统计机器学习中参数可辨识性研究及其关键问题

冉智勇 [1] 胡包钢 [2]

摘 要 参数可辨识性研究在统计机器学习中具有重要的理论意义和应用价值. 参数可辨识性是关于模型参数能否被惟一确定的性质. 在包含物理参数的学习模型中, 可辨识性不仅是物理参数获得正确估计的前提条件, 更重要的是, 它反映了学习机器中由参数决定的物理特征. 为扩展到未来类人智能机器研究的考察视角, 我们将学习模型纳入 "知识与数据共同驱动模型" 的框架中讨论. 在此框架下, 我们提出两个关键问题. 第一是参数可辨识性准则问题. 该问题考察与可辨识性密切相关的各种判断准则, 其中知识驱动子模型与数据驱动子模型的耦合方式为参数可辨识性问题提供了新的研究空间. 第二是参数可辨识性与机器学习理论和应用相关联的研究. 该研究包括可辨识性对参数估计、模型选择、学习算法、学习动态过程、奇异学习理论、贝叶斯推断等内容的深刻影响.

关键词 可辨识性, 统计机器学习, 参数估计, 奇异学习理论, 贝叶斯推断

引用格式 冉智勇, 胡包钢. 统计机器学习中参数可辨识性研究及其关键问题. 自动化学报, 2017, **43**(10): 1677−1686

DOI 10.16383/j.aas.2017.c160720

Parameter Identifiability and Its Key Issues in Statistical Machine Learning

RAN Zhi-Yong[1] HU Bao-Gang[2]

Abstract The study of parameter identifiability has important theoretical meaning and practical value in statistical machine learning. Parameter identifiability is a property that concerns whether the model parameters can be uniquely determined. In learning models containing physical parameters, identifiability is a prerequisite for estimating those parameters; more importantly, it reflects the physical characteristic determined by those parameters. In order to extend our perspective to future human-like intelligent machines, we put the learning models into the framework of "knowledge- and data-driven models". Within this framework, we propose two key issues. The first is about identifiability criteria which aim to study various criteria closely related to identifiability; the coupling manner between knowledge-driven submodel and data-driven submodel provides novel topics for identifiability study. The second focuses on identifiability relevant to theory and application in machine learning; this involves the deep influence of identifiability on parameter estimation, model selection, learning algorithms, learning dynamics, Bayesian inference.

Key words Identifiability, statistical machine learning, parameter estimation, singular learning theory, Bayes inference

Citation Ran Zhi-Yong, Hu Bao-Gang. Parameter identifiability and its key issues in statistical machine learning. *Acta Automatica Sinica*, 2017, **43**(10): 1677−1686

"辨识" 概念起源于统计学, 并在控制科学中得到了巨大的发展. 根据文献 [1−2] 记载, 上世纪 20 年代, 一些经济学家在统计推断模型中首次提出了 "辨识" 概念及其相关问题. 从 1934 年到 1975 年, 一批经济学家对此做出了里程碑式的理论化奠基工作; 这其中包括多个诺贝尔经济学奖得主, 如 1969 年荷兰经济学家 Jan Tinbergen 与挪威经济学家 Ragnar A. K. Frisch, 1975 年美国经济学家 Tjalling Koopmans, 1989 年挪威经济学家 Trygve M. Haavelmo, 以及 2007 年美国经济学家 Leonid Hurwicz. 在此期间, 以 "辨识" 为标题的论文有 Koopmans 于 1949 年针对经济模型发表的文献 [3]; 而文献 [4] 认为 Haavelmo 是辨识问题中给出通用与准确数学公式化的首位研究者, 并明确指出它与估计问题的不同 ("Haavelmo was the first to give a general and precise mathematical formulation of the identification problem, and to distinguish it clearly from the estimation problem").

在此之后, 控制学界发展了基于控制理论的 "辨识" 研究. 开创性工作有 Zadeh 于 1956 年针对黑箱系统建模问题[5] 以及 1962 年针对 "电阻 – 电感 – 电容" 网络系统[6] 提出的辨识问题. 根据文献 [4] 评述, Zadeh 与 Kalman 提出的辨识问题更强调 (黑

收稿日期 2016-10-15 录用日期 2017-06-07
Manuscript received October 15, 2016; accepted June 7, 2017
国家自然科学基金 (61573348, 61620106003) 资助
Supported by National Natural Science Foundation of China (61573348, 61620106003)
本文责任编委 朱军
Recommended by Associate Editor ZHU Jun

1. 重庆邮电大学计算机科学与技术学院计算智能重庆市重点实验室 重庆 400065 2. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190
1. Chongqing Key Laboratory of Computational Intelligence, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065 2. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

箱) 模型的选择与估计 (“For Zadeh or Kalman, identification is the selection and estimation of a model”), 或者更侧重 “从样本到总体 (From sample to population)”. 这不同于当时已有经济模型中辨识问题更强调 “可辨识性 (Identifiability)”, 即更侧重 “从总体到结构 (From population to structure)”. 在发展出状态空间表示方法之后, 控制学科为辨识理论体系产生出了许多独有的学术贡献和系统性研究工作[7−12].

从辨识研究的发展史可以看出, “辨识” 是数学模型与控制系统中最核心和最基础的研究内容之一. 虽然学界认为线性系统和模型辨识研究已经成熟 (“System identification for linear systems and models is a well-established and mature topic)”, 然而, Ljung 指出辨识非线性模型是 “更加丰富且严苛的领域” (“Identifying nonlinear models is a much more rich and demanding problem area”)[13]. 辨识概念及其研究主题不仅持续扩展其应用领域, 而且已经滋润了其他学科的迅速发展. 本文正是在统计机器学习研究背景下进行关键问题讨论[14−15], 这不同于经典统计模型和控制模型[16−17]. 对于这样十分活跃且为当下最为 “显学” 的研究领域进行辨识主题的全面综述无疑是很大的挑战, 因此, 我们把范围限定为统计机器学习模型中的参数可辨识性. 参数可辨识性是关于模型参数能否被唯一确定的性质, 也是系统辨识理论研究中不可或缺的子问题.

机器学习研究本质涉及到多种数学空间的学习. 如果考察各个空间的关联 (图 1), 可以把机器学习模型视为一个参数学习机; 这也说明可辨识性是机器学习理论研究中的核心内容之一. 然而, 目前它并没有得到学术界的广泛重视和深入系统性研究. 本文从新视角定义的模型类别开始讨论该专题, 研究其中的两个关键问题, 特别强调对未来发展轨迹提出新的见解.

