การทำเหมืองข้อมูล เล่ม 1
การค้นหาความรู้จากข้อมูล

Data Mining 1 :
Discovering Knowledge in Data

พิมพ์ครั้งที่ 2 ฉบับปรับปรุง

ข้อสอบ 8 พ.ศ.

DM DM DM DM DM

สำหรับ

สถิติ คณิตศาสตร์ การวิจัยการดำเนินงาน
คอมพิวเตอร์ สารสนเทศ วิศวกรรมคอมพิวเตอร์
วิศวกรรมสารสนเทศ ครุศาสตร์อุตสาหกรรม
บริหารธุรกิจ เศรษฐศาสตร์ สังคมศาสตร์
มนุษยศาสตร์ ศึกษาศาสตร์ วิจัยการศึกษา
แพทยศาสตร์ เภสัชศาสตร์ เกษตรศาสตร์

เนื้อหาประกอบด้วย

- บทนำเกี่ยวกับการทำเหมืองข้อมูล
- การจัดเตรียมข้อมูลก่อนประมวลผลข้อมูล
- การวิเคราะห์ข้อมูลโดยการสำรวจ
- วิธีการเชิงสถิติในการประมาณค่าและการทำนาย
- ความใกล้เคียงกันมากที่สุด
- ต้นไม้ตัดสินใจ
- โครงข่ายประสาทเทียม
- การจัดกลุ่ม
- โครงข่ายโคโฮเนน
- กฎความสัมพันธ์
- วิธีการประเมินตัวแบบ
- การทำเหมืองข้อมูลบนเวป
- โปรแกรมในการทำเหมืองข้อมูล

BEST SELLER

รศ.สายชล สินสมบูรณ์ทอง

ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

สารบัญ

	หน้า
บทที่ 1 บทนำเกี่ยวกับการทำเหมืองข้อมูล	1
1.1 การทำเหมืองข้อมูลคืออะไร	1
1.2 ทำไมจึงต้องใช้การทำเหมืองข้อมูล	3
1.3 วิวัฒนาการในการทำเหมืองข้อมูล	4
1.4 ความจำเป็นของมนุษย์ในการทำเหมืองข้อมูล	5
1.5 ปัจจัยที่ทำให้การทำเหมืองข้อมูลได้รับความนิยม	5
1.6 ประเภทข้อมูลที่สามารถทำเหมืองข้อมูล	6
1.7 ลักษณะเฉพาะของข้อมูลที่สามารถทำเหมืองข้อมูล	6
1.8 ตัวอย่างแสดงผลสำเร็จของการนำการทำเหมืองข้อมูลไปใช้	7
1.9 ประเภทของการทำเหมืองข้อมูล	7
1.10 กระบวนการที่เป็นมาตรฐานในอุตสาหกรรม	8
1.11 ความผิดพลาดของการทำเหมืองข้อมูล	11
1.12 งานที่ทำให้การทำเหมืองข้อมูลประสบความสำเร็จ	12
แบบฝึกหัดที่ 1	19
เฉลยแบบฝึกหัดที่ 1	23
บทที่ 2 การจัดเตรียมข้อมูลก่อนประมวลผลข้อมูล	29
2.1 ความจำเป็นของการจัดเตรียมข้อมูลก่อนประมวลผลข้อมูล	29
2.2 การกลั่นกรองข้อมูล	30
2.3 การจัดการกับข้อมูลสูญหาย	32
2.4 การจำแนกผิดพลาด	36
2.5 วิธีการใช้กราฟสำหรับหาค่าผิดปกติ	36
2.6 การแปลงข้อมูลหรือการปรับเปลี่ยนรูปแบบข้อมูล	38
2.7 วิธีการเชิงตัวเลขสำหรับหาค่าผิดปกติ	43
แบบฝึกหัดที่ 2	45
เฉลยแบบฝึกหัดที่ 2	49

สารบัญ (ต่อ)

