Systems AI
The computational and mathematical modeling of complex AI systems

Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines
Kristian Kersting
Deep Neural Networks

Potentially much more powerful than shallow architectures, represent computations

Deep Neural Networks

Potentially much more powerful than shallow architectures, represent computations


The Moral Choice Machine

<table>
<thead>
<tr>
<th>Dos</th>
<th>WEAT</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>smile</td>
<td>0.116</td>
<td>0.348</td>
</tr>
<tr>
<td>sightsee</td>
<td>0.090</td>
<td>0.281</td>
</tr>
<tr>
<td>cheer</td>
<td>0.094</td>
<td>0.277</td>
</tr>
<tr>
<td>celebrate</td>
<td>0.114</td>
<td>0.264</td>
</tr>
<tr>
<td>picnic</td>
<td>0.093</td>
<td>0.260</td>
</tr>
<tr>
<td>snuggle</td>
<td>0.108</td>
<td>0.238</td>
</tr>
<tr>
<td>hug</td>
<td>0.115</td>
<td>0.233</td>
</tr>
<tr>
<td>brunch</td>
<td>0.103</td>
<td>0.225</td>
</tr>
<tr>
<td>gift</td>
<td>0.130</td>
<td>0.186</td>
</tr>
<tr>
<td>serenade</td>
<td>0.094</td>
<td>0.186</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Don’ts</th>
<th>WEAT</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>rot</td>
<td>-0.099</td>
<td>-1.118</td>
</tr>
<tr>
<td>negative</td>
<td>-0.101</td>
<td>-0.763</td>
</tr>
<tr>
<td>harm</td>
<td>-0.110</td>
<td>-0.730</td>
</tr>
<tr>
<td>damage</td>
<td>-0.105</td>
<td>-0.664</td>
</tr>
<tr>
<td>slander</td>
<td>-0.108</td>
<td>-0.600</td>
</tr>
<tr>
<td>slur</td>
<td>-0.109</td>
<td>-0.569</td>
</tr>
<tr>
<td>contaminate</td>
<td>-0.102</td>
<td>-0.544</td>
</tr>
<tr>
<td>brutalise</td>
<td>-0.118</td>
<td>-0.529</td>
</tr>
<tr>
<td>poison</td>
<td>-0.131</td>
<td>-0.520</td>
</tr>
<tr>
<td>murder</td>
<td>-0.114</td>
<td>-0.515</td>
</tr>
</tbody>
</table>

[Jentzsch, Schramowski, Rothkopf, Kersting 2018]
Deep neural networks may not be faithful probabilistic models.
Can we borrow ideas from deep learning for probabilistic graphical models?

Judea Pearl, UCLA
Turing Award 2012
Deep Probabilistic Modelling using Sum-Product Networks

Computational graph (kind of TensorFlow graphs) that encodes how to compute probabilities

Inference is Linear in Size of Network
SPFlow: An Easy and Extensible Library for Sum-Product Networks

[Molina, Vergari, Stelzner, Peharz, Poupart, Di Mauro Kersting 2018]

https://github.com/SPFlow/SPFlow

Compile SPNs into TF and PyTorch and even into flat, library-free code even suitable for running on devices: C/C++, GPU, FPGA [Sommer et al ICDD 2018]
Random sum-product networks

[Peharz, Vergari, Molina, Stelzner, Trapp, Kersting, Ghahramani UDL@UAI 2018]
We can even automatically discover the statistical types and parametric forms of the variables.
That is, the machine understands the data with few expert input ...

Exploring the Titanic dataset

This report describes the dataset Titanic and contains general statistical information and an analysis on the influence of different features and subgroups of the data have on each other. The first part of the report contains general statistical information about the dataset and an analysis of the variables and probability distributions. The second part focuses on a subgroup analysis of the data. Different clusters identified by the network are analyzed and compared to give an insight into the structure of the data. Finally, the influence different variables have on the predictive capabilities of the model are analyzed. The whole report is generated by fitting a sum product network to the data and extracting all information from this model.