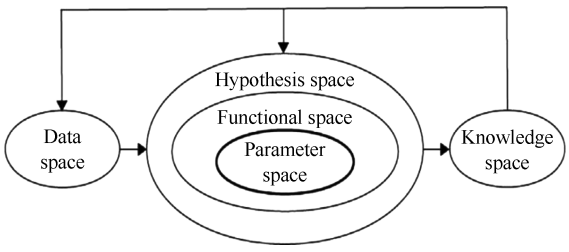

图 1 机器学习中各个空间关系示意图[14]

Fig. 1 The relationship between various spaces in machine learning[14]

我们主要讨论统计机器学习中有关参数可辨识性的关键问题. 所以, 我们把机器学习模型置于统计框架下. 按照 Amari 等信息几何的观点[18], 可以把参数化的统计分布族看作具有几何结构的统计流形, 每一个具体的统计分布被视为统计流形上的一个点. 因此, 可辨识性问题考虑的是: 对于给定的参数统计模型, 是否存在不同的参数值, 它们对应的统计分布是相同的.

在机器学习文献中, 与局部不可辨识等价的一个概念是奇异性 (Singularity), 如果一个统计模型的 Fisher 信息矩阵不是处处严格正定的, 我们称此模型为奇异学习机[19]. 奇异学习机在机器学习理论和应用中具有重要的地位[20−23]. 如果一个学习机具有层次结构[24−26]、隐变量[27−28]、状态变量[29−31]、讨厌参数[32]、语法规则[23]、耦合的子模型[15, 33] 等, 那么这个学习机通常是奇异的. 基于奇异学习机在机器学习领域普遍存在, Watanabe 指出: 几乎所有的学习机都是奇异的 (Almost all learning machines are singular)[22]. 奇异学习机包括多层感知器、径向基函数、高斯混合模型、玻尔兹曼机、Bayes 网络等.

在机器学习领域, 相对于各种各样具体的学习算法而言, 与可辨识性有关的研究则显得相对缺乏. 长期以来, 此问题并没有得到广泛的关注. 直到近几年, 随着日本学者 Amari、Watanabe 与 Fukumizu 等关于奇异学习理论的完善, 越来越多的学者开始关注这一理论; 这一点可以从近年来机器学习主流期刊和国际会议的文章数量看出来. 这些理论成果主要得益于两个方面: 1) Amari 的信息几何理论[18]; 2) Watanabe 基于代数几何 (Algebraic geometry) 和代数分析的奇异学习理论[23].

针对机器学习领域中的可辨识性研究, 本文以机器学习和神经计算为应用背景, 主要讨论参数可辨识性研究在统计机器学习中的两大关键问题, 简述其研究进展和研究难点, 并提出若干瓶颈问题. 第一个问题主要讨论与可辨识性准则相关的几个问题, 其中包括判断学习机器奇异的准则 (因为模型奇异和模型局部不可辨识是等价的[34]), 也包括判断参数全局可辨识的准则 (这对模型的可解释性和透明度有重要意义). 第二个问题主要讨论奇异性对机器学习各个方面的影响; 这种影响包括机器学习理论、参数估计、模型选择、学习算法、学习过程动态分析、Bayesian 推断等.

1 新视角下机器学习模型研究框架

对于可辨识性理论的研究, 目前学界有不同的研究视角和问题侧重. 研究视角的差异源于数学模型类型的不同, 或者由此限定的侧重范围, 比如基于非线性性质、静动态性质、统计性质、数据类型、应用领域范围等方式的传统划分. 在本文中, 我们试

图根据统计机器学习推理原理和极具发展前景的新模型 (见图 2) 视角进行考察. 目前, 多数机器学习模型应用 “归纳 (Induction)” 推理原理 (图 3); 这包括当下流行的深度学习模型. 由于该类模型依赖大量训练数据和很少的先验知识, 所以称为数据驱动模型. 当完全没有先验知识时, 模型参数可以任意给定并且不对应真实物理系统的物理意义, 因此又称 “非参模型 (Nonparametric model)” 或 “黑箱模型”. 因此, 使用者无法直接理解并解释黑箱模型的物理内涵和刺激–响应行为. 与此对应的另一类模型是基于 “演绎 (Deduction)” 推理原理的知识驱动模型; 它通常根据第一性原理 (First principle) 或实际系统机理知识而构造, 其中模型参数个数不仅固定, 而且具有明确的物理意义.

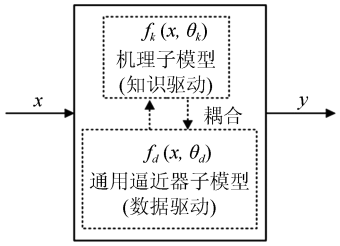

图 2　基于知识与数据共同驱动的机器学习模型 (其中, 两个子模型通过耦合算子互相联结[14−15])

Fig. 2　Knowledge- and data-driven machine learning model (within which two submodels are connected by a coupling operation[14−15])

本文讨论的模型对象是 “基于知识与数据共同驱动的模型”. 图 2 以简化的方式示意了该类模型的结构原理, 并用如下函数形式表示:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}, \theta) = \boldsymbol{f}_k(\boldsymbol{x}, \theta_k) \oplus \boldsymbol{f}_d(\boldsymbol{x}, \theta_d) \quad (1)$$

$$\theta = (\theta_k, \theta_d), \theta_k \cap \theta_d = \emptyset \quad (2)$$

其中, $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 分别是模型的输入和输出变量, $\boldsymbol{f}$ 是整体 (Complete) 模型函数, 而 $\boldsymbol{f}_k$ 与 $\boldsymbol{f}_d$ 分别对应知识驱动 (Knowledge-driven, KD) 与数据驱动 (Data-driven, DD) 子模型函数, 它们相应的参数变量是 θ_k 和 θ_d. 由于 θ_k 对应物理意义的参数, 我们将其称为物理参数变量. 符号 “$\oplus$” 记为 “耦合算子 (Coupling operator)”, 它表示两个子模型之间相互作用的运算. 数据驱动子模型可以是机器学习中的决策树、神经元网络或支持向量机等. 对于式 (1), 如果模型函数 $\boldsymbol{f}$ 中包括时间变量 t, 则模型是动态系统.

参数可辨识性研究在知识与数据共同驱动机器学习模型中具有重要的理论意义. 知识驱动 (子) 模型参数变量 θ_k 的可辨识性不仅是模型具有可解释性的必要前提, 也是参数获得正确估计的重要条件. 数据驱动 (子) 模型参数变量 θ_d 的可辨识性对机器学习诸多方面有着重要的影响 (我们将在第 4 节详细讨论). 本文在此模型中开展参数可辨识性研究主要基于以下三个方面的原因.

Prior domain knowledge	Zero knowledge	Partial knowledge	Complete knowledge
Inference methodology	Induction	Abduction or others	Deduction
Suitable model	A neural-net! work-based model	Semi-analytical model	Analytical model
Model parameters	Non-parametric	Semi-parametric	Parametric
Model transparency	Black	Grey	White

图 3　根据先验领域知识、推理方法、模型类型, 模型参数, 模型透明度等划分的模型方法[15]

Fig. 3　The modeling approaches that are based on prior domain knowledge, inference methodology, model type, model parameter and model transparency[15]

1) 新模型提供了更加统一而广义的建模框架, 因为传统的知识驱动或数据驱动模型将成为其特例. 模糊系统与概率图模型等由单一结构化知识表达的模型也可以理解为是新模型中的某一种. 新模型可以包容非结构化的先验知识、任意类型的动态模型等.