	หน้า
บทที่ 3 การวิเคราะห์ข้อมูลโดยการสำรวจ	59
3.1 การทดสอบสมมติฐานและการวิเคราะห์ข้อมูลโดยการสำรวจ	59
3.2 การเข้าใจถึงความรู้ที่มีอยู่ในข้อมูล	60
3.3 การจัดการกับตัวแปรที่มีสหสัมพันธ์กัน	62
3.4 การสำรวจข้อมูลโดยพิจารณาจากตัวแปรเชิงกลุ่ม	64
3.5 การใช้การวิเคราะห์ข้อมูลโดยการสำรวจเพื่อหาขอบเขตข้อมูลที่ผิดปกติ	71
3.6 การสำรวจข้อมูลโดยพิจารณาจากตัวแปรเชิงตัวเลข	72
3.7 การสำรวจข้อมูลโดยพิจารณาจากความสัมพันธ์ระหว่างตัวแปรหลายตัว	82
3.8 การเลือกเซตย่อยของข้อมูลที่น่าสนใจสำหรับการตรวจสอบเพิ่มเติม	84
3.9 การแบ่งกลุ่มข้อมูล	86
3.10 การสรุปผล	87
แบบฝึกหัดที่ 3	89
เฉลยแบบฝึกหัดที่ 3	95
บทที่ 4 วิธีการเชิงสถิติในการประมาณค่าและการทำนาย	101
4.1 งานในการทำเหมืองข้อมูลในการค้นหาความรู้จากข้อมูล	101
4.2 วิธีการเชิงสถิติในการประมาณค่าและการทำนาย	102
4.3 วิธีการตรวจสอบด้วยตัวแปรตัวเดียว : มาตรวัดแนวโน้มเข้าสู่ส่วนกลางและการกระจาย	102
4.4 การอนุมานเชิงสถิติ	106
4.5 ความเชื่อมั่นในค่าประมาณที่ได้	108
4.6 การประมาณค่าช่วงความเชื่อมั่น	108
4.7 วิธีการตรวจสอบด้วยตัวแปรสองตัว : การถดถอยเชิงเส้นอย่างง่าย	110
4.8 ข้อควรระมัดระวังในการประมาณค่า	116
4.9 ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของ Y เมื่อกำหนดค่าของ X ให้	117
4.10 ช่วงการทำนายสำหรับค่าที่เลือกอย่างสุ่มของ Y เมื่อกำหนดค่าของ X ให้	118

สารบัญ (ต่อ)

หน้า

4.11 การถดถอยเชิงซ้อน 121
4.12 การตรวจสอบข้อสมมติเบื้องต้นของตัวแบบ 124
แบบฝึกหัดที่ 4 128
เฉลยแบบฝึกหัดที่ 4 135

บทที่ 5 ความใกล้เคียงกันมากที่สุด 139
5.1 วิธีการที่มีผู้สอนและวิธีการที่ไม่มีผู้สอน 139
5.2 ระเบียบวิธีการสร้างตัวแบบที่มีผู้สอน 140
5.3 ความไม่สอดคล้องกันระหว่างความเอนเอียงและความแปรปรวน 143
5.4 งานในการจำแนกกลุ่ม 145
5.5 อัลกอริทึมความใกล้เคียงกันมากที่สุด 147
5.6 ฟังก์ชันระยะห่าง 150
5.7 ฟังก์ชันการรวมกัน 153
5.8 การหาจำนวนเกี่ยวข้องกันในเชิงคุณลักษณะ : การขยายแกน 156
5.9 การพิจารณาจากฐานข้อมูล 158
5.10 อัลกอริทึมความใกล้เคียงกันมากที่สุดสำหรับการประมาณค่าและการทำนาย 158
5.11 การเลือกค่า k 160
แบบฝึกหัดที่ 5 161
เฉลยแบบฝึกหัดที่ 5 166

บทที่ 6 ต้นไม้ตัดสินใจ 171
6.1 อัลกอริทึมต้นไม้การจำแนกและการถดถอย 174
6.2 อัลกอริทึม C4.5 182
6.3 กฎการตัดสินใจ 191
6.4 การเปรียบเทียบของการประยุกต์ใช้อัลกอริทึม C5.0 และ CART 192
แบบฝึกหัดที่ 6 197
เฉลยแบบฝึกหัดที่ 6 203

สารบัญ (ต่อ)

หน้า

บทที่ 7 โครงข่ายประสาทเทียม 207

7.1 การลงรหัสข้อมูลเข้าและข้อมูลออกหรือผลลัพธ์ 208

7.2 โครงข่ายประสาทเทียมสำหรับการประมาณค่าและการทำนาย 210

7.3 ตัวอย่างของโครงข่ายประสาทเทียมอย่างง่าย 211

7.4 ฟังก์ชันกระตุ้นรูปตัว S 214

7.5 การแพร่แบบย้อนกลับ 215

7.6 วิธีลดองศา 216

7.7 กฎการแพร่แบบย้อนกลับ 217

7.8 ตัวอย่างของการแพร่แบบย้อนกลับ 218

7.9 เกณฑ์การหยุด 221

7.10 อัตราการเรียนรู้ 222

7.11 โมเมนตัม 223

7.12 การวิเคราะห์ความไว 225

7.13 การประยุกต์ใช้การสร้างตัวแบบโครงข่ายประสาทเทียม 226

แบบฝึกหัดที่ 7 229

เฉลยแบบฝึกหัดที่ 7 233

บทที่ 8 การจัดกลุ่ม 237

8.1 งานในการจัดกลุ่ม 237

8.2 วิธีการจัดกลุ่มแบบเป็นขั้นตอน 240

8.3 วิธีการจัดกลุ่มแบบไม่เป็นขั้นตอน 253

8.4 ตัวอย่างการจัดกลุ่มแบบเฉลี่ย k กลุ่ม 254

แบบฝึกหัดที่ 8 260

เฉลยแบบฝึกหัดที่ 8 266

สารบัญ (ต่อ)