...and can compile data reports automatically
P( heart attack | EHR )?
Statistical Relational AI: Logic, Probability, and Computation (or Bibel meets Bayes)


- building general-purpose thinking and learning machines
- make the AI/ML expert more effective
- increases the number of people who can successfully build AI/ML applications

P(heart attack)?
Probabilistic Models of EHRs

Atherosclerosis is the cause of the majority of Acute Myocardial Infarctions (heart attacks)

**Logical Variables (Abstraction)**

**Rule/Database view**

**Probability**

---

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Likelihood</th>
<th>AUC-ROC</th>
<th>AUC-PR</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boosting</td>
<td>0.81</td>
<td>0.96</td>
<td>0.93</td>
<td>9s</td>
</tr>
<tr>
<td>LSM</td>
<td>0.73</td>
<td>0.54</td>
<td>0.62</td>
<td>93 hrs</td>
</tr>
</tbody>
</table>

The higher, the better

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
<th>AUC-ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>J48</td>
<td>0.667</td>
<td>0.607</td>
</tr>
<tr>
<td>SVM</td>
<td>0.667</td>
<td>0.5</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.667</td>
<td>0.608</td>
</tr>
<tr>
<td>Bagging</td>
<td>0.677</td>
<td>0.613</td>
</tr>
<tr>
<td>NB</td>
<td>0.75</td>
<td>0.653</td>
</tr>
<tr>
<td>RPT</td>
<td>0.667*</td>
<td>0.778</td>
</tr>
<tr>
<td>RFGB</td>
<td>0.667*</td>
<td>0.819</td>
</tr>
</tbody>
</table>

The higher, the better

25%

---


[Kersting, Driessens ICML ´08; Karwath, Kersting, Landwehr ICDM ´08; Natarajan, Joshi, Tadepelli, Kersting, Shavlik. IJCAI ´11; Natarajan, Kersting, Ip, Jacobs, Carr IAAI ´13; Yang, Kersting, Terry, Carr, Natarajan AIME ´15; Khot, Natarajan, Kersting, Shavlik ICDM ´13, MLJ ´12, MLJ ´15, Yang, Kersting, Natarajan BIBM ´17]
This establishes a novel “Deep AI”

Data and Feature Programming
(Un-)Structured Data Sources
External Databases

Statistical AI Knowledge Base
(data, weighted rules, loops and data structures)

Model Rules and Domain Knowledge
Data Programming
Machine Learning and AI Algorithms

Graph Kernels
Diffusion Processes
Random Walks
Decision Trees
Frequent Itemsets
SVMs
Graphical Models
Topic Models
Gaussian Processes
Autoencoder
Matrix and Tensor
Factorization
Reinforcement Learning
...

Inference Results

And connects well to other communities

Jim Gray  Turing Award 1998
“Automated Programming”

Mike Stonebraker  Turing Award 2014
“One size does not fit all”
... also Cognitive Science, the twin science of Artificial Intelligence

"How do we humans get so much from so little?" and by that I mean how do we acquire our understanding of the world given what is clearly by today's engineering standards so little data, so little time, and so little energy.

Josh Tenenbaum, MIT
“Bayesian Programming”

Lake, Salakhutdinov, Tenenbaum, Science 350 (6266), 1332-1338, 2015
Tenenbaum, Kemp, Griffiths, Goodman, Science 331 (6022), 1279-1285, 2011
Since we need **Systems AI**, the computational and mathematical modeling of complex AI systems.

The next breakthrough in AI may not just be a new AI algorithm...

...but may be in the ability to rapidly combine, deploy, and maintain existing algorithms.

Since we need Systems AI, the computational and mathematical modeling of complex AI systems...
There are strong investments into probabilistic programming. RelationalAI, Apple, Microsoft, and Uber are investing hundreds of millions of US dollars. And it appears in industrial-strength solvers such as CPLEX and GUROBI.
AI is a team sport!
Core AI at TU Darmstadt

Internationally leading at Statistical Relational AI, Natural Language Processing, Robot Learning, Computer Vision and Machine Learning in general.
Complementary Expertises at TUDA
TUDA Centre for Cognitive Science

“The Machines have to serve the humans.”
AI Strategy of the German government, 15 Nov. 2018
Systems AI
The computational and mathematical modeling of complex AI systems

Artificial Intelligence at TU Darmstadt
http://www.ai-da.tu-darmstadt.de/

#1 German and #2 European AI institution, according to csrankings.org