2) 不同于传统模型, 新模型中耦合方式及 DD 子模型选择为参数可辨识性提供了新的、特有的研究空间. 调整耦合方式或 DD 子模型可以改变整体模型中参数可辨识性的实际状况, 这也为某些具有物理意义的不可辨识参数转变为可辨识参数提供了可能性[15].

3) 未来的机器学习模型或类人智能模型必然需要同时最大化利用知识与数据. 人类有效结合 “归纳” 与 “演绎” 推理体系而实现的智能范式为新模型提供了参照依据. 未来类人智能机器在利用已有知识与数据方面均会超过人类平均水平 (图 4), 这也预示大数据 (Big data) 下成功发展的深度学习模型必将依赖利用大知识 (Big knowledge) 而产生突破性发展.

2　关键问题一: 可辨识性准则

本节主要讨论与可辨识性准则相关的几个问题: 可辨识性的判定, 参数冗余的判定和重参数化方法.

2.1　无约束参数模型的可辨识性准则

在此情况下, 我们假设可容许的参数空间是某个欧氏空间. 早在 1971 年, Rothenberg[34] 就推导了参数统计模型局部可辨识的充分必要条件, 正是该文献证明了模型局部不可辨识和模型奇异是等价的. 特别地, 文献 [34] 还得到了结论: 对于指数分布

族, 此条件也是模型全局可辨识的充分必要条件. 从信息论 (Information theory)[35] 的观点来看, 不可辨识的原因在于缺乏足够的 "信息" 把一个参数点和另外一个参数点区分开, 因此可以利用信息论的相关工具来研究可辨识性. 在文献 [36] 中, Bowden 利用信息论中的 Kullback-Leibler (KL) 散度, 将可辨识性问题等价地转化为求解非线性方程根的个数问题. 文献 [37] 等利用 KL 散度, 把无约束参数模型的可辨识性问题等价地转化为一个无约束最优化 (Unconstrained optimization) 问题[38], 用最优化理论的观点研究了可辨识性问题. 进而, 文献 [39] 利用辨识函数 (Identifying function) 的方法研究了可辨识性问题. 通常, 这些方法的复杂度都很高 (比如, 需要显式地求解 Fisher 信息矩阵或者 KL 散度), 或者所得结果只能用于具体的模型 (比如, 文献 [40] 用复分析的方法研究了多元高斯分布的可辨识性问题), 或者对模型施加了某些限定 (比如, 文献 [39] 假定模型辨识函数的维数是有限的). 所以, 对于任意的参数统计模型, 缺乏一个简便高效的解决办法.

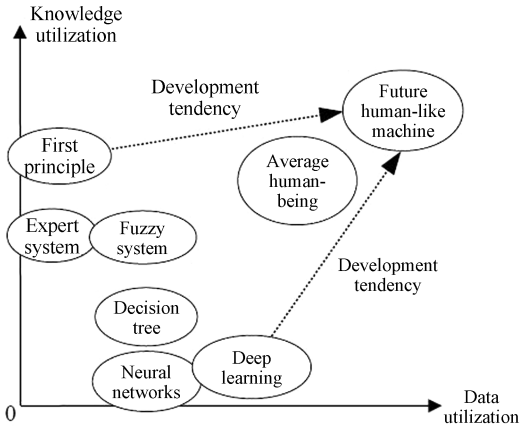

图 4 现有智能模型与未来类人机器在知识与数据利用中的相对关系示意图

Fig. 4 The relationship between current intelligent models and the future human-like machines which is based on the use of knowledge and data

2.2 参数受限 (Parameter-constrained) 模型的可辨识性准则

如果无约束参数模型不可辨识, 我们可以改变建模方式使得模型参数变得可辨识. 通常有两种方式实现这一目的: 第一种方式是在模型中加入先验知识[41−42], 使得模型在受限的参数空间中可辨识. 第二种方式是对参数引入先验分布, 把参数学习置于 Bayes 统计框架[43−44]. 在一些特殊情形下, 这两种方式具有等价性[43−44].

在第一种方式中, 目前学界研究最普遍的情形是: 假设模型参数满足一族等式约束, 从而把可容许参数空间限制在其低维流形. 文献 [34] 研究了参数受限统计模型的可辨识性问题. 文献 [39] 利用 KL 散度, 把参数受限模型的可辨识性问题等价于一个约束最优化 (Constrained optimization) 问题, 用最优化理论的观点研究了可辨识性问题.

参数约束研究中的难点是不再局限于 "刚性 (即严格满足)" 约束条件, 而是寻求广义约束. 以包容各种 "柔性 (松弛满足)" 约束条件, 如语义 (Semantic) 约束、概率约束、非精确约束等. 关于此类问题的可辨识性准则, 学界至今还缺乏系统性研究.

2.3 全局可辨识性准则

学界关于可辨识性准则研究的另一难点是全局可辨识问题. 目前只有很少的几个结论, 主要是关于线性模型, 前馈型神经网络等简单模型或者具有特殊结构的模型. 两个常用的方法是: 估计量方法[45−46] 和特征量 (Characteristic) 方法[34]. 但是, 如何寻找 (或构造) 满足条件的估计量或者特征量, 统计学界也没有得到一个通用的方法. 根据文献 [37], 我们可以利用 KL 散度, 把全局可辨识性问题等价地转化为全局最优化问题. 然而, 除了凸优化 (Convex optimization)[47] 问题之外, 在非线性最优化理论中还没有关于全局最小值 (即唯一极小值) 的有效而普适的理论 (很多理论结果都需要各种约束规范[38], 或者结果不具有实用性).

2.4 参数冗余

模型不可辨识的最主要原因之一是参数冗余 (Parameter redundancy). 如果模型能用更少的参数 $\beta = \boldsymbol{g}(\theta)$ 等价地表达, 则称模型是参数冗余的[48−50]. 否则, 称模型是满秩的. 从几何上看, 如果模型是参数冗余的, 那么对任何训练数据集而言, 似然函数具有完全平坦的岭线[50] (这个结论也可以从 Amari 研究奇异模型学习动态轨迹的实验中得到验证[51]).

