

计算机系统项目综合实践课程的几点教改研究

苗 军

北京信息科技大学计算机学院，北京
Email: jmiao@bistu.edu.cn

收稿日期：2020年11月19日；录用日期：2020年12月1日；发布日期：2020年12月8日

摘 要

计算机系统项目综合实践课程是本教学单位针对高年级本科生开设的一门旨在运用多门课程所学知识、设计和实现一个复杂信息处理系统的综合性实践课程。该课程的教学目标是使学生掌握运用传感器、互联网或移动应用技术获取数据并对它们进行智能化信息处理，从而培养学生解决复杂工程问题的能力。为达到此目标，本文以光纤传感信号的数据获取和识别这一复杂系统为例，研究讨论了引入数据可视化、数学理论类比解析和智能信息处理的几何解释等较为直观形象的教改手段和措施，使学生在数据获取、数据清洗、样本生成、模型理解和系统设计等一系列环节上有效地学习和培养解决复杂工程问题的能力。

关键词

综合实践课程，教学改进，形象直观教学

Improvement Study of the Course of Comprehensive Practice of Computer System Project

Jun Miao

School of Computer Science, Beijing Information Science and Technology University, Beijing
Email: jmiao@bistu.edu.cn

Received: Nov. 19th, 2020; accepted: Dec. 1st, 2020; published: Dec. 8th, 2020

Abstract

The Course of Comprehensive Practice of Computer System Project is set for junior students to design and implement a complex information processing system. The goal of this course is to make students grasp the skills to acquire data from sensors, web or mobile device and build an intelligent information system to handle the data. In this paper, several improved teaching measures for

students to understand data acquiring, data cleaning, sample generating, model and system designing are discussed, in order to training the students' ability to handle the complex project problems.

Keywords

Comprehensive Practice Course, Education Improvement, Visual Teaching

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

计算机系统项目综合实践课程是本院针对高年级本科生开设的一门旨在运用多门课程所学的知识、设计和实现一个复杂信息处理系统的综合性实践课程。该课程旨在使学生掌握运用传感器、互联网或移动应用技术获取相应类型的数据并对它们进行智能化信息处理,从而培养学生解决复杂工程问题的能力。

在不少院校中有类似课程的设立,相关教改研究文献[1] [2] [3] [4]从各个角度提出了相应的措施和方法以提高学生进行综合项目实践的能力。例如,文献[1]提出一种“案例引导,项目驱动”模式下的计算机综合实践能力提升策略。该策略强调通过引入实际的工程项目,从案例出发讲解和引导学生进行综合性的工程项目实践。与文献[1]不同,除了项目驱动的教学模式,文献[2]中还引入了“兴趣导向”的教学机制,强化对综合实践的教学效果。在综合实践类课程的教学理论和方法上,文献[3]提出了一套“四维”教学模型,包含理论与方法基础、分析设计工具、应用领域与软件产品、实现技术与拓展。文献[4]则研究探讨了在教育教学形式上具体的方式方法问题,该文指出形象直观的教学在培养学生抽象思维能力方面起着事半功倍的作用。

2. 本课程的教学现状和改进方向

与文献[1]类似,在本教学单位开设的《计算机系统项目综合实践》课程上,我们也采用了“案例引导、项目驱动”的模式对学生进行教学。与文献[2]类似,我们在教学中也引入了兴趣机制,具体就是提供多个任务案例和实践项目供学生根据自己的兴趣进行选择参与;此外,我们还允许学生根据课程大纲的要求自己设计一个综合实践项目或任务,在任课教师审核通过后,可以开展这一自定项目的学习实践,这样可以最大限度地调动学生参与实践的积极性。与文献[3]相通的是,我们在教学过程中,也强调理论与方法基础在综合项目实践中的重要性,尤其是强调,对智能信息处理理论方法的理解和运用在我们的综合项目实践中起着至关重要的作用。

然而,在过去教授基础理论和方法的过程中,有学生反馈“知其然不知其所以然”,说明在基础理论和方法的教学方式方法上需要改进创新,我们需要采用“形象直观”易于理解的方法[4]对学生阐述相关的基础理论方法,使他们在运用这些理论和方法进行项目实践时做到“心中有数”。为达到此目的,本文以光纤传感信号的数据处理和识别这一复杂系统项目的设计和实现为例,研究讨论引入“数据可视化的物理意义教学”、“样本维数的数学意义教学”和“数据分类的几何意义教学”等较为直观形象的教改手段和措施,使学生在数据获取、数据清洗、样本生成、模型和系统设计等一系列环节上有效地学习和培养解决复杂工程问题的能力。

