인공지능과 비즈니스

2020. 7. 18

윤 종 영

학기일정 (안)

W1 (7/4)	강좌소개 및 개요
W2 (7/11)	비즈니스 아이디어 구상과 발굴을 위한 문제 정리 및 분석
W3 (7/18)	사업 아이디어 타당성 검토 및 구체적 실행방안과 계획 수립
W4 (7/25)	User Feedback 분석 및 Prototyping 계획 수립
W5 (8/1)	프로젝트 중간발표

W6 ( )	시제품 제작 아이디어에 대한 피드백 및 수정
W7 ( )	Prototyping (시제품 제작 및 멘토링)
W8 (8/8)	Prototyping (시제품 제작 및 멘토링)
W9 ( )	시제품 제작 완료
W10 (8/22)	프로젝트 최종발표

Contents

일주일동안 어떻게 지내셨나요

Practical Artificial Intelligence

An Enterprise Playbook

Alan Pelz-Sharpe & Kashyap Kompella

Deep Publishing

Table of Contents

3. How It Works, Step-by-Step

Step 1: Problem Definition

- Describe how this same problem could be solved manually
    - To help surfacing existing domain knowledge and identify data requirements.

Step 2: Data Preparation

- Identify, format, clean, sample the data to make it ready for use.
- Get it ready for the machine learning model.

Step 3: Apply Machine Learning

- May need to try out several different algorithms and pick out a few that look promising.
- Split the data into a training and a test subset.

Step 4: Tune the Model

- Tune the data and algorithms
- May need to combine multiple models to get your optimized state

Step 5: Deploy the Model

- Consider scalability, load balancing, hardware requirements, security, and auditability.

Step 6: Monitor the Model

Building a Data Strategy

- AI is based on data: good quality data, and lots of it.

- Clean data is just as important as more data.

- Consider aggregating or even gradually getting rid of some existing data and features.

- Data is messy.

Challenges of Data

- Data Availability
  - Historically, we do not collect data with the intention of using it in machine learning.

- Data accessibility
  - Permission, Regulation, Security, Delays

- Data usability
  - Is there enough data?
  - How accurate is the data?
  - How well structured and labeled is it?
  - Volume vs. variety
  - Data may not be missing, but it may be messy.

- Data Understanding
  - Ongoing collaboration between machine learning and business analyst teams is important.

(You need to think about not just the data you need to kick the project off, but what you will require in the future to continue to feed the model.)

Data Best Practices

- Create a data inventory.
- Create a data pipeline.
- Ensure that the data you locate is actually available to you.
- Assess the usability of the data.
- Ensure that your machine learning experts, the data owners, and the business teams work together to ensure the data is understandable for machine learning work.

Deployment and management

Challenges

- Complexity

Best Practices

- Continuous integration
- Continuous deployment
- Micro-services and serverless capabilities
- Monitor system performance metrics (latency and error rates)
- Central registry for model discovery and re-use
- Automatic versioning of models

Algorithms and Infrastructure

- The most complex may not necessarily be the best.

- Review the academic literature to find out which algorithms achieved the best performance for the type of problems you're looking to solve.

- Ensemble methods
    - Combine predictions from multiple models.
    - "If you ask three AI systems the same question, what do you get?" —> Three different answers
    - Compare, contrast, and combine outputs from multiple methods (ensemble methods)

Human in the Loop

- Semi-supervised machine learning method

- Deployed in use cases where there is lots of inexpensive unlabeled data, but the cost of labeling is expensive. *(example?)*

- Underlying philosophy: Machines can learn more economically if they ask humans for some context.

- In theory, the use of HITL decrease over time. But things change in the real world over time, and the human remains in the loop for a much longer time.

- Human supervision of AI is the best practice.

- In some cases, supervision may even be mandated by law.

Key Points from this Chapter

1. Machine Learning systems follow a series of six steps: problem definition, data preparation, applying machine learning, tuning the model, deploying the model, and monitoring the model.
2. You need a data strategy with access to lots of high-quality data.
3. Data scientists spend the majority of their time cleaning and organizing data.
4. The data you need may not be easily accessible or usable.
5. Machine learning systems need to be fed data on an ongoing basis.

6. Create a data inventory and pipeline.
7. Ensure your experts, data owners, and business teams work together to ensure the data is usable and understandable for machine learning work.
8. Use best practices such as continuous integration and continuous deployment.
9. The biggest and most complex algorithm may not be the best for your needs.
10. AI systems tend to require ‘humans in the loop’ to resolve exceptions. You should plan for that.

Case Study & Project Mentoring

by Taek Young Kang

공지사항

- 수업일정 논의
    - 보강일정 (네이버, AI양재허브)
    - 최종발표 8/29?