# Hybrid learning

2022年6月6日 8:58

我们所有的知识都开始于感性，然后进入到知性，最后以理性告终。没有比理性更高的东西了。——康德

## 一、what and why

1. What
Hybrid learning, also known as neuro-symbolic AI, is an advanced version of artificial intelligence that improves how a neural network arrives at a decision by adding classical rules-based (symbolic) AI to the process.
This hybrid approach **requires less training data** and makes it possible for humans to track how AI programming made decision
Hybrid learning is a better model for leveraging machine learning and human Expertise

2. Why hybrid learning
For achieving high accuracy, deep learning requires a large amount of data that is sometimes difficult, expensive, or impractical to obtain. Integrating human knowledge into machine learning can significantly reduce data requirement, increase reliability and robustness of machine learning, and build explainable machine learning systems.

[What Is Neuro-Symbolic AI And Why Are Researchers Gushing Over It (analyticsindiamag.com](What Is Neuro-Symbolic AI And Why Are Researchers Gushing Over It (analyticsindiamag.com)

3. Neurosymbolic AI: The 3rd Wave https://arxiv.org/pdf/2012.05876.pdf
• Knowledge should be grounded onto vector representations for efficient learning from data based on message passing in neural networks as an efficient computational model.
• Symbols should become available as a result of querying and knowledge extraction from trained networks, and offer a rich description language at an adequate level of abstraction, enabling infinite uses of finite means, but also compositional discrete reasoning at the symbolic level allowing for extrapolation beyond the data distribution.
• The combination of learning and reasoning should offer an important alternative to the problem of combinatorial reasoning by learning to reduce the number of effective combinations, thus producing simpler symbolic descriptions as part of the neurosymbolic cycle.

4. Future
While the complexities of tasks that neural networks can accomplish have reached a new high with GANs, neuro-symbolic AI gives hope in performing more complex tasks. By combining the best of two systems, it can create AI systems which require fewer data and demonstrate common sense, thereby accomplishing more complex tasks.
   - The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence https://arxiv.org/ftp/arxiv/papers/2002/2002.06177.pdf
     Gary Marcus， 在robust AI, 反对deep learning，这篇看不进去
   - A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence
     https://vossen.info/wp-content/uploads/2020/08/akata-2020-research.pdf
     This lack of alignment with human val ues is impacting us more frequently. Now that AI technologies affect our everyday lives at an ever-increasing pace, there is a greater need for AI systems that work synergistically with humans rather than ones that simply replace them
     It is better to view AI systems not as “thinking machines” but as cognitive prostheses that can help humans think and act better, 不是把AI当成一个会思考 的机器，而是把它当成识别的外挂，将帮助人们更好地思考和行动

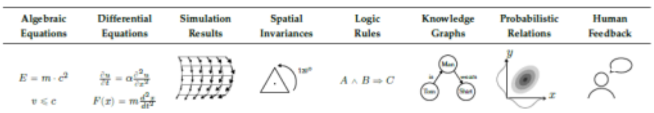
## 二、唱唱反调

5. 三位DL的先驱都不同意 hybrid learning，The future of deep learning, according to its pioneers – TechTalks (bdtechtalks.com)
6. The Bitter Lesson (incompleteideas.net)， 强化学习之父 Rich Sutton

## 三、Review

7. Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems
范老师説好文章， 一篇全面的综述， 可以作为Hybrid learning 文章的大索引
https://arxiv.org/pdf/1903.12394.pdf

将知识归为8种




8. The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems
https://arxiv.org/ftp/arxiv/papers/2105/2105.03354.pdf
归纳了人机协同系统的分类体系，有4个维度，从任务类型，目标，共享的数据表示，人入场的时机， 范老师说是 good paper for new year's gift, 我没看出好来。
杨洲说“这样的文章是给出一种*完备性*或是*系统性*的思考，它把已经有的工作(pieces)，放到一个未来的大厦当中，让人看到宏伟的建筑结构，那些已经做了，那些还没做，下一步应该做什么。有了这种思考，就知道AGI离我们还有多远，难度如何，我们应该如何做。”

## 四、方法

目前我们探索了三种改loss function的方法: semantic loss；GCN, LTN

9. Semantic loss 原理是约束，模型给出的结果如果满足约束LOSS小，反之Loss大
A Semantic Loss Function for Deep Learning with Symbolic Knowledge [A Semantic Loss Function for Deep Learning with Symbolic Knowledge (arxiv.org)]
   a. 主要提到WMC，将规则转换为可导的方法是论文的精华 A Differential Approach to Inference in Bayesian Networks, [765568.765570 (acm.org)]
   b. [SDD: A New Canonical Representation of Propositional Knowledge Bases (ijcai.org)]
   这里面提到谓词逻辑的形式化，以合取范式CNF为基础，提出一种形式化的逻辑规则表示方法
   合取范式： 任何命题公式，最终都能够化成 ( A 1 ∨ A 2 ) ∧ ( A 3 ∨ A 4 ) 的形式，这种先 ∨ 析取再 ∧ 合取的范式，被称为 “ 合取范式”
   c. 进一步计算SDD概率的方法 Probabilistic Sentential Decision Diagrams， http://reasoning.cs.ucla.edu/fetch.php?id=136&type=pdf
   算法已经开源 https://github.com/art-ai/pypsdd ， 自2018年后没有维护，存在BUGS， Mohsen已经fixed