หน้า

บทที่ 9 โครงข่ายโคโฮเนน 271

9.1 แผนที่จัดการตัวเอง 271

9.2 โครงข่ายโคโฮเนน 274

9.3 การตรวจสอบความถูกต้องของการจัดกลุ่ม 281

9.4 การประยุกต์ใช้การจัดกลุ่มโดยใช้โครงข่ายโคโฮเนน 282

แบบฝึกหัดที่ 9 289

เฉลยแบบฝึกหัดที่ 9 291

บทที่ 10 กฎความสัมพันธ์ 293

10.1 คำนำ 293

10.2 ลักษณะข้อมูลที่ต้องการของการหากฎความสัมพันธ์ 293

10.3 การวิเคราะห์ความสัมพันธ์และการวิเคราะห์ตะกร้าตลาด 293

10.4 ซัพพอร์ท ความเชื่อมั่น กลุ่มรายการสินค้าที่เกิดขึ้นบ่อยและคุณสมบัติ Apriori 297

10.5 อัลกอริทึม Apriori ทำงานอย่างไร (ส่วนที่ 1)? การสร้างกลุ่มรายการสินค้าที่เกิดขึ้นบ่อย 299

10.6 อัลกอริทึม Apriori ทำงานอย่างไร (ส่วนที่ 2)? การสร้างกฎความสัมพันธ์ 301

10.7 ส่วนขยายจากข้อมูลตรรกนีสำหรับข้อมูลเชิงกลุ่มโดยทั่วไป 305

10.8 วิธีการเชิงทฤษฎีในการใช้สารสนเทศ : วิธีการเหนี่ยวนำกฎโดยทั่วไป 306

10.9 กฎความสัมพันธ์ใช้เมื่อไร 311

10.10 กฎความสัมพันธ์แสดงการเรียนรู้แบบมีผู้สอนหรือการเรียนรู้แบบไม่มีผู้สอน 314

10.11 รูปแบบ local เทียบกับตัวแบบ global 315

แบบฝึกหัดที่ 10 317

เฉลยแบบฝึกหัดที่ 10 323

บทที่ 11 วิธีการประเมินตัวแบบ 329

11.1 วิธีการประเมินตัวแบบสำหรับงานในการพรรณนา 330

11.2 วิธีการประเมินตัวแบบสำหรับงานในการประมาณค่าและการทำนาย 330

สารบัญ (ต่อ)

	หน้า
11.3 วิธีการประเมินตัวแบบสำหรับงานในการจำแนกกลุ่ม	332
11.4 อัตราความคลาดเคลื่อน ความผิดพลาดเชิงบวก และความผิดพลาดเชิงลบ	332
11.5 การปรับค่าใช้จ่ายในการจำแนกผิดพลาด	335
11.6 การวิเคราะห์ค่าใช้จ่าย/ผลกำไรในการตัดสินใจ	337
11.7 แผนภูมิอกระดับและแผนภูมิผลกำไร	339
11.8 การประเมินตัวแบบด้วยการสร้างตัวแบบ	342
แบบฝึกหัดที่ 11	343
เฉลยแบบฝึกหัดที่ 11	345
บทที่ 12 การทำเหมืองข้อมูลบนเวป	347
12.1 คำนำ	347
12.2 การประยุกต์ใช้การทำเหมืองข้อมูลบนเวปทางด้านการบริการเวป	347
12.3 การประยุกต์ใช้การทำเหมืองข้อมูลบนเวปทางด้านการทำธุรกิจอีคอมเมิร์ซ	348
12.4 การประยุกต์ใช้การทำเหมืองข้อมูลบนเวปทางด้านการตลาดบนอินเตอร์เน็ต	349
แบบฝึกหัดที่ 12	350
เฉลยแบบฝึกหัดที่ 12	351
บทที่ 13 โปรแกรมในการทำเหมืองข้อมูล	353
13.1 ซอฟแวร์ในการทำเหมืองข้อมูล	353
13.2 การติดตั้งโปรแกรม Weka	354
13.3 การเริ่มการทำงานโปรแกรม Weka	355
13.4 ข้อมูลที่ใช้ในโปรแกรม Weka	359
13.5 การจำแนกกลุ่ม	385
13.6 การจัดกลุ่ม	510
13.7 กฎความสัมพันธ์	535
แบบฝึกหัดที่ 13	543
เฉลยแบบฝึกหัดที่ 13	545
บรรณานุกรม	548