对模型参数冗余的研究是参数可辨识性问题研究的重点之一. 在文献 [50] 中, Catchpole 等证明了: 在指数分布族中, 模型局部不可辨识和模型参数冗余是完全等价的. 但是, 对于一般的参数模型, 二者是否等价仍然没有得到证明. 文献 [52] 首次推导了 Bayesian 分类器和代价敏感学习中代价矩阵的独立参数定理. 该定理不仅为代价敏感学习中参数冗余提出新的理论见解 (即代价矩阵中最大独立参数个数), 也为以泛函形式表达的机器学习目标函数中独立参数个数推导提供了范例. 在参数冗余研究中, 主要的瓶颈问题如下:

1) 参数冗余的判定准则. 参数冗余的判定最终可以归结为求辨识函数导矩阵 (Derivative matrix) 的符号秩 (Symbolic rank) 问题[39]. 由于导矩阵的

每个元素都是参数的函数, 从而导矩阵是一个泛函矩阵. 但是, 如何在模型中提取辨识函数, 则需要具体情况具体分析. 目前, 只能对若干常用模型 (比如线性模型、指数族、滑动平均模型等) 提取其辨识函数[53]. 如果模型通过非线性状态–空间方程或者更复杂的微分方程 (尤其是偏微分方程) 表示, 学界没有求解模型辨识函数的通用方法. 而在生物、医学或工程领域, 模型的输入输出关系常常不是以参数统计分布的形式显式地给出, 而是以泛函形式、时间序列、微分方程、积分方程、差分方程或者混合形式隐式地给出, 这都大大增加了研究的复杂度.

2) 机器证明. 参数冗余研究中的一个重要内容是机器证明的应用. 当冗余判定定理存在时, 人们希望应用计算机来自动推断模型是否参数冗余, 而非人工推断. 目前主要的方法有解析计算法、符号计算法和数值计算法三大类. 解析计算法和数值计算法自身的缺点使得符号计算法成为最主要的方法.

在计算导矩阵符号秩的时候, 通常需要利用符号计算软件 (比如 Maple、Mathematica、Maxima) 求解. 从理论上说, 只要辨识函数是有限维的, 我们都可以写出其导矩阵. 但是, 在模型的参数结构比较复杂或参数数量非常多的情况下, 由于计算机计算能力和内存的限制, 计算导矩阵符号秩的任务往往很难实现. 比如在文献 [54] 中, 作者研究了一个线性分室 (Compartmental) 模型, 该模型仅仅只有 11 个参数, 但计算机不能求出其导矩阵的秩. 这就说明, 导矩阵符号秩计算的复杂度依赖于模型的非线性程度 (包括模型关于输入、输出变量的非线性程度, 以及模型关于参数的非线性程度), 参数维数等. 根据文献, 到目前为止, 唯一可以 “全自动” 判断模型参数冗余的符号计算方法只有基于微分代数的 DAISY (Differential algebra for identifiability of SYstems) 软件[55], 但是该方法只能解多项式或者有理型, 不能适用于任意非线性模型. 所以, 对于复杂的参数模型, 如何发展高效、自动的计算方法也是可辨识性研究中需要解决的问题. 当符号计算无法胜任时, 数值计算可能是必须的.

3) 重参数化 (Reparametrization). 如果模型已经被判定为参数冗余的, 我们可以对模型进行重参数化 (即: 把原参数 θ 的函数 $\boldsymbol{r}(\theta)$ 作为模型新的参数, 从而把模型表示为 $\boldsymbol{r}(\theta)$ 的函数)[39, 56]. 通过重参数化, 我们可以减少模型 “内在参数” 的个数, 得到模型最精简的参数表示. 重参数化的过程可以通过求解一个和导矩阵相关的一阶线性偏微分方程组实现[39].

重参数化可以理解为参数空间更新问题, 这也是未来机器学习中必不可少的智能进化方面的研究内容. 从已有的文献数量来看, 关于重参数化的研究相对比较缺乏, 主要集中在一些相对简单的模型[57]. 目前, 更多的研究停留在判定模型是否可以辨识或模型是否冗余的层面, 而没有进一步很好地解决参数约束或重参数化问题.

4) 各种可辨识性概念之间的联系. 在可辨识性研究中, 由于理论自身和实际应用的推动, 各种新的观点、概念和模型层出不穷, 其中很重要的一点是理清各种已有概念之间的联系. 虽然已有大量文献 (比如文献 [11, 58−59]) 对此问题做了深入的研究, 但是, 仍然有很多问题没有得到圆满的解决. 在文献 [50] 中, Catchpole 研究了参数冗余和非 δ-辨识的关系, 即: 如果模型是参数冗余的, 则模型一定是非 δ-可辨识的. 如果更深入地研究满秩 (非冗余) 模型, 我们可以把满秩模型划分为本性满秩 (Essentially full rank) 和条件满秩 (Conditionally full rank) 两类, 即: 如果模型导矩阵在整个参数空间处处满秩, 我们称此模型为本性满秩的; 否则, 称为条件满秩的. 在文献 [50] 中, Catchpole 等证明了一个重要事实, 即: 如果一个模型是本性满秩的, 则此模型是 δ-可辨识的. Catchpole 进而提出猜想: 如果模型是本性满秩的, 则该模型一定是全局可辨识的. 此猜想还是一个公开问题.

3 关键问题二: 可辨识性研究对机器学习各个方面的影响

在经典统计理论中, 可辨识性是一个基本假设, 很多重要统计性质的推导都依赖于此假设. 比如, 极大似然估计和 Bayes 后验分布的渐近正态性, Cramér-Rao 不等式等[51]. 可辨识性研究和机器学习领域的很多主题也有密切的联系, 比如概率主成分分析[27]、隐因子模型[28]、变分 Bayesian 矩阵分解[60]、低秩矩阵完备化[61] 等. 然而, 在机器学习领域, 几乎所有的学习机都是奇异的. 这种奇异性将对机器学习的很多方面产生深刻的影响, 比如参数学习、模型选择、学习算法、学习动态分析、Bayes 推断等. 因此, 我们需要一种全新的统计理论来研究这类奇异学习机. 直到现在, 仍然还有很多艰深的理论问题没有得到完全地解决. 在本节中, 我们主要讨论奇异性对机器学习各个方面的影响.

3.1 奇异统计模型参数空间的几何结构

根据 Amari 信息几何的观点, 通过模型的一阶近似, 正则 (Regular, 即非奇异) 统计模型参数空间的局部结构可以用其统计流形的切空间表示, 高阶近似可以根据信息几何中的仿射连接 (Affine connection) 和相关的 e-曲线和 m-曲线得到[18]. 但是, 奇异模型的统计流形在奇异点处没有切空间, 而只能通过切锥表示. Dacunha 等[62] 以统一的方式

研究了奇异统计模型的局部结构, 在奇异模型中, 所有观测等价的参数以锥 (Cone) 的形式嵌入到正则流形中. 关于此问题的文献甚至可以追溯到 1976 年 Brockett 关于线性系统的研究[63].

在机器学习领域, 学界对此也有大量研究, 比如, Amari 等[51] 通过研究层次神经网络, 发现该模型的参数等价类并不是孤立集合, 而是形成一个连续统 (Continuum). 这些结果从理论上揭示了奇异模型参数空间的几何结构, 也表明此类模型的参数空间比正则模型更加困难. 文献 [56] 研究了奇异统计模型参数空间的几何结构. 总之, 奇异模型的代数结构和几何结构需要更系统深入的理论研究.

3.2 参数估计

奇异性对参数估计的影响体现在以下几个方面.