   Semantic loss 公式
   The semantic loss L s (α, p) is a function of a sentence α in propositional logic, defined over variables X = {X1, . . . , Xn}, and a vector of probabilities p for the same variables X.

   **Definition 1** (Semantic Loss). Let p be a vector of probabilities, one for each variable in **X**, and let $\alpha$ be a sentence over **X**. The semantic loss between $\alpha$ and p is

$$L^s(\alpha, p) \propto -\log \sum_{\mathbf{x} \models \alpha} \prod_{i:\mathbf{x} \models X_i} p_i \prod_{i:\mathbf{x} \models \neg X_i} (1 - p_i).$$

$$\sum_{\mathbf{x} \models \alpha}$$

   方法是把逻辑规则转换为结果对应的真值表， 其中 x, 是α 的真值表的一行， 任意一种满足真值表的输出结果，都认为是符合约束条件的。 这个LOSS的含义就是让所有符合约束条件的输出结果的概率和尽可能大， 约束规则是对分类结果的约束
   比如 5分类问题， 逻辑规则转换后的真值表有两个，011000， 11100； 则Loss为 (1-p1)*p2*p3*(1-p4)*(1-p5) + p1*p2*p3*(1-p4)*(1-p5)

   [Guy Van den Broeck - News (ucla.edu)]
   AI can learn from data. But can it learn to reason [OSU (ucla.edu)]
   Bridging Data and Knowledge in Neuro-Symbolic Learning [Caltech summer school (ucla.edu)]
   pylon A PyTorch Framework for Learning with Constraints

   [TaxoClass: Hierarchical Multi-Label Text Classification Using Only Class Names (aclanthology.org)]

10. Graph knowledge
Multi-view Factorization AutoEncoder with Network Constraints for Multi-omic Integrative Analysis [1809.01772.pdf (arxiv.org)]
SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS [1609.02907.pdf (arxiv.org)]

11. A review of some techniques for inclusion of domain-knowledge into deep neural networks
https://www.nature.com/articles/s41598-021-04590-0.pdf
重点总结了科学研究中积累的逻辑规则和数值约束类型的知识
we are interested in implicit or explicit sources of domain-knowledge, represented either as logical or numeric constraints, and used at the model-construction stage by DNNs
   ▪ Transforming the input data

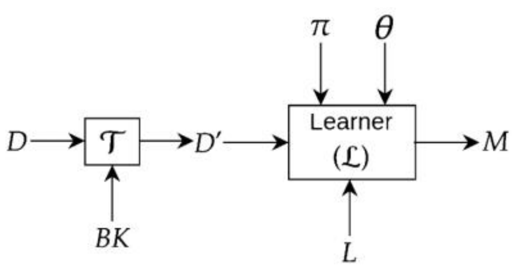
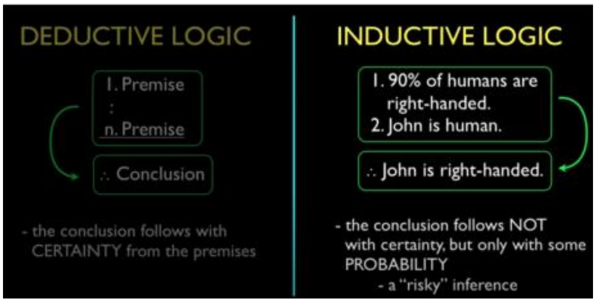



**Figure 6.** Introducing background knowledge into deep network by transforming data. $\mathcal{T}$ is a transformation block that takes input data $D$, background knowledge ($BK$) and outputs transformed data $D'$ that is then used to construct a deep model using a learner $\mathcal{L}$.

