

O'REILLY®

การพัฒนาแอปพลิเคชัน Machine Learning

เริ่มจากแนวคิดสู่ตัวผลิตภัณฑ์

เรียบเรียงด้วย
สำนวนไทย
อ่านเข้าใจง่าย



CORE
FUNCTION

Building Machine Learning
Powered Applications

Emmanuel Ameisen

สารบัญ

ส่วนที่ 1 ค้นหาแนวทางที่ถูกต้องสำหรับ ML	15
บทที่ 1 จากเป้าหมายสู่การกำหนดกรอบของ ML	19
ประเมินสิ่งที่เป็นไปได้	20
โมเดล (Model)	21
ข้อมูล	28
กรอบการทำงานของ ML Editor	31
พยายามทำทุกอย่างด้วย ML: เฟรมเวิร์คแบบ End-to-End	31
วิธีที่ง่ายที่สุด: ให้อยู่ในรูปของอัลกอริทึม	33
เรียนรู้จากประสบการณ์	34
สัมภาษณ์ Monica Rogati: วิธีเลือกและจัดลำดับความสำคัญของโครงการ ML	36
สรุป	38
บทที่ 2 การสร้างแผนงาน	41
การประเมินความสำเร็จ	41
ประสิทธิภาพทางธุรกิจ	43
ประสิทธิภาพของโมเดล	43
ความใหม่และการกระจายของข้อมูลที่เปลี่ยนไป	45
ความเร็ว	47
ประเมินขอบเขตและอุปสรรค	48
ใช้ประโยชน์จากความเชี่ยวชาญ	48
ต่อยอดจากสิ่งที่มีคนคิดไว้แล้ว	49
การวางแผนสำหรับ ML Editor	53
เริ่มต้น	53
เริ่มต้นด้วยโมเดลที่ง่ายก่อนเสมอ	54
เพื่อสร้างความก้าวหน้าอย่างสม่ำเสมอ: เริ่มสร้างแบบง่ายๆ	55
เริ่มต้นด้วยไปป์ไลน์ง่ายๆ	55

ไปป์ไลน์สำหรับ ML Editor	57
สรุป	59
ส่วนที่ 2 สร้างลำดับการทำงาน	61
บทที่ 3 เริ่มสร้างไปป์ไลน์	65
โครงสร้างที่ง่ายที่สุด	65
ต้นแบบของ ML Editor	67
การแยกอาร์กิวเมนต์และคลีนข้อมูล	67
การแบ่งข้อความออกเป็นคำ	68
การสร้างคุณสมบัติ	69
ทดสอบเวิร์คโฟลว์ของคุณ	72
ประสบการณ์ที่ผู้ใช้ได้รับ	72
ผลลัพธ์จากโมเดล	73
การประเมินต้นแบบของ ML Editor	74
โมเดล	75
ประสบการณ์ของผู้ใช้	76
สรุป	76
บทที่ 4 จัดหาชุดข้อมูลเริ่มต้น	79
การทำซ้ำๆ กับชุดข้อมูล	79
กระบวนการทางวิทยาศาสตร์ข้อมูล	80
สำรวจชุดข้อมูลแรกของคุณ	81
เริ่มจากจุดเล็กๆ แล้วขยับไปสู่ประสิทธิภาพ	81
ข้อมูลเชิงลึกเมื่อเทียบกับผลิตภัณฑ์	82
เกณฑ์การประเมินคุณภาพข้อมูล	82
ป้ายกำกับเพื่อค้นหาแนวโน้มข้อมูล	90
ค่าสถิติเชิงสรุป	91
สำรวจและติดป้ายกำกับอย่างมีประสิทธิภาพ	93
ทำให้เป็นอัลกอริทึม	109

แนวโน้มข้อมูล	111
ให้ข้อมูลกำหนดพีเจอร์และโมเดล	112
สร้างคุณลักษณะจากรูปแบบ	112
พีเจอร์สำหรับ ML Editor	115
Robert Munro: คุณค้นหา, ตัดป้ายกำกับ และใช้ประโยชน์จากข้อมูลได้อย่างไร?	116
สรุป	117
ส่วนที่ 3 กระบวนการทำซ้ำกับโมเดล	119
บทที่ 5 การฝึกและประเมินความสามารถของโมเดล	123
รูปแบบที่เหมาะสมของโมเดลที่ง่ายที่สุด	123
โมเดลแบบง่าย	123
จากรูปแบบข้อมูลไปสู่โมเดล	126
แยกชุดข้อมูลของคุณ	127
การแบ่งข้อมูลสำหรับ ML Editor	132
ประเมินประสิทธิภาพ	134
ประเมินโมเดลของคุณ: ดูมากกว่าค่าความแม่นยำ	137
เทียบเคียงข้อมูลกับการคาดการณ์	137
Confusion Matrix	137
เส้นโค้ง ROC	139
เส้นโค้งสำหรับการปรับแต่ง	142
การลดมิติสำหรับข้อผิดพลาด	143
วิธี Top-k	143
โมเดลอื่นๆ	148
ประเมินความสำคัญของพีเจอร์	148
ส่งตรงจากตัวจำแนกประเภท	149
ตัวอธิบายแบบ “กล่องดำ”	150
สรุป	153