1) 极大似然估计. 根据 Vapnik 的观点, 在传统统计理论的框架中, 函数估计都是基于极大似然方法的, 极大似然方法是传统体系下的一个归纳引擎[64−65]. 比如参数估计的极大似然法, 模式识别的判别分析法和回归估计的最小二乘法都可以纳入这个框架. 从另一角度看, 根据 Amari 的信息几何理论, 在正则统计理论中, 分布族的 Fisher 信息矩阵处处正定, 从而在参数流形 (神经网络文献里被称为神经流形 (Neuromanifold)) 上确定了一个 Riemann 度量, 我们可以根据这个度量所确定的拓扑结构, 研究参数空间的各种几何性质.

如果一个参数分布族非奇异, 则 Fisher 信息矩阵处处正定, 并且被广泛地用来度量样本包含的总体信息. 比如, 似然函数可以用参数的二次型逼近, 极大似然估计是渐近有效估计量 (Asymptotically efficient estimator), 以速度 $o(1/n)$ 渐近收敛到正态随机变量 (n 为样本数量), 著名的 Cramér-Rao 不等式成立等[51].

但是, 在奇异统计模型中, Fisher 信息矩阵并非处处正定, 从而导致一系列截然不同的性质. 比如, 极大似然估计不再收敛到正态变量, 著名的 Cramér-Rao 不等式不再成立. 再比如, 在正则模型中, 极大似然估计通过极大化似然函数得出; 然而, 在奇异模型中, 求解极大似然估计需要随机场理论. 总之, 这些奇异性使得参数空间的几何结构更加复杂, 从而需要一种全新的理论来处理这个问题.

2) 大样本性质. 通常情况下, 我们很难推导出有限样本条件下估计量精确的概率分布. 然而, 借助于经典统计的标准工具 (比如大数定律和中心极限定律), 我们可以得到估计量的大样本性质. 其中, 很关键的性质是一致性 (Consistency). 在正则模型中, 一致性表明估计量几乎必然 (Almost surely, a. s.) 收敛到最优参数. 然而, 在奇异模型中, 一致性表明估计量 a. s. 收敛到一个参数集合 (该集合中的参数都能使误差函数达到极小)[66]. 这意味着, 在奇异模型中, 误差函数可以在参数空间的某些曲线上取恒定不变的值. 这个结论和 Amari 的观察[51] 同样也是吻合的.

大样本性质中另一个重要的问题是极限分布 (Limiting distribution). 在正则模型中, 估计量的极限分布是高斯分布. 然而, 在奇异模型中, 估计量的极限分布属于极限混合高斯分布族 (Limiting mixed Gaussian family)[66]. 另外, 在假设检验问题中, 奇异模型也体现了和正则模型完全不一样的性质[51]. 所以, 对奇异统计模型而言, 经典数理统计领域几乎所有的结果都需要重新认识.

3.3 模型选择

模型选择是机器学习中的一个重要问题. 在模型选择方面, 目前没有一般的方法可以推荐, 通常根据可辨识性、灵活性、吝啬性等原则进行统筹兼顾. 经常使用的模型选择准则是 Akaike information criterion (AIC), Bayesian information criterion (BIC), Minimum description length (MDL) 等. Hagiwara 等[67] 最先注意到, 在神经网络建模时, 如果使用 AIC 作为模型选择标准, 其结果并不理想. 后来, Hagiwara 注意到这是由于神经网络模型的奇异性导致的. 从而, 文献 [68] 提出了网络信息准则 (Network information criterion, NIC), 由于 NIC 考虑到了奇异性的影响, 其理论和数值实验的结果都明显好于 AIC. 只是 NIC 的基本思想类似 Vapnik 的结构风险极小化归纳原则[64−65], 需要构造一个嵌套的函数集合, 从而实现起来很困难.

从奇异统计的角度来看, 因为 AIC 旨在极小化模型的泛化误差[69], 而在奇异模型中, 泛化误差较正则模型具有更复杂的性质. 在奇异模型中, BIC、MDL 的缺陷和 AIC 类似. 另外, 在 Bayes 模型比较方面, 模型的奇异性也是一个必须要考虑的因素[44]. 所以, 如何在奇异模型中设计好的模型选择标准, 是机器学习领域的一个重要而困难的问题.

3.4 学习算法

随着学习模型中参数维数和非线性程度的急剧增加, 学习过程中的计算资源开销也越来越大. 为了加快参数的学习过程, 与奇异性有关的学习算法问题也成了一个热点. 学界很早就注意到, 多层感知器网络中的后向传播 (Back propagation, BP) 算法收敛速度非常慢. 为了加快 BP 算法的收敛速度, 很多学者相继提出了大量的方法, 典型的方法有 Duda 等提出自适应步长和动量方法等[70], 但是这些方法本质上都是基于梯度下降的一阶算法, 所以无法从根本上克服平台现象 (Plateau phenomenon). 各种

二阶方法, 比如牛顿法、共轭梯度法和拟牛顿法虽然利用了参数空间的曲率信息, 但这些算法计算开销极大, 只具备局部收敛性, 而且绝大部分算法只适合批量方式, 不适合大规模数据或在线学习的场景. Amari 注意到导致此平台现象 (或慢流形) 的主要原因是模型的奇异性. 为克服慢收敛现象, Amari 提出自然梯度下降 (Natural gradient descent) 算法[71], 该算法考虑到了参数空间的流形结构, 从而获得了更快的收敛速度. 文献 [73] 研究了自然梯度下降算法的计算复杂性问题. 文献 [73−74] 研究了自然梯度下降算法的统计物理机理. 然而, 由于自然梯度下降算法 (包括其自适应形式[71]) 本身的复杂度也很高, 所以, 在奇异模型中, 期待有更高效 (至少二阶收敛速度) 的算法出现.

3.5　学习过程的动态轨迹分析

学习过程的动态轨迹分析对研究模型的奇异性有重要的意义. 在奇异模型中, 由于参数不可辨识, 使得所有观测等价的参数具有同样的误差函数, 从而误差曲面在很多地方是平坦的 (图 5), 正是这个原因使得学习过程非常缓慢, 并导致两个严重后果: 1) 学习过程非常缓慢, 通常学习轨迹会在平坦的低维流形上逗留很长时间, 随着训练数据中的随机噪声影响, 才可能继续下降[71]. 2) 最终结果陷入局部极小值.

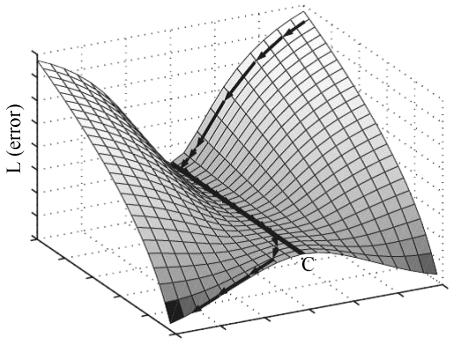

图 5　在奇异点附近, 参数的学习轨迹在误差曲面上有完全平坦的岭线[51]

Fig. 5　Learning trajectory of parameters near the singularities has completely flat ridge in error surface[51]

从理论分析的角度看, 在奇异模型中, 批 (Batch) 学习方式会平滑掉训练数据中的随机噪声, 使得学习过程容易陷入平坦的低维流形和局部极小点, 而 Online 学习方式会使学习过程的动态轨迹更容易从平坦区域中逃离. 因此, Online 学习方式更适合奇异模型. 为了揭示参数空间的学习轨迹, 文献 [71, 75−76] 分别研究了多层感知器网络、径向基函数网络和高斯混合模型中参数学习的动态过程, 并分析了参数空间中学习过程的动态向量场. 然而, 对一般奇异模型中参数学习过程的动态轨迹分析, 则没有一个普适的结果.

