08.1999. – С.41 – 43.
9. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во ИМ СО РАН, 1999. – 270 с.
10. Туманов В. Data Warehouse: с чего начать? // PC WEEK, 1999. - № 29. – С. 15 – 16.
11. Губарев В.В. Вероятностные модели: Справочник. В 2-х ч. – Новосибирск: НЭТИ, 1992. – 422 с.



12. Рузайкин Г.И. Орудие Data Mining: успех в анализе данных // Мир ПК, 1997. – № 1. – С. 102 – 103.
13. Аркадьев А. Г., Браверман Э. М. Обучение машины классификации объектов. – М., 1971. – 192 с.
14. Горелик А.Л., Гуревич И.Б., Скрипкин В.А. Современное состояние проблемы распознавания. М.: Радио и связь, 1985, 160 с.
15. Бонгард М.М. Проблема узнавания. – М.: Наука, 1967.- 320 с.
16. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. М.: Наука, 1974. – 415 с.
17. Дуда Р., Харт П. Распознавание образов и анализ сцен. – М.: Мир, 1976. – 511 с.
18. А.Головко, под ред. проф. А.И.Галушкина "Нейронные сети: обучение, организация и применение", ИПРЖР, Москва 2001
19. С.Короткий "Нейронные сети" – Режим доступа до інформації:
<http://lii.newmail.ru/>
20. Л.Г.Комарцова, А.В. Максимов "Нейрокомпьютеры" Москва, МГТУ им. Н.Э.Баумана, 2002
21. Хайкин С. Нейронные сети: полный курс, 2-е издание. Пер. с англ. – М.: Издательский дом «Вильямс», 2006 г. – 1104с.
22. Долгополов Н.В., Яблоков М.Ю. "Электронный нос" – новое направление индустрии безопасности. – Мир и безопасность. 2007, № 4, с. 54–59.
23. Боровиков В.П. STATISTICA Neural Networks – Техническое описание. М.: Мир, 1999. – 239 с.
24. Уоссермен Ф. Нейрокомпьютерная техника / Уоссермен Ф. – М.: Мир, 1992. – 240с..
25. Розенблатт Ф. Принципы нейродинамики / Розенблатт Ф. – М.: Мир, 1965. – 480 с.

