Choosing AI techniques

a Fundamental Topic of Software Engineering for Artificial Intelligence

Outline

1. Making intelligence
2. Machine learning fundamentals
3. Machine learning approaches
4. Choosing the right approach
5. Common solution for common problems

Outline

1. Making intelligence
   a. What is Intelligence?
   b. Representing Intelligence?
   c. Creating Intelligence?
2. Machine learning fundamentals
3. Machine learning approaches
4. Choosing the right approach
5. Common solution for common problems

Making Intelligence

- intelligence = decision making component of a system
- maps context to a prediction
- context must be implemented in the intelligence runtime
- intelligence should be available to the creator
    - intelligent systems work best if the data comes from users

Making Intelligence

- intelligence can represent as:
  - program that tests multiple conditions
  - hand-labelling specific context and store it in lookup table
  - building models with ML
  - combination of the above
- criteria for a good representation
  - compact: can be loaded into the intelligence runtime
  - easy to load and execute in a runtime
  - safe for frequent updates, no system crashing bugs

Making Intelligence

- intelligence creating process is iterative
    1. Understanding the environment
    2. Defining success
    3. Getting data
    4. Getting ready to evaluate
    5. Trying simple heuristics
    6. Trying simple machine learning
    7. Assessing and iterating until you succeed

Outline

Machine learning fundamentals

TECHNISCHE UNIVERSITÄT DARMSTADT

Factors (among others) that intelligence creators must control :

- Feature Engineering
- Model structure complexity
- Model search complexity:
- Data Size

Machine learning fundamentals

- Feature Engineering:
  - Converting data to useful form.
  - Helping the model use the data.
  - Normalizing feature values.
  - Exposing hidden information.
  - Expanding the context.
  - Eliminating misleading things.

Machine learning fundamentals

TECHNISCHE UNIVERSITÄT DARMSTADT

Structure- & Search complexity:

- complex models to solve hard, big, open-ended problems.
- more complexity → higher chance the model misunderstands the concept.
    - It might underfit the concept, in that its understanding of the concept might too simple.
    - It might overfit the concept, in that its understanding of the concept is wrong; it happens to work in some contexts, but it just isn’t right.

Outline

1. Making intelligence
2. Machine learning fundamentals
3. Machine learning approaches
4. Choosing the right approach
5. Common solution for common problems

Machine learning approaches

TECHNISCHE UNIVERSITÄT DARMSTADT

- Start with a simple model.
- Try slightly refined versions of the model (informed by the training data).
- Check to see if the refined versions are better (using the training data).
- And iterate (roughly) until their search procedure can’t find better models.

Machine learning approaches

TECHNISCHE UNIVERSITÄT DARMSTADT

- Supervised machine learning
  - models relationship between target prediction and input features (labeled data)
- Unsupervised machine learning
  - explores input without an explicit output (pattern detection & descriptive modeling)
- Semi-supervised learning
  - part of the data is labeled, model learns the structures to organize the data
- Reinforcement learning :
  - learning a task by trying to maximize the rewards in iterative fashion

Outline

1. Making intelligence
2. Machine learning fundamentals
3. Machine learning approaches
4. Choosing the right approach
   a. Set the parameters
   b. Choose the technique
5. Common solution for common problems

Choosing the right approach

1. Know your problem
   a. is it solvable?
   b. define if the problem is hard
   c. define success scenarios
   d. Software requirements (functional-, non-functional- and domain specific requirements)
2. Know your data
   a. Look at summary statistics and visualizations
   b. Visualize the data
   c. Look for main features

Choosing the right approach

3. Process the data
   a. preprocessing
   b. clean the data
   c. profile the data
   d. Aggregate if needed
4. Transform the data: data → features
   a. weight the important features
   b. eliminate misleading data samples
   c. normalize the data

Choosing the right approach

Tweak the complexity parameters:

- how fast should the training be?
- how fast should the prediction be?
- how accurate should the module predict?
- how flexible should the model be?
- how much memory /computing power is available?
- how much training data is available?
- how many useful features are available in the data?

Choosing the right approach

tradeoffs

- **accuracy vs decision speed**
- **simple vs flexible**
- **training set size vs required memory**
- **training set size vs training time**
- **variance vs bias**
- **scalability vs # of Features**
- **interpretability vs accuracy**
- **# of Features vs prediction time**
- ...

Outline

1. Making intelligence
2. Machine learning fundamentals
3. Machine learning approaches
4. Choosing the right approach
    a. Set the parameters
    b. Choose the technique
5. Common solution for common problems

Choosing the right approach

- Categorize the problem by input
  a. labeled data → supervised learning
  b. unlabeled data, want to find structure → unsupervised learning
  c. labeling all data too expensive, labels serve as boundary → semi-supervised
  d. optimize a function by interacting with environment → reinforcement learning

Choosing the right approach

- Categorize problem by output
  a. output is a pattern → data correlation
  b. output is a number → regression
  c. output is a class → classification
  d. output is a set of groups → clustering
  e. output is an anomaly → anomaly detection

Choosing the right approach

- Find the right algorithm:
  1. define problem
  2. choose the right set of data
  3. set complexity parameters, in regards to trade-offs
  4. define suitable candidates (as we will see in example in the next slides)
  5. implement and analyze in regards to one another
     - ⟳ if we didn't achieve at least one success scenario go back to 2 (lack of data) or 3 (tweak complexity parameters)
  6. choose best performing one(s)
- many different algorithms for different problems
- no one-size-fits-all approach