演绎逻辑和归纳逻辑




归纳逻辑程序设计：ILP算法以概念(如同源蛋白质)的示例E 和背景知识B (例如分子动力学的定义)为例，构建一个假设h，该假设以B 来解释E , 详细介绍见 周志华 机器学习西瓜书

2. Knowledge Infused Learning (K nowledge Infused Learning (K-IL): T -IL): Towards Deep Incorpor ds Deep Incorporation of Knowledge in Deep Learning

https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=1511&context=aii_fac_pub

Heterogeneous Embedding Space (HES)

即文本中的词嵌入向量和从知识图谱中实体的词嵌入向量，它们的向量空间不一致

3. Judgment Prediction via Injecting Legal Knowledge into Neural Networks

   知识为一阶逻辑规则，其中的一个要点是将一阶逻辑用fuzzy logic转化为连续值

   Fuzzy logic: https://www.jianshu.com/p/b316acff0f02

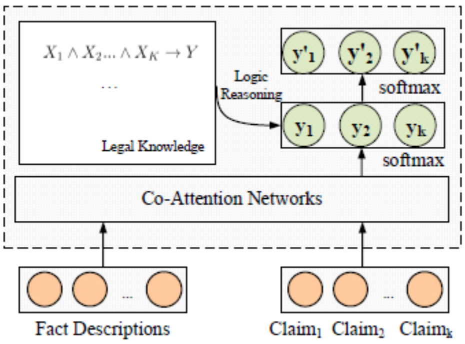
   


   Figure 2: The overall architecture.

   特色是用规则直接re-weight 模型的输出，然后再进行一次softmax

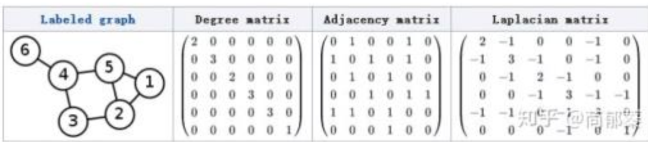
4. Embedding Symbolic Knowledge into Deep Networks

   把逻辑规则转换为图，再通过GCN把规则embeding ,其中用到规则求解器构建正反例

   这篇文章的Key point 在于把data 和 knowledge 投影到一个共同的空间，在那个共同的空间整合，训练模型

   图卷积入门

   https://zhuanlan.zhihu.com/p/71200936

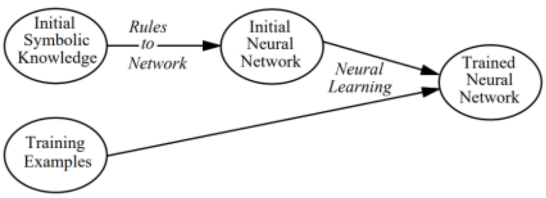
   


   有什么东西来度量节点的邻居节点这个关系呢，学过图论的就会自然而然的想到邻接矩阵和拉普拉斯矩阵。举个简单的例子，对于下图中的左图（为了简单起见，举了无向图且边没有权重的例子）而言，它的度矩阵 [公式]，邻接矩阵 [公式] 和拉普拉斯矩阵 [公式] 分别如下图所示，度矩阵 [公式] 只有对角线上有值，为对应节点的度，其余为0；邻接矩阵[公式]只有在有边连接的两个节点之间为1，其余地方为0；拉普拉斯矩阵 [公式] 为 [公式] 。但需要注意的是，这是最简单的一种拉普拉斯矩阵，除了这种定义，还有接下来介绍的几种拉普拉斯矩阵。

5. Knowledge-based artificial neural network

   94年的Paper，大概可以算做hybrid learning的鼻祖

   The translation of rules into neural structures

6. Multi-view Factorization AutoEncoder with Network Constraints for Multi-omic Integrative Analysis
   综述有介绍

7. Deep Neural Networks with Massive Learned Knowledge
   作者有CMU的邢波教授
   主要思想是将knowledge作为teacher模型，运用模型蒸馏，将教师知识transfer给学生
   创新主要在于它是老师与学生相互蒸馏
   教师模型的优化目标是

$$\min_{q \in \mathcal{P}} \mathrm{KL}(q(\boldsymbol{Y}) \| p_\theta(\boldsymbol{Y}|\boldsymbol{X})) - C \sum_l \lambda_l \mathbb{E}_q[f_l(\boldsymbol{X}, \boldsymbol{Y})], \quad (1)$$

   这个有点类似交叉熵与KL散度和A的信息熵的关系

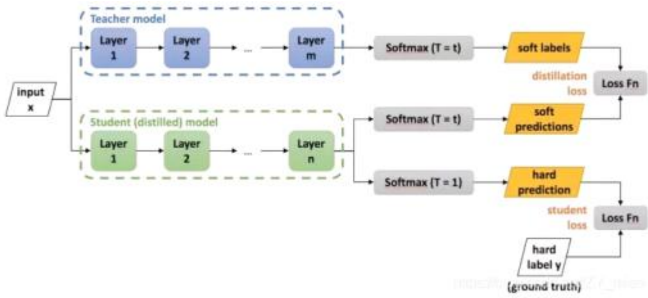
$$H(A, B) = D_{KL}(A \| B) + S_A$$

8. Distilling the Knowledge in a Neural Network
   我们可以训练非常复杂的模型，其易于从数据中提取出结构。这个复杂的模型可以是独自训练模型的集成，也可以是一个用强大正则器如dropout训练的单个大模型。一旦复杂模型训练完毕，之后我们可以使用一种不同的训练方式，称之为“蒸馏”，将知识从复杂的模型（称之为teacher模型）转移到更易于部署的小模型（称之为student模型）中