3. 教改措施

本课程教学所采用的综合项目实践案例之一是“光纤传感信号的数据处理与识别”。如图 1 所示，该项目的应用背景是：为了防止盗挖石油运输管道、窃取石油的事件发生，将光纤线缆埋在石油管道旁侧用以采集地面震动信号并由远程信号处理中心实施分析和监控。信息处理过程包括：震动信号的数据采集、数据清洗、特征提取、样本生成和分类识别。该项目最终将引起震动的事件划分为两类：挖掘机挖掘事件和其它非破坏性背景事件。

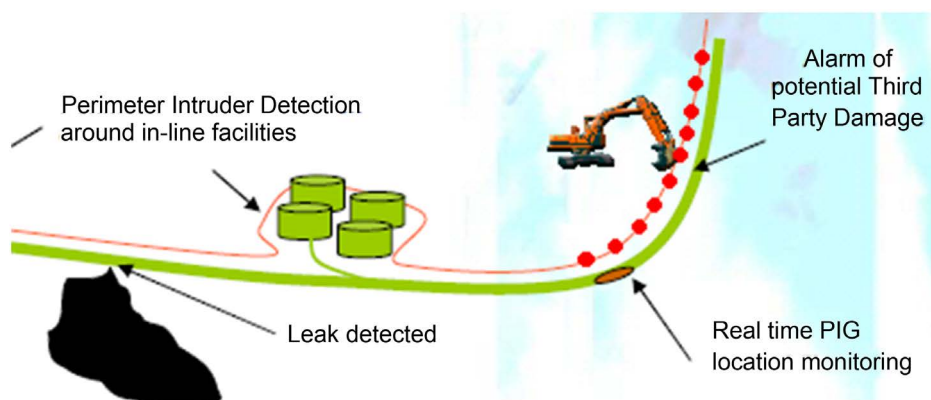


Figure 1. Illustration of opticfiber vibration sensors for excavation monitoring
图 1. 光纤振动传感器用于防范石油管道破坏和石油盗采的监控示意图

我们将以该项目为例，探讨对本课程的几点教学改进措施，主要包括：1) 数据可视化的物理意义教学：用于形象直观地了解数据的特征形态和分布特点，以帮助学生理解对数据进行前、背景分割的操作实践和处理；2) 样本维数的数学意义教学：用于解释说明数据样本生成和维度大小的确定方法；3) 数据分类的几何意义教学：用于阐述分类模型算法的几何运算机理，帮助理解智能信息处理背后的计算原理。

3.1. 数据可视化的物理意义教学

对于初学者而言，往往不清楚待处理的数据信号究竟有什么特点，以及如何针对这些特点采取相应的数据清洗和特征提取操作。例如，对光纤传感信号，初学者只知道该信号可用来传递不同震动的信号，但对不同的震动信号有什么形态和性质、如何去除噪声、如何提取特征等都感到高深莫测。如果用可视化的工具和手段对数据的时空分布形态进行展示，就可以让初学者对光纤信号的特点产生形象直观的认识和了解，进而容易对数据采取相应的清洗和特征提取操作。

如图 2 所示，“挖掘机挖掘”事件可由两个二维时空图可视化表示。每张图上的横坐标代表光纤传感器的序列位置，纵坐标代表传感器的采样时刻；左右两幅图分别表示 250 个光纤传感器在 500 余个时刻上采集到的频率(过零率)和振幅(能量)数值；这些数值通过伪彩色可以获得直观显示。由图上可以看出，在多数传感器上两种特征沿时间轴快速变化，构成前景信号。同时在图片的两侧有大片夹杂着点状分布图案的蓝色区域，这些属于背景信号，与前景信号区域形成鲜明对比。

根据此图，我们告知学生，采集到的数据有两个特点：1) 数据在纵向方向有明显的频率变化特征可以提取；2) 数据区域混杂有前景信号和背景噪声，需要将前景信号区域从背景中分割出来。进一步，我们向学生展示可视化图 3，教授学生用红色方框分割出前景部分，这一操作可以通过编程用程序实现。

