DD2434 Machine Learning, Advanced Course

Exercise 06: State-of-the-Art in Machine Learning

Hedvig Kjellström
hedvig@kth.se
https://www.kth.se/social/course/DD2434/

Trend 1

Expert-Designed ➔ Example-Based

NIPS 2015

Largest ML conference

Trend 2015: Large interest from industry: 10% increase in papers, 60% increase in attendants

Traditionally

Design a model that explains the world (using expert knowledge)

Fit to data

Example 1: Language translation

Build grammar in Language 1, fit paragraph to be translated to this grammar

Build grammar in Language 2, with connections from Language 1 to 2

Map grammatic model of paragraph to Language 2

Since a language is quite contained … Voila!

Example 2: Reconstruction of articulated 3D motion

E.g. the Microsoft human tracker [Shotton et al. CVPR 2011]

Build 3D model of human, fit this model to an RGB-D video sequence (using Random Forests), infer model parameters = joint angles and limb 3D positions

Well known structure … Voila!

Turn-of-the-century:

A lot more data available!

Why bother with this model building?

Instead, go example-based (essentially, do Nearest Neighbor search)

Techniques for optimization/search in chronological order:

- Inverted index (search engines)
- Approximative Nearest Neighbor
- Deep Neural Networks

Example 1: Language translation

Collect tons of examples of corresponding statements in Language 1 and Language 2, and build an enormous inverted index (*or, we do not exactly know – maybe a Deep Neural Network (DNN) these days?)*

For a new statement in Language 1, find the nearest neighbors (or closest matches if you think about it in a search engine way)

Return the corresponding utterances in Language 2

With enough data, good when language does not follow grammar/dictionary … Voila!

Example 2: Reconstruction of articulated 3D motion

E.g. *Deep Visual Analogy-Making* [Reed et al. NIPS 2015]

[let us have a look]

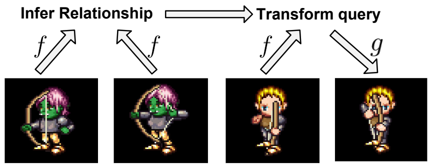


Figure 1: Visual analogy making concept. We learn an encoder function f mapping images into a space in which analogies can be performed, and a decoder g mapping back to the image space.

At least for this class of virtual image data, might be hard to scale … Voila!

Discussion point

The benefits of the data driven approach is that it is really flexible and learns everything about the mapping from input (data) to output (label)

Discuss with your neighbor (5 min):
What are the limitation of this approach?

Possible answers:

- Does not generalize well, need all kinds of possible data
- Huge search space, highly nonlinear and complex mapping, need many many parameters – **regularization**
- Need a lot of data – how get labeled data? (Reed synthesis, others crowd-sourcing)

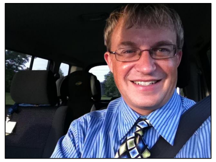
Trend 2

Have we reached the limit of purely data driven?
(*Yann LeCun et al: NO, Neil Lawrence et al: YES*)

Data driven way towards intelligence

“Visual Turing test”
E.g. *Visual Question Answering* [Antol et al. ICCV and NIPS-workshop 2015]

[let us have a look]



Does this man have children?	yes yes yes	yes yes yes
Is this man crying?	no no no	no yes yes

Well? Voila? Discuss with your neighbor what you think? Is this intelligence?

A great source of information on ML!

`https://nips.cc/Conferences/current`