   Neural networks typically produce class probabilities by using a “softmax” output layer that converts the logit, $z_i$, computed for each class into a probability, $q_i$, by comparing $z_i$ with the other logits.

$$q_i = \frac{exp(z_i/T)}{\sum_j exp(z_j/T)} \quad (1)$$

   使用T的原理是为了增加分类的难度， 对于一个学的好的大网络，错误的概率要比正确的概率小很多很多， 也就是正确、错误的区别是明显的。 但小网络正确类与错误类没有这么明显的区分度。 所以就在小网络的softmax加一个T参数，加上这个T参数以后错误分类再经过softmax以后输出会变大,同样的正确分类会变小。这就人为的加大了训练的难度，一旦将T重新设置为1，分类结果会非常的接近于大网络的分类效果。
   如同你负重登山，虽然过程很辛苦，但是当有一天你取下负重，正常的登山的时候，你就会变得非常轻松，可以比别人登得高登得远。

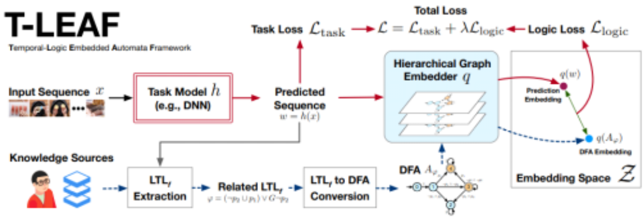



$$\mathcal{L} = \mathcal{L}_{hard} + \lambda \mathcal{L}_{soft}$$

$$cost\ function = CroEntropy(y_s, y_t) + \alpha CrossEntropy(y_s, y)$$

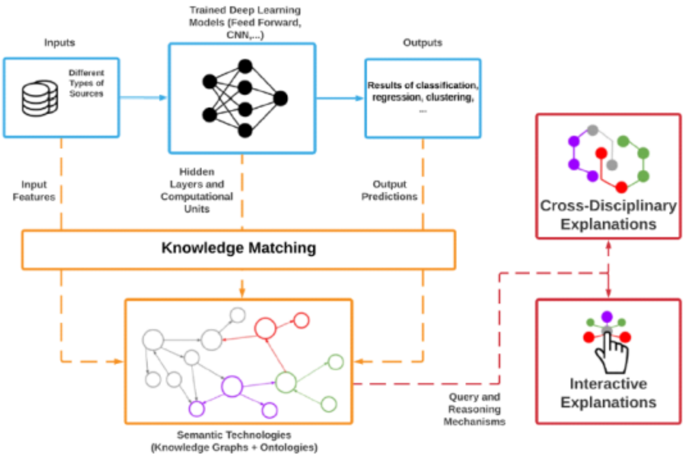
9. Logic Tensor Networks
   Mohen最初探索的rule-based hybrid learning，支持一级谓词逻辑

10. Embedding Symbolic Temporal Knowledge into Deep Sequential Models, [2101.11981.pdf (arxiv.org)](2101.11981.pdf (arxiv.org))
   Dr. Fan 説这篇文章可以帮助我们理解 hierarchical loss




   LTL, 线性时序逻辑， 时间逻辑（temporal logic）描述的是关于过去发生了什么，现在在发生什么，未来会发生什么的逻辑。 LTL可以转换为确定有限自动机（deterministic finite automaton, DFA），对DFA进行embedding，比较模型预测的状态序列和LTL定义的状态序列的编码的距离，构建损失函数; 图的embedding使用GCN

13. Explainable Artificial Intelligence (AI) system




14. 2023, 范老师推的最新hybrid AI 相关papers

A. https://hal.science/hal-03622260/document , Combining Data-Driven and Knowledge-Based AI Paradigms for Engineering AI-Based Safety-Critical Systems

B. PRBOOST: Prompt-Based Rule Discovery and Boosting for InteractiveWeakly-Supervised Learning

C. 智能计算的最新进展，挑战与未来， https://spj.science.org/doi/epdf/10.34133/icomputing.0006?adobe_mc=MCMID%3D22206250801421742214231095253443011171%7CMCORGID%3D242B6472541199F70A4C98A6%2540AdobeOrg%7CTS%3D1674912512

D. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence， 2002.06177.pdf (arxiv.org)， Gary Marcus的文章

E. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence , https://vossen.info/wp-content/uploads/2020/08/akata-2020-research.pdf