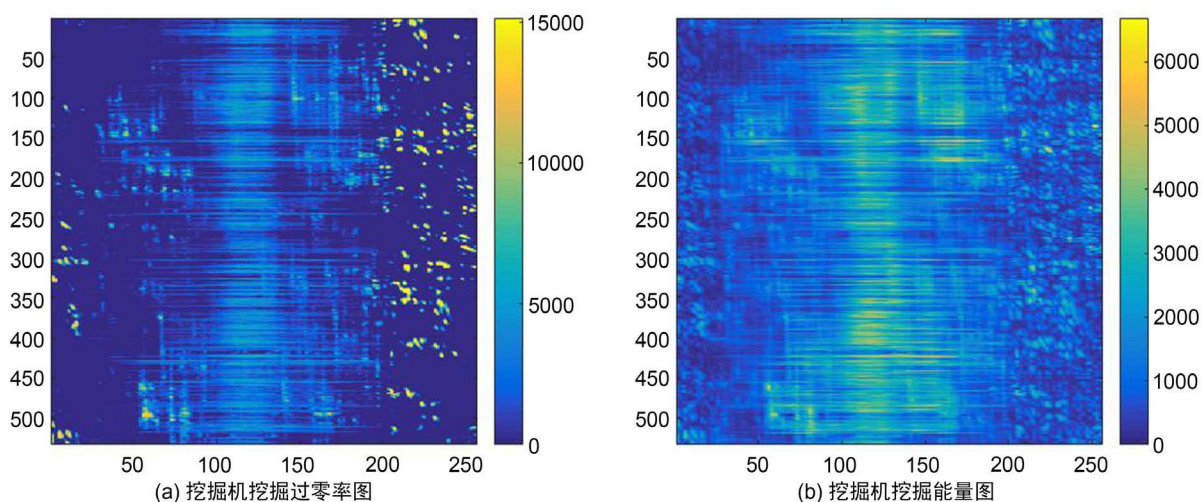


Figure 2. Signal diagram of opticfiber vibration caused by excavator's excavation

图 2. 光纤振动传感器采集的由挖掘机挖掘事件导致的二维时空信号图

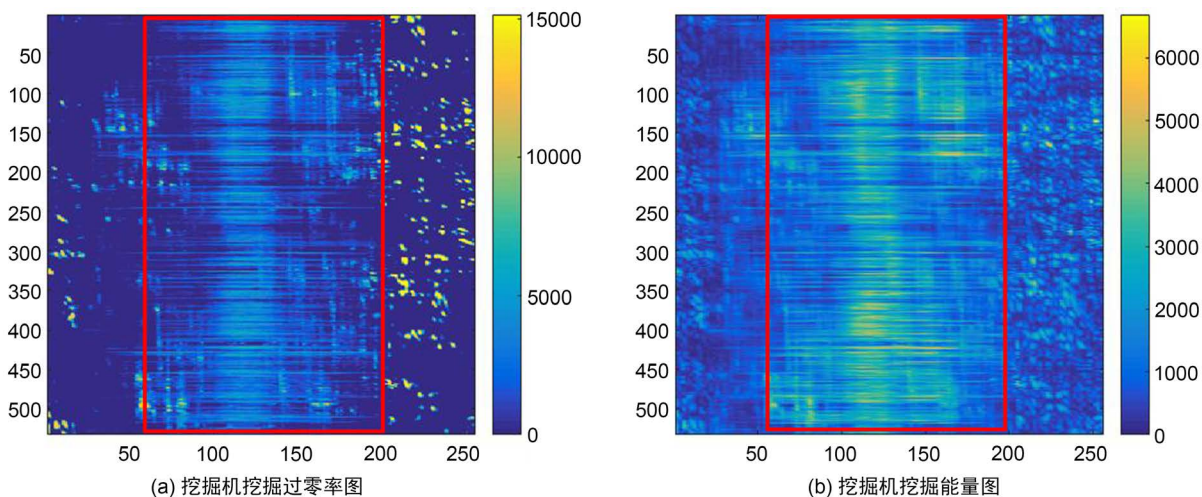


Figure 3. Segmentation of foreground signal region

图 3. 前景信号区域分割

3.2. 样本维数的数学意义教学

在将数据信号从背景中分割出来之后, 下一步涉及到的是数据样本的制作问题: 对于送入一个智能信息处理模型对其进行学习训练或测试的数据样本, 如何从数据信号区域中获得? 是选择整个数据信号区域(如图 3 的红框)内的所有数据作为一个样本, 还是选择区域中的部分数据形成一个样本? 被选定用来形成一个数据样本的数据及其个数可以用一个数据样本向量 $\vec{A} = (a_1, a_2, \dots, a_n)$ 和它的维数 n 表示, 这样问题转化为: 样本在数据信号前景区域的什么地方取以及 n 取多大值?

