

# OBE视域下《数据结构》思政教学案例探究 ——以“串的模式匹配算法”为例

邱骏达, 蔡秋茹\*, 史培中, 由从哲

江苏理工学院计算机工程学院, 江苏 无锡

收稿日期: 2024年2月8日; 录用日期: 2024年3月6日; 发布日期: 2024年3月13日

## 摘要

目前, 大部分《数据结构》教材中选用的案例较为陈旧, 与国家发展战略所涉及的技术领域需求存在脱节现象。针对这一现状, 本文基于OBE教学理念, 聚焦计算机网络舆情管理的应用热点, 结合课程思政设计了一个“串的模式匹配算法”的教学案例。本案例采用真实网络热点事件评论文章为素材构建数据样本, 结合本教学团队公开发表在SCI期刊上的相关算法, 讲授建模思想、模型求解过程及信息后处理方法, 落实“研”为“教”所用的教学科研理念。本文提出的思政教学案例共需4个理论课时和2个实验课时, 学时安排合理, 另外, 该案例已在2022~2023学年第一学期相关专业班级进行实验教学, 得到了学生良好的学习反馈。

## 关键词

数据结构, OBE教学理念, 网络舆情管理, 课程思政, 建模思想

# Case Study on Ideological and Political Teaching of “Data Structure” from the Perspective of OBE

## —Taking “String Pattern Matching Algorithm” as an Example

Junda Qiu, Qiuru Cai\*, Peizhong Shi, Congzhe You

School of Computer Engineering, Jiangsu Institute of Technology, Wuxi Jiangsu

Received: Feb. 8<sup>th</sup>, 2024; accepted: Mar. 6<sup>th</sup>, 2024; published: Mar. 13<sup>th</sup>, 2024

\*通讯作者。

## Abstract

At present, most of the cases selected in the “Data Structure” textbook are relatively old, and there is a phenomenon of disconnection with the needs of the technical field involved in the national development strategy. In view of this situation, based on the OBE teaching concept, this paper designs a teaching case of “string pattern matching algorithm” by focusing on the application hot spots of computer network public opinion management and combining curriculum ideology and politics. This case uses real online hot event review articles as the material to build data samples, combined with the relevant algorithms published by the teaching team in SCI journals, to teach modeling ideas, model solving processes and information post-processing methods, and to implement the teaching and research concept used by “research” for “teaching”. The ideological and political teaching case proposed in this paper requires a total of 4 theoretical hours and 2 experimental hours, and the class time arrangement is reasonable. In addition, the case has been experimentally taught in relevant professional classes in the first semester of the 2022~2023 academic year, and good learning feedback has been obtained from students.

## Keywords

Data Structure, OBE Teaching Concept, Network Public Opinion Management, Curriculum Ideological and Political, Modeling Thought

Copyright © 2024 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

党的二十大的顺利召开进一步明确了国家实施科教兴国、人才强国战略的坚定决心。高等教育作为承担教育图强、创新制胜、人才引领光荣使命的重要一环，需要在日常教学中坚持以学生为中心、以成果为导向、以持续改进为重点的教育理念。落实到高校计算机专业的日常教学工作层面，针对学生理论知识的实践化能力培养尤为重要。

《数据结构》作为计算机专业极其重要的专业基础课，对学生的逻辑思维、编程能力、算法设计能力起到奠基意义。国内高校许多优秀教师对该课程的教学方法进行了深入的探究。于真[1]制定了“学习产出”式教学目标，依据教学目标制定相应教学内容并开展灵活多变的教学活动，同时，他也探讨了过程性考核的实施方法。王霞[2]等探讨了构建基于 CDT (案例驱动教学)的数据结构教学内容体系和学教结合的梯进式教学过程体系，通过满意度调查证明了所提教学改革方案的有效性。徐航[3]等以线上教学平台为基础，以实际教学经验为参考，提出了一种有助于提高学生的学习效率和积极性的教学改革方案。司国东[4]等分析了《数据结构》课程教学中存在的问题，提出了基于认知负荷理论的实验题库设计和教学方法，对更新教学案例进行了初步探究。然而，目前针对如何将高校教学团队科研成果与日常教学有机结合，如何利用理论知识解决国家战略发展面临的实际困难等问题研究成果较少。本文结合本教学团队科研成果，结合大数据环境下网络舆情信息分析问题，设计出了一个适用于“串的模式匹配算法”章节学习的教学案例。