3.6　泛化误差

迄今为止, 关于泛化误差的计算仍然是基于 Cramér-Rao 范例的. 而在奇异统计模型中, 泛化误差的计算需要新的方法 (需要奇异模型中对数似然比的特殊性质). 人们很早就认识到在高斯混合模型中, 对数似然比呈现奇异的性质[77], 这主要是因为奇异统计模型对数似然比的性质与正则情况截然不同. 在神经网络领域, Fukumizu[78] 首先通过一个简单的线性模型揭示了多层感知器的泛化误差和一般正则模型的泛化误差明显不同, 进而利用高斯随机场理论得到多层感知器模型的对数似然比的渐近性质和泛化误差的准确结果.

目前, 学界所得到的结果非常有限[23, 79], 多数都是针对特殊模型来研究. 至于一般奇异模型的通用结果, 还需要学界更加深入地研究.

3.7　Bayes 推断

对于参数的学习, 如果我们不使用极大似然方法, 而改用 Bayes 学习方法, 那么, 由于 Bayes 方法引入了先验知识, 在很多场合会使过拟合问题得到大大的缓解, 显示卓越的泛化性能. 但是在奇异学习机中, 我们将不可避免地遇到理论上的困难. Watanabe 最先注意到在奇异统计模型中使用 Bayes 推理的理论问题[23], 比如, 如果我们使用的先验知识是 “无信息先验 (Non-informative prior)”, 则光滑的先验密度在由奇异点构成的等价类上是无穷大, 这使得模型的后验分布将对奇异点有所偏好, 这又违背了 “无信息先验” 的原则, 这显然不合常理. 另外, 在奇异模型中, Bayes 后验分布也不再渐近地收敛于正态分布.

Watanabe 最先研究了奇异性对 Bayes 推断的影响. 通过利用 Hironaka 奇异性分解定理和 Sato 公式, Watanabe 将代数几何和代数分析的工具引入奇异机器学习理论, 以此来研究各种层次奇异学习机中 Bayes 预测分布的渐近性能, 并得出了一系列结果, 部分结果总结在其专著[23] 和网站 http://watanabe-www.math.dis.titech.ac.jp/users/swatanab/singstat-intro.html. 其主要理论结果可以总结为 4 个方面: 对数似然比函数的标准型、随机复杂度的收敛问题、Bayes 估计中的两个方程以及训练误差和泛化误差的对称关系[23]. 关于奇异性对学习理论各个方面的影响, 应该成为机器学习中的重要基础知识, 并有待理论研究上的突破.

4 结论

本文试图说明参数可辨识性研究是统计机器学习中的基础理论内容, 并给出其中的两个关键问题. 文献 [80] 主要讨论了统计框架下两个关键问题的技术解决, 本文则侧重研究框架的选择, 并将可辨识性问题纳入此框架下进行讨论, 期望加深对深度学习以及人工智能在方法论发展前景下的问题认知. 同时, 我们也试图说明参数可辨识性研究的大背景是人类对生物智能与机器智能的深刻认知与理解. 由于生物智能本质是非透明的, 我们必须借助机器智能仿真揭示生物智能的内在机理. 这也表明机器智能不是简单地复制生物智能功能, 而是可以超越现有生物智能的知识与众多功能. 目前深度学习网络在若干大数据应用中取得了卓越成功, 包括超出人类平均水平的模式识别精度[81]. 但是, 当下深度学习网络仍然属于黑箱模型范畴, 并缺乏理论解释的支撑. 在文献 [82] 中, 作者指出: 无论是基于 "工具论 (追求效率或预测)" 或 "认知论 (追求解释或理解)" 为导向的研究, 以人工神经元网络为代表的数据驱动模型必然要走向包容 "增加模型透明度 (或可理解性)" 的学习目标. 回到本文图 2 中示意的机器学习模型, 在给定的参数集里, 哪些参数是可辨识的, 哪些参数是不可辨识的? 对于不同的奇异机器学习模型, 它们各自的参数空间几何结构有何不同? 这些问题都会指向参数可辨识性的研究内容. 这些研究会带来对模型自身物理意义的解释, 以及对输入输出响应关系的解释, 它是导致我们对学习机器持续完善以及对人工智能深刻认知的必然路径.

因此, 本文侧重于从参数可辨识性研究角度为机器学习理论和实践带来新的研究视角, 希望能够引起不同学科领域中研究者的进一步思考和质疑. 我们认为今后对各个层面问题给出完整而严谨的数学表达定义是必不可少的内容和挑战, 更大的挑战是如何将控制科学中的重要概念结合到其他学科领域研究中. 比如, 在金融、经济、教育、社会发展中, 某些自变量的设定 (如存贷款基准利率) 或公共政策的调整 (如发展指标) 可以借鉴控制科学中提出的 "调节"、"反馈"、"可控性"、"可观性" 以及 "平行管理与控制[83]" 的概念而开展研究. 对于这些典型的复杂巨系统[84−87] 问题 (它们的演变机理在本质上同样不透明, 并与智能及非智能人类行为相关), 知识与数据共同驱动的建模思想及其参数可辨识性研究会对深度学习、人工智能以及其他学科发展产生独到的贡献.

References

1 Matzkin R L. Nonparametric identification. *Handbook of Econometrics*. New York: Elsevier Science Ltd, 2007.

2 Dufour J M, Hsiao C. Identification. *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan Ltd., 2008.

3 Koopmans T C. Identification problems in economic model construction. *Econometrica*, 1949, **17**(2): 125−144

4 Aldrich J J. Haavelmo′s identification theory. *Econometric Theory*, 1994, **10**: 198−219

5 Zadeh L A. On the identification problem. *IRE Transactions on Circuit Theory*, 1956, **3**(4): 277−281

6 Zadeh L A. From circuit theory to system theory. *Proceedings of the Institute of Radio Engineers*, 1962, **50**(5): 856−865

7 Ljung L. *System Identification: Theory for the User* (Second Edition). Upper Saddle River, NJ: Prentice-Hall, 1999.

8 Chen H F, Guo L. *Identification and Stochastic Adaptive Control*. Boston, MA: Birkhauser, 1991.

9 Walter E, Pronzato L. *Identification of Parameter Models from Experimental Data*. London: Springer-Verlag, 1997.

10 Zhou Tong. *Introduction to Control-oriented System Identification*. Beijing: Tinghua University Press, 2002.
(周彤. 面向控制的系统辨识导论. 北京: 清华大学出版社, 2002.)