对于初学者而言, 往往不知道数据样本应该如何选取, 包括样本在数据区域中的位置、大小和数量的选定。为了向学生讲明样本制作的要点和依据, 我们借用《线性代数》课程的相关理论进行说明。

《线性代数》课程中关于线性方程组求解的内容涉及到“解的唯一性”条件问题: 对于一个 n 元一次方程组, 如果方程的个数小于未知量的个数 n , 那么该方程组有无穷多解; 如果方程的个数等于未知变量的个数 n (且系数向量都线性无关), 那么该方程组有唯一解。

以对数据样本进行智能信息处理的两层神经网络模型为例(模型结构参见图 5), n 个输入神经元负责输入任意一个样本向量 $\vec{A} = (a_1, a_2, \dots, a_n)$, 同时, n 个输入神经元与输出神经元相连, 它们之间有 n 个待求解的连接权系数 $\vec{X} = (x_1, x_2, \dots, x_n)$, 那么对应任意一个输入样本 $\vec{A}$, 输出神经元的响应值为神经网络对输入样本的加权和:

$$\vec{A}^T \vec{X} = a_1 x_1 + a_2 x_2 \cdots a_n x_n = c$$

如果要获得上述方程中的 n 个未知参数 $(x_1, x_2, \dots, x_n)$ 的唯一解, 根据上文中线性代数的知识, 需要有至少 n 个数据样本 $(\vec{A}_1, \vec{A}_2, \dots, \vec{A}_n)$ 输入到该神经网络模型中, 产生由 n 个对应的方程组成的方程组并对之求解:

$$\begin{aligned}\vec{A}_1^T \vec{X} &= a_{11} x_1 + a_{12} x_2 \cdots a_{1n} x_n = c_1 \\ \vec{A}_2^T \vec{X} &= a_{21} x_1 + a_{22} x_2 \cdots a_{2n} x_n = c_2 \\ &\dots\dots\dots \\ \vec{A}_n^T \vec{X} &= a_{n1} x_1 + a_{n2} x_2 \cdots a_{nn} x_n = c_n\end{aligned}$$

由上述方程组可知, 对应输入样本的数据向量 $\vec{A} = (a_1, a_2, \dots, a_n)$ 的维数 n 设定为多少, 那么需要的样本数同样为多少, 即需要的样本数目为相同大小的 n 。

以图 3 中所示的数据为例, 红色方框中所含的数值个数约为 $500 \times 150 = 75,000$ 个, 如果把这 75,000 个数值作为一个样本 $\vec{A} = (a_1, a_2, \dots, a_n)$, 那么该样本的维数 $n = 75,000$ 。对应的神经网络的输入神经元的个数同样需要设定为 75,000 个, 同时需要有 75,000 个连接权系数 $(x_1, x_2, \dots, x_{75000})$ 对输入样本向量作加权求和形成对应的一个方程:

$$\vec{A}^T \vec{X} = a_1 x_1 + a_2 x_2 \cdots a_{75000} x_{75000} = c$$

根据前述方程组求解的要求, 需要有 75,000 个图 3 所示的红色方框样本, 形成对应的 75,000 个方程构成的方程组, 才能求解出神经网络模型中从输入神经元到输出神经元之间的 75,000 个连接权未知系数 $(x_1, x_2, \dots, x_{75000})$ 的唯一解。

然而, 在有些实际的工程项目中, 很难采集到如此多的数据样本。例如, 图 3 显示的是挖掘机执行一次挖土动作所产生的光纤震动数据, 要获得 75,000 个这样的样本数据, 需要挖掘机执行 75,000 次挖掘动作, 这样的操作很难达到。通常情况下, 操作挖掘机以获得相应的数据, 次数往往非常有限, 例如可能最多只能重复 100 次这样的操作, 换句话说, 只能获得 100 张图 3 所示的数据。如果将图 3 中红色方框内的 75,000 个数值作为一个样本, 其对应的神经网络模型需要求解 75,000 个未知参数, 而所能提供的样本数只有 100 个, 由于 100 远小于 75,000, 也就是说由 100 个样本对应的 100 个方程所组成的方程组去求解方程中的 75000 个未知权系数, 将产生不确定性的无穷多解, 无法获得神经网络模型参数较精确的唯一解。