## 2. 案例分析及数据建模

### 2.1. 案例引入

BP 和 KMP 是串的模式匹配问题中非常经典的两个算法，然而它们均只能实现串的完全匹配，并不能根据相似程度完成串的模糊匹配，在解决实际问题时很难得到准确的结果。本文结合大数据环境下网络舆情信息聚类问题，设计了相关的讲解案例。

**案例描述：**针对某一网络热点事件报道文章下的 3 万条用户评论进行识别，根据评论偏好信息将用户进行分类，并根据相关报道的作者观点将评论分为基本赞成、保持中立、基本反对三类，为进一步的舆情信息分析打下基础。

**案例分析：**由于网络用户表达意见方式存在差异，无法利用串的完全匹配算法来识别主串(用户评论)中是否存在模式串(作者观点)。另外，由于文章及评论中包含了大量与事件态度无关的表述，先对舆情信息进行特征向量提取可以大幅提高信息分析速度及准确率。最后，应先根据特征向量利用聚类算法对评论用户进行识别，再通过设置阈值来判断各分类与作者观点之间的匹配程度，从而完成最终的评论分类。

### 2.2. 数据建模

实际网络热点事件评论信息进行特征向量抽取，得到一个具有三个事件评判属性的特征向量数据集(此工作由教学团队预先利用 Transformer 模型进行提取，然后直接下发给学生进行建模)。引导学生将其转换成如图 1 所示的非线性网络舆情信息串(为了方便表述，该图只取 3 条信息串，每个串由 10 个特征值组成，其中黄色串为相关报道作者观点)：

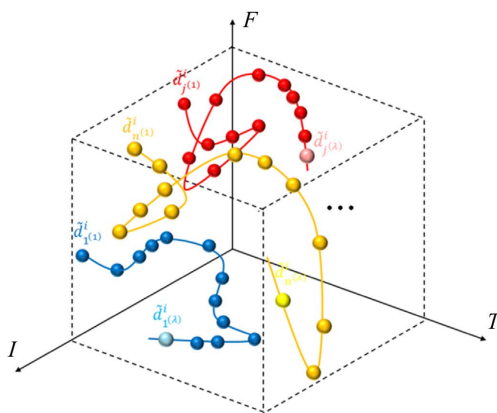


Figure 1. Non-linear network public opinion information string  
图 1. 非线性网络舆情信息串

不同用户观点特征值之间的差异被描述为三维空间中点的距离，距离越大说明用户观点分歧越大，反之亦然。

## 3. 模型求解

课堂教学中，模型求解需分解成：改进 nextval 数组生成函数，用户评论信息串聚类，用户评论信息串识别三个部分进行讲解。

### 3.1. 改进 nextval 数组生成函数

由于实际问题中，评论用户观点的特征值均存在差异，故需要对改良后的 KMP 算法(使用 nextval 数

组)进行进一步改进使之可以完成模糊匹配。当两个特征串某一对应位置的特征值点之间的距离小于阈值  $\mu$  时,即认为两者匹配。修改后的 nextval 数组求解函数如下所示:

```

Void get_nextval (String T, int *nextval)
    int i = 1; int k = 0;
    float d = 0; //不同特征值之间的距离
    nextval [1] = 0;
    while (I < T[0])
        D = get_distance (T[i], T[k]); //此为距离求解函数, 求解不同特征值之间的距离
        if (k == 0 || d ≤ 0.35) //本例中模糊匹配阈值为 0.35
            ++ i;
            ++ k;
            if (d > 0.35)
                nextval [i] = k;
            else
                nextval [i] = nextval [k];
        else
            k = nextval [k];