Outline

1. Making intelligence
2. Machine learning fundamentals
3. Machine learning approaches
4. Choosing the right approach
5. Common solution for common problems
   a. Supervised algorithms
   b. Unsupervised algorithms
   c. Reinforcement learning

Supervised machine learning

Algorithm	How does it work?	Exemplary use	Advantages	Disadvantages
Linear Regression	- approximate input using polynomial regression - continuous output	understand product sales drivers. optimize price points.	- fast to implement and train - prone to overfit, but solvable with L1 or L2 reduction - performs well on linearly separable data	- can suffer from outliners - problematic with nonlinear functions
Logistic Regression	- approximate input using logistic regression - discrete output	Classify customers based on how likely they are to repay the loan.	- fast to implement and train - transparent - updates weight gradually over examples → no need to retrain - gives probabilistic output	- needs a lot of data(for supervised) - assumption of linearity in the logit - suffers from outliners
Naive Bayes	- build a classifier - Bayes Assumption: value is independent of values of other features	a framework for hiring new employees	- easy and fast to train and implement - memory efficient - easy to understand - takes prior knowledge into account	- strong feature independence - fails classifying rare occurrences - irrelevant features

Supervised machine learning

Algorithm	How does it work?	Exemplary use	Advantages	Disadvantages
Random Forest	- ensemble method for classification, regression, ao. - constructs multiple decision trees in runtime	Object detection	- can work in parallel - no need to tweek parameters - can handle missing values - seldom overfits - easy to use	- biased in multiclass problems - difficult to interpret - weaker on regression when dealing with extrema
SVN	- used for regression and classification - find hyperplane in a N-dimensional space with maximum margin	Predict how likely someone is to click on an online ad	- relatively memory efficient - good at approximating complex functions - works well for clear separation margin between classes	- difficult to interpret - not suited for large datasets - suffers from noise, overlapping classes

Unsupervised machine learning

Algorithm	How does it work?	Exemplary use	Advantages	Disadvantages
K-means clustering	- group similar data points and discover patterns - look at k clusters - starts with random centroid and optimizes from it	Computer vision Spell checking problems	- efficient O(K*n*d) - tight clusters - easy to implement	- order of data and seed impact result - scale sensitive - difficult to predict number of clusters
Hierarchical clustering	- Agglomerative: compute proximity matrix → each point is a cluster → merge two closest clusters → repeat - Divisive: opposite, rarely used	Charting Evolution through Phylogenetic Trees Clustering politicians by Twitter posts	- easy to implement - hierarchical → easy to decide on number of clusters (dendrogram) - flexible	- inefficient O((n^3)*d) - very sensitive to outliers - can’t handle big data - can’t step back during procedure

Unsupervised machine learning

TECHNISCHE UNIVERSITÄT DARMSTADT

Algorithm	How does it work?	Exemplary use	Advantages	Disadvantages
SVD	- factorizes a matrix to 3 matrices - reduce R rank matrix to K rank matrix	Recommender systems	- restructure data to: - reduce overfitting - improve performance - remove correlated features	- difficult to interpret - information loss - difficult to understand restructuring
Gaussian mixture model	- cluster according to a different Gaussian distribution - each data point approximated using a gaussian distribution using EM	feature extraction from speech data object tracking in video data	- fastest algorithm for learning mixture models - maximizes only the likelihood, it will not bias the means towards zero	- will always use all the components it has access to

Reinforcement learning

Algorithm	How does it work?	Exemplary use	Advantages	Disadvantages
Dynamic Programming	- compute each value in a state by using values of surrounding states - once done move to next state - repeat until changes are less than defined threshold	Optimize trading strategy for an options-trading portfolio Stock and pick inventory using robots Optimize the driving behavior of self-driving cars Optimize pricing in real time for an online auction of a product with limited supply	- efficient, finds policy in polynomial time - guaranteed to find an optimal policy	- not suitable for large problems - requires knowledge of transition probability matrix
Temporal Difference (TD)	- choose an action A on State S with X% according to policy and Y% random - calculate S' and reward - compute average at V(s) - update S to S'		- does not need to know the transition probability matrix - due to incremental updates no need to wait for the end of an episode	- problem with local minima - faster but unstable - might converge to wrong solution if model starts wrong
Neural Networks	- simulates human brain - takes in the weight of connections between neurons - trains weights based on automatically generated character. of processed samples		- Can approximate any nonlinear function - robust to outliers - can work with only support vectors	- easy to overfit - difficult to setup and tune - difficult to interpret - very ressource and memory intensive

Sources

[1] **Chapters 16,17,18 and 20** of Hulten, Geoff. Building Intelligent Systems: A Guide to Machine Learning Engineering. Apress. 2018.

[2] Machine Learning —Fundamentals

[3] An Executive's Guide to Artificial Intelligence

[4] Choosing the Right Algorithm for Machine Learning

[5] Software Engineering | Classification of Software Requirements

[6] 10 Machine Learning Methods that Every Data Scientist Should Know

[7]Gaussian Mixture Model

[8]Three reasons to choose the right Machine Learning algorithm

Sources

[9] How to choose the right Machine Learning Algorithm?

[10] Summary of Tabular Methods in Reinforcement Learning

[11] Do you know how to choose the right machine learning algorithm among 7 different types?

[12]An easy guide to choose the right Machine Learning algorithm

[13] Foundations of Machine Learning : Singular Value Decomposition (SVD)

[14] Choosing the Right Machine Learning Algorithm(Hackernoon)

[15] Pros and Cons of K-Means Clustering

[16]Understanding the concept of Hierarchical clustering Technique

Discussion

Acknowledgements & License

-