11 Miao H Y, Xia X H, Perelson A S, Wu H L. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*, 2011, **53**(1): 3−39

12 Wang Le-Yi, Zhao Wen-Xiao. System identification: new paradigms, challenges, and opportunities. *Acta Automatica Sinica*, 2013, **39**(7): 933−942
(王乐一, 赵文虓. 系统辨识: 新的模式、挑战及机遇. 自动化学报, 2013, **39**(7): 933−942)

13 Ljung L. Perspectives on system identification. *Annual Reviews in Control*, 2010, **34**(1): 1−12

14 Ran Z Y, Hu B G. Determining structural identifiability of parameter learning machines. *Neurocomputing*, 2014, **127**: 88−97

15 Hu B G, Qu H B, Wang Y, Yang S H. A generalized-constraint neural network model: associating partially known relationships for nonlinear regressions. *Information Sciences*, 2009, **179**(12): 1929−1943

16 Koopmans T C, Reiersol O. The identification of structural characteristics. *Annuals of Mathematical Statistics*, 1950, **21**(2): 165−181

17 Bellman R, Aström K J. On structural identifiability. In: Proceedings of the Mathematical Biosciences. Amsterdam: Elsevier, 1970, **7**: 329−339

18 Amari S I, Nagaoka H. *Methods of Information Geometry*. New York: AMS and Oxford University Press, 2000.

19 Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 2010, **11**: 3571−3594

20 Watanabe S. Algebraic geometrical methods for hierarchical learning machines. *Neural Networks*, 2001, **14**(8): 1049−1060

21 Watanabe S. Algebraic geometry of singular learning machines and symmetry of generalization and training errors. *Neurocomputing*, 2005, **67**: 198−213

22 Watanabe S. Almost all learning machines are singular. In: Proceedings of the 2007 IEEE Symposium on Foundations of Computational Intelligence. Piscataway, USA: IEEE, 2007.

23 Watanabe S. *Algebraic Geometry and Statistical Learning Theory*. Cambridge: Cambridge University Press, 2009.

24 Chen A M, Lu H, Hecht-Nielsen R. On the geometry of feedforward neural network error surfaces. *Neural Computation*, 1993, **5**(6): 910−927

25 Kurková V, Kainen P C. Functionally equivalent feedforward neural networks. *Neural Computation*, 1994, **6**(3): 543−558

26 Sussmann H J. Uniqueness of the weights for minimal feedforward nets with a given input-output map. *Neural Networks*, 1992, **5**(4): 589−593

27 Bishop C M. *Pattern Recognition and Machine Learning*. Berlin: Springer, 2006

28 Henao R, Winther O. Sparse linear identifiable multivariate modeling. *Journal of Machine Learning Research*, 2011, **12**: 863−905

29 Walter E, Lecourtier Y. Unidentifiable compartment models: what to do? *Mathematical Biosciences*, 1981, **56**(1−2): 1−25

30 Wu H L, Zhu H H, Miao H Y, Perelson A S. Parameter identifiability and estimation of HIV/AIDS dynamic models. *Bulletin of Mathematical Biology*, 2008, **70**(3): 785−799

31 Xia X, Moog C H. Identifiability of nonlinear systems with application to HIV/AIDS models. *IEEE Transactions on Automatic Control*, 2003, **48**(2): 330−336

32 Fortunati S, Gini F, Greco M S, Farina A, Graziano A, Giompapa S. On the identifiability problem in the presence of random nuisance parameters. *Signal Processing*, 2012, **92**(10): 2545−2551

33 Yang S H, Hu B G, Cournède P H. Structural identifiability of generalized-constraint neural network models for nonlinear regression. *Neurocomputing*, 2008, **72**(1−3): 392−400

34 Rothenberg T J. Identification in parametric models. *Econometrica*, 1971, **39**(3): 577−591

35 Cover T M, Thomas J A. *Elements of Information Theory* (Second Edition). Chichester: Wiley-Blackwell, 1991.

36 Bowden R J. The theory of parametric identification. *Econometrica*, 1973, **41**: 1069−1074

37 Ran Z Y, Hu B G. Determining parameter identifiability from the optimization theory framework: a Kullback-Leibler divergence approach. *Neurocomputing*, 2014, **142**: 307−317

38 Luenberger D G, Ye Y Y. Linear and nonlinear programming. *International Series in Operations Research & Management Science* (Second Edition). New Jersey: Addison-Wesley, 1984

39 Ran Z Y, Hu B G. An identifying function approach for determining parameter structure of statistical learning machines. *Neurocomputing*, 2015, **162**: 209−217

40 Hochwald B, Nehorai A. On identifiability and information-regularity in parameterized normal distributions. *Circuits Systems Signal Processing*, 1997, **16**(1): 83−89

41 Stoica P, Ng B C. On the Cramér-Rao bound under parametric constraint. *IEEE Signal Processing Letters*, 1998, **5**(7): 177−179

42 Yao Y W, Giannakis G. On regularity and identifiability of blind source separation under constant modulus constraints. *IEEE Transactions on Signal Processing*, 2005, **53**(4): 1272−1281

43 Murphy K P. *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT Press, 2012.

44 Bishop C M. *Pattern Recognition and Machine Learning*. Berlin: Springer, 2006.

45 Paulino C D M, de Bragança Pereira C A. On identifiability of parametric statistical models. *Journal of the Italian Statistical Society*, 1994, **3**(1): 125−151

46 Ernesto S M, Fernando Q. Consistency and identifiability, revisited. *Brazilian Journal of Probability and Statistics*, 2002, **16**(1): 99−106

47 Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge: Cambridge University Press, 2004.

48 Catchpole E A, Morgan B J T, Freeman S N. Estimation in parameter redundant models. *Biometrika*, 1998, **85**(2): 462−468

49 Catchpole E A, Morgan B J T. Deficiency of parameter-redundant models. *Biometrika*, 2001, **88**(2): 593−598

50 Catchpole E A, Morgan B J T. Detecting parameter redundancy. *Biometrika*, 1997, **84**(1): 187−196

51 Amari S I, Park H, Ozeki T. Singularities affect dynamics of learning in neuromanifolds. *Neural Computation*, 2006, **18**(5): 1007−1065

52 Hu B G. What are the differences between Bayesian classifiers and mutual information classifiers. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **25**(2): 249−264

53 Ran Z Y, Hu B G. An identifying function approach for determining structural identifiability of parameter learning machines. In: Proceedings of International Joint Conference on Neural Networks. Beijing: IEEE, 2014.