บทที่ 6 แก้ไขปัญหา ML ของคุณ	155
แนวทางปฏิบัติที่ดีที่สุดสำหรับซอฟต์แวร์	155
แนวทางปฏิบัติที่ดีที่สุดสำหรับ ML โดยเฉพาะ	157
แก้ไขข้อบกพร่องในขั้นตอนเชื่อมต่อข้อมูล	158
เริ่มต้นด้วยข้อมูลเดียว	158
การทดสอบโค้ด ML	165
แก้ไขข้อบกพร่องในขั้นตอนฝึกโมเดล	170
ความยากของงาน	171
ปัญหาการเพิ่มประสิทธิภาพ	173
แก้ไขข้อบกพร่องในขั้นตอนใช้งาน	175
การรั่วไหลของข้อมูล	176
การเกิด Overfit	176
พิจารณาปัญหา	179
สรุป	180
บทที่ 7 การใช้ตัวจำแนกสำหรับให้คำแนะนำในการเขียน	183
คำแนะนำจากโมเดล	184
เราจะทำอะไรได้บ้างหากไม่มีโมเดล?	184
หาค่าความสำคัญของพีเจอร์โดยรวม	186
การใช้คะแนนของโมเดล	187
หาค่าความสำคัญของพีเจอร์เฉพาะที่	187
การเปรียบเทียบโมเดล	190
เวอร์ชัน 1: ใบบริยายาน	191
เวอร์ชัน 2: มีประสิทธิภาพมากขึ้น แต่ความไม่ชัดเจนก็เพิ่มขึ้น	191
เวอร์ชัน 3: ให้คำแนะนำที่เข้าใจได้	193
การสร้างคำแนะนำ	194
สรุป	200

ส่วนที่ 4 การนำไปใช้และตรวจดูการทำงาน	203
บทที่ 8 ข้อควรพิจารณาเมื่อนำโมเดลไปใช้งาน	207
ประเด็นเกี่ยวกับข้อมูล	207
ความเป็นเจ้าของข้อมูล	208
ความเอนเอียงของข้อมูล	208
ความเอนเอียงเชิงระบบ	210
ข้อควรคำนึงเกี่ยวกับการสร้างโมเดล	210
ลูปของ feedback	211
ประสิทธิภาพโดยรวมของโมเดล	212
พิจารณาบริบท	213
ฝ่ายตรงข้าม	213
การใช้งานแบบไม่ถูกต้อง	214
สนทนากับ Chris Harland	215
สรุป	216
บทที่ 9 ทางเลือกในการนำโมเดลไปใช้งาน	219
การติดตั้งบนเซิร์ฟเวอร์	219
แอปพลิเคชันแบบสตรีมมิง	220
การคาดการณ์แบบแบตช์	223
การติดตั้งบนฝั่งไคลเอนต์	224
โมเดลบนอุปกรณ์	226
ใช้งานบนบราวเซอร์	227
แนวทางแบบผสมผสาน	227
สรุป	228
บทที่ 10 สร้างมาตรการป้องกันให้โมเดล	231
ความล้มเหลวมีอยู่รอบตัว	231
การตรวจสอบอินพุตและเอาต์พุต	231
การถอยกลับเมื่อโมเดลล้มเหลว	236

การพัฒนาเพื่อประสิทธิภาพ	239
การปรับให้รองรับผู้ใช้หลายคน	239
การบริหารวงจรชีวิตของโมเดลและข้อมูล	242
การประมวลผลข้อมูลและ DAG	244
ขอความคิดเห็นจากผู้ใช้	245
เพิ่มขีดความสามารถให้นักวิทยาศาสตร์ข้อมูลในการนำโมเดลไปใช้	248
สรุป	249
บทที่ 11 การตรวจสอบและอัปเดตโมเดล	251
เฝ้าตามการทำงานของโมเดล	251
ตรวจสอบว่าถึงเวลาต้องอัปเดตหรือไม่	251
ตรวจสอบการใช้งานผิดประเภท	252
เลือกตัวชี้วัด	253
ตัวชี้วัดด้านประสิทธิภาพ	253
ตัวชี้วัดทางธุรกิจ	255
CI/CD สำหรับ ML	256
การทดสอบและการทดลองแบบ A/B	258
แนวทางอื่นๆ	260
สรุป	261