解决的办法就是缩小样本向量的维数, 令维数近似等于可获得的样本个数, 从而求取较精确的近似唯一解。例如, 如图 4 所示, 将红色方框内的数据划分成数据小块, 每个小块设为 10×10 大小, 这样每个小块内含有 100 个数值; 将每个小块作为一个数据样本, 那么该样本对应的数据向量的维数 $n = 100$; 对应的神经网络模型的输入神经元的个数或者待求解的从输入层到输出神经元的未知加权系数也变为 100 个大小; 在这种情况下, 只需要制作生成 100 个样本, 建立一个由 100 个方程组成的方程组, 就可以实现对神经网络模型中 100 个未知加权参数的较精确的求解。

采样如上的方式教学, 由学生的反应可以看出, 他们较容易地掌握了样本大小和样本个数的关系, 同时对样本制作和生成有了形象直观的理解和认识。

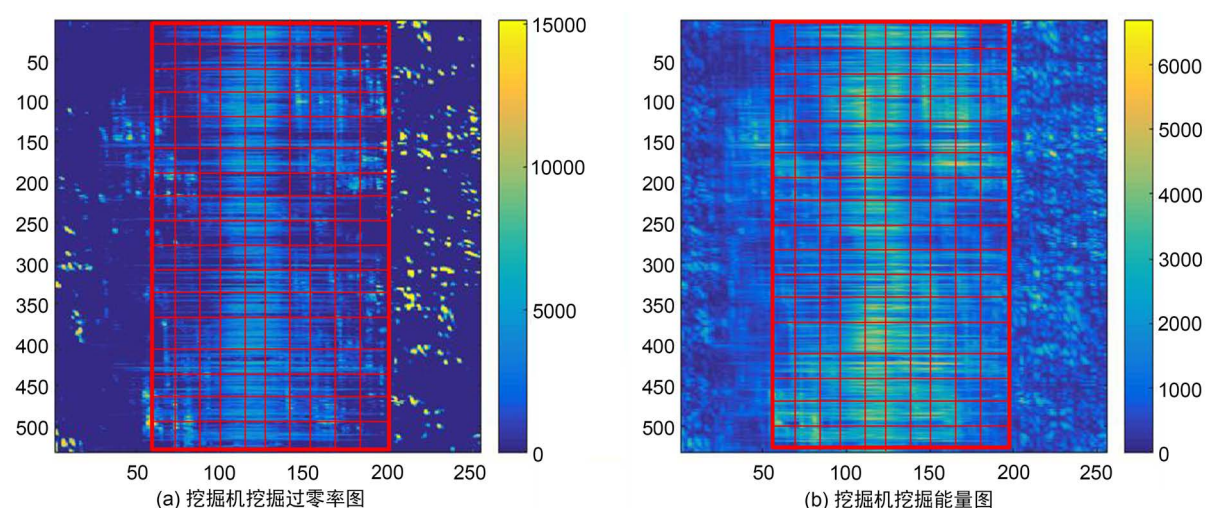


Figure 4. Generating data samples by dividing the foreground signal region into blocks

图 4. 前景信号区域分割成小块用于生成样本

3.3. 数据分类的几何意义教学

在对数据样本的生成原理和实践方法有了一个清晰的认识之后，下一步教师需要告诉学生输入数据样本的神经网络模型是如何对数据样本进行分类或决策的。

如图 5 所示，在以前的教学 PPT 中，我们会将两层神经网络或感知机的模型结构、结点所代表的数学意义、以及模型对应的分类或决策函数公式列出来供学生们领会学习。但学生们反映看了这些公式“知其然不知其所以然”。