54 Jiang H H, Pollack K H, Brownie C, Hightower J E, Hoeing J E, Hearn W S. Age-dependent tag return models for estimating fishing mortality, natural mortality and selectivity. *Journal of Agricultural, Biological, and Environmental Statistics*, 2007, **12**(2): 177−194

55 Saccomani M P, Audoly S, Bellu G, D′Angiò L. Examples of testing global identifiability of biological and biomedical models with the DAISY software. *Computers in Biology and Medicine*, 2010, **40**(4): 402−407

56 Dasgupta A, Self S G, Gupta S D. Unidentifiable parametric probability models and reparameterization. *Journal of Statistical Planning and Inference*, 2007, **137**(11): 3380−3393

57 Evans N D, Chappell M J. Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems. *Mathematical Biosciences*, 2000, **168**(2): 137−159

58 Little M P, Heidenreich W F, Li G Q. Parameter identifiability and redundancy in a general class of stochastic carcinogenesis models. *PLoS ONE*, 2009, **4**: e8520

59 Little M P, Heidenreich M F, Li G. Parameter identifiability and redundancy: theoretical considerations. *PLoS ONE*, 2010, **5**: e8915

60 Nakajima S, Sugiyama M. Implicit regularization in variational Bayesian matrix factorization. In: Proceedings of the International Conference on Machine Learning. Piscataway, USA: IEEE, 2010.

61 Kiraly F, Tomioka R. A combinatorial algebraic approach for the identifiability of low-rank matrix completion. In: Proceedings of the 29th International Conference on Machine Learning. Edinburgh, Scotland: ACM, 2012. 755−762

62 Dacunha-Castelle D, Gassiat É. Testing in locally conic models, and application to mixture models. *Probability and Statistics*, 1997, **1**: 285−317

63 Brockett R W. Some geometric questions in the theory of linear systems. In: Proceedings of the IEEE Conference on Decision and Control including the 14th Symposium on Adaptive Processes. New York: IEEE, 1975. 71−76

64 Vapnik V N. *The Natural of Statistical Learning Theory*. New York: Springer, 1995.

65 Vapnik V N. *Statistical Learning Theory*. New York: John Wiley and Sons, 1998.

66 White H. Learning in artificial neural networks: a statistical perspective. *Neural Computation*, 1989, **1**(4): 425−464

67 Hagiwara K. On the problem in model selection of neural network regression in overrealizable scenario. *Neural Computation*, 2002, **14**(8): 1979−2002

68 Murata N, Yoshizawa S, Amari S I. Network information criterion-determining the number of hidden units for an artificial network model. *IEEE Transactions on Neural Networks*, 1994, **5**(6): 865−872

69 Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1974, **19**(6): 716−723

70 Duda R O, Hart P E, Stork D G. *Pattern Classification*. New York: Wiley, 2001.

71 Amari S I. Natural gradient works efficiently in learning. *Neural Computation*, 1998, **10**(2): 251−276

72 Yang H H, Amari S I. Complexity issues in natural gradient descent method for training multi-layer perceptrons. *Neural Computation*, 1998, **10**(8): 2137−2157

73 Rattray M, Saad D. Analysis of natural gradient descent for multilayer neural networks. *Physical Review E*, 1999, **59**(4): 4523−4532

74 Rattray M, Saad D, Amari S I. Natural gradient descent for online learning. *Physical Review Letters*, 2000, **81**(24): 5461−5464

75 Cousseau F, Ozeki T, Amari S I. Dynamics of learning in multilayer perceptrons near singularities. *IEEE Transactions on Neural Networks*, 2008, **19**(8): 1313−1328

76 Wei H K, Zhang J, Cousseau F, Ozeki T, Amari S I. Dynamics of learning near singularities in layered networks. *Neural Computation*, 1989, **20**(3): 813−843

77 Weyl H. On the volume of tubes. *American Journal of Mathematics*, 1939, **61**(2): 461−472

78 Fukumizu K. Generalization error of linear neural networks in unidentifiable cases. In: Proceedings of the 10th International Conference on Algorithmic Learning Theory. Berlin: Springer-Verlag, 1999. 51−62

79 Liu X, Shao Y Z. Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statics*, 2003, **31**(3): 807−832

80 Ran Z Y, Hu B G. Parameter identifiability in statistical machine learning: a review. *Neural Computation*, 2017, **29**(5): 1151−1203

81 He K M, Zhang X Y, Ren S Q, Sun J. Delving deep into rectifiers: surpassing human-level performance on imageNet classification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 1026−1034

82 Hu Bao-Gang, Wang Yong, Yang Shuang-Hong, Qu Han-Bing. How to add transparency to artificial neural networks. *Pattern Recognition and Artificial Intelligence*, 2007, **20**(1): 72−84
(胡包钢, 王泳, 杨双红, 曲寒冰. 如何增加人工神经元网络的透明度. 模式识别与人工智能, 2007, **20**(1): 72−84)

83 Wang Fei-Yue, Liu De-Rong, Xiong Gang, Cheng Chang-Jian, Zhao Dong-Bin. Parallel control theory of complex systems and applications. *Complex Systems and Complexity Science*, 2012, **9**(3): 1−12
(王飞跃, 刘德荣, 熊刚, 程长建, 赵冬斌. 复杂系统的平行控制理论及应用. 复杂系统与复杂性科学, 2012, **9**(3): 1−12)

84 Ma Shi-Jun, Wang Ru-Song. The Social-economic-natural Complex Ecosystem. *Acta Ecologica Sinica*, 1984, **4**(1): 1−9
(马世骏, 王如松. 社会－经济－自然复合生态系统. 生态学报, 1984, **4**(1): 1−9)

85 Qian Xue-Sen, Yu Jing-Yuan, Dai Ru-Wei. A new discipline of science — the study of open complex giant system and its methodology. *Chinese Journal of Nature*, 1990, **13**(1): 3−10
(钱学森, 于景元, 戴汝为. 一个科学新领域 — 开放的复杂巨系统及其方法论. 自然杂志, 1990, **13**(1): 3−10)

86 Dai Ru-Wei, Cao Long-Bing. Internet — a open complex giant system. *Scientia Sinica*, 2003, **33**(4): 289−296
(戴汝为, 操龙兵. Internet — 一个开放的复杂巨系统. 中国科学, 2003, **33**(4): 289−296)

87 Wan Bai-Wu. Rethinking macroeconomic modeling in the viewpoint of Cybernetics: review and trend of modeling for predicting the next economic crisis. *Control Theory and Applications*, 2015, **32**(9): 1132−1142
(万百五. 控制论视角下对宏观经济建模的再思考: 为能预测经济危机, 对建模的审视及趋势评述. 控制理论与应用, 2015, **32**(9): 1132−1142)

冉智勇 博士, 重庆邮电大学计算机学院讲师. 主要研究方向为机器学习, 模式识别, 系统辨识. 本文通信作者.
E-mail: ranzy@cqupt.edu.cn
(**RAN Zhi-Yong** Ph. D., lecturer at the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications. His research interest covers machine learning, pattern recognition, and system identification. Corresponding author of this paper.)

胡包钢 博士, 中国科学院自动化研究所研究员. 主要研究方向为人工智能, 计算机建模.
E-mail: hubg@nlpr.ia.ac.cn
(**HU Bao-Gang** Ph. D., professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers artificial intelligence and computational modeling.)