```

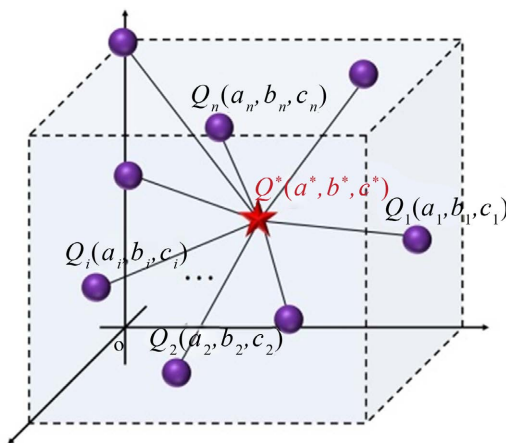
### 3.2. 用户评论信息串聚类

首先, 让学生理解最优集结点(聚类中心)的定义。

**定义 1:** 设  $n$  维空间闭区域内点集  $Q = \{Q^1, Q^2, \dots, Q^m\}$  ( $m \geq 3$ ) 有  $m$  个带权重的点  $Q^i(a_1^i, a_2^i, \dots, a_n^i)$ ,  $Q^i$  的权重  $\zeta^i \in [0, 1]$ , 且  $\sum_{i=1}^m \zeta^i = 1$ 。若存在点  $Q^*(a_1^*, a_2^*, \dots, a_n^*)$  与  $Q$  内所有点之间的加权 Euclidean 距离之和  $D_{Q^*}$  最小, 即

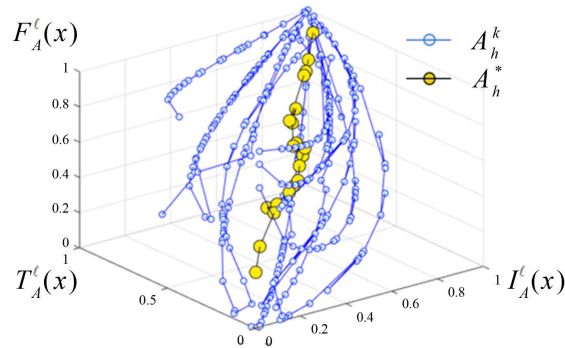
$$D_{Q^*} = \min \sum_{i=1}^m \left( \zeta_i \sqrt{(a_1^* - a_1^i)^2 + (a_2^* - a_2^i)^2 + \dots + (a_n^* - a_n^i)^2} \right) \quad (1)$$

则称  $Q^*(a_1^*, a_2^*, \dots, a_n^*)$  为点集  $Q$  的  $n$  维空间最优集结点。三维空间最优集结点如图 2 所示:



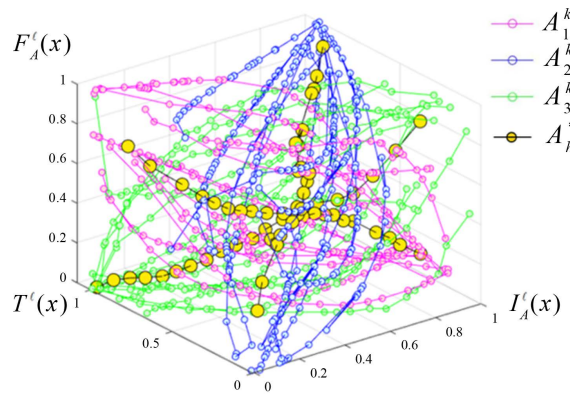
**Figure 2.** Optimal assembly point in three-dimensional space  
**图 2.** 三维空间最优集结点

在此基础上，向学生下发教学团队开发的自适应算法[5] [6] [7]代码，让学生结合改进后的 KMP 算法对自适应算法进行修改，计算出如图 3 所示的实际案例中网络用户的综合评论信息串：



**Figure 3.** Web user comprehensive review information string  
**图 3.** 网络用户综合评论信息串

根据本项目所选案例的要求及实际情况，设置 2 个阈值，将用户评论信息串与综合评论信息串之间的距离与阈值进行比较，将用户评论信息划分为 3 类，并再次调用改进后的自适应算法求解出如图 4 所示每一类的综合评论信息串：



**Figure 4.** Three types of user comprehensive comment information string  
**图 4.** 三类用户综合评论信息串

### 3.3. 用户信息评论串识别

选取 2.2 中各类综合评论信息串与报道作者信息串进行比对，利用基于投影理论的模糊信息投影公式计算出各类综合评论信息串在报道作者信息串上的投影，投影值越大，证明此类评论信息对作者观点的赞成度越高。计算过程如下：

**定义 2:** 设  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$  和  $\beta = (\beta_1, \beta_2, \dots, \beta_m)$  是两个向量，则

$$\text{Pr}_{j_\beta}(\alpha) = |\alpha| \cos \langle \alpha, \beta \rangle = \sqrt{\sum_{i