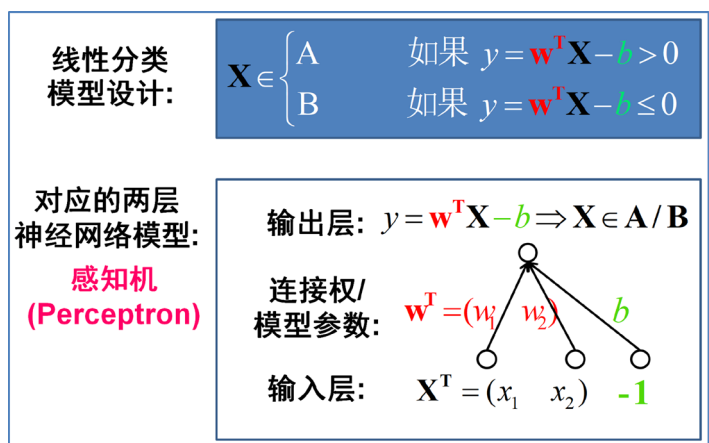


Figure 5. Original PPT
图 5. 原 PPT 教学设计

为此，我们增加了若干页 PPT，从不同类别数据样本在空间的分布及神经网络输出结点函数对它们的划分入手，对神经网络的基本原理进行形象直观的几何意义解释。具体如图 6 所示，在我们的教学实例中，不妨将挖掘机挖掘事件所产生的数据样本设为 A 类样本，这些样本在简化的数据空间中的分布如图中的红点所示；将其它事件如背景噪声产生的数据样本设为 B 类样本，这些样本在数据空间中的分布如图中的蓝点表示。当神经网络输出层结点的输出函数 $y = \mathbf{w}^T \mathbf{X} - b = 0$ 时，此时的神经网络描述的等式方程恰好对应于图中的绿色分割线，其物理意义是此时的样本落在绿线上，该样本处于 A 类和 B 类样本

的中间过渡位置。表达式中的 $\mathbf{w}^T \mathbf{X}$ 表示输入的数据样本 $\mathbf{X}$ 与神经网络连接权向量 $\mathbf{w}$ 所代表的 A 类事件或模板向量进行内积或相似度计算，其几何意义是数据样本向量的长度向 A 类事件模板向量上做垂直投影。当神经网络输出层结点的输出函数 $y = \mathbf{w}^T \mathbf{X} - b > 0$ 或 ≤ 0 时，此时的神经网络描述的不等式方程对应于图中绿色分割线上方或下方的区域，其物理意义是此时的样本落在绿线上方或下方，该样本处于 A 类事件样本或 B 类事件样本所在的区域，即数据样本 $\mathbf{X}$ 被分类为 A 类或 B 类。

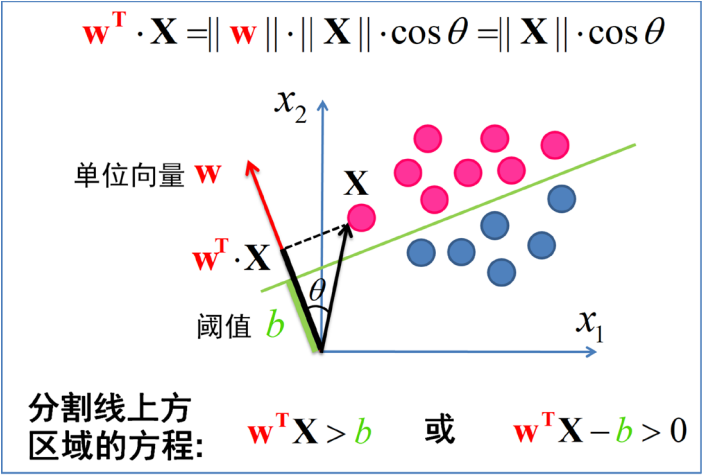


Figure 6. Added PPT for geometrical explanation
图 6. 新增的直观几何意义教学 PPT

经过图 6 形象直观的原理解释，学生们反映，对神经网络机器学习模型的分类功能有了本质的理解和掌握。

4. 结束语

综上所述，在计算机系统项目综合实践的课程教学中，为了能较好地培养学生解决复杂工程问题的能力，我们根据学生对课程教学效果的反馈，在教学中引入了几点教学改进措施，强调从数据可视化、数学理论类比解析和智能信息处理模型几何意义解释这几个角度，对复杂工程问题的解决进行循序渐进、由浅入深的分析和剖析，得出既有科学依据又易于理解掌握的分析 and 处理问题的方式方法，使得学生的学习和实践效果得到改善。随着现代科学知识的增多和工程问题的愈加复杂化，我们的教学仍旧面临挑战。今后在教学改进研究上，我们仍将不断探索，摸索总结出使学生获得更好实践效果的教学方法。

基金项目

北京信息科技大学 2019 年教改项目(2019JGYB16)。

参考文献

- [1] 成江荣. “案例引导, 项目驱动”模式下的计算机综合实践能力提升策略[J]. 软件导刊, 2015(2): 190-192.
- [2] 王健, 黄敏. “兴趣导向 - 任务驱动 - 能力提升”模式的统计教学改革与实践[J]. 教育教学论坛, 2016(15): 275-276.
- [3] 魏恒义, 安健, 鲍军鹏. 综合实践课程对本科生系统能力的培养[J]. 计算机教育, 2017(12): 86-89.
- [4] 赵月秀. 重视形象直观教学, 培养抽象思维能力[J]. 青海师专学报, 1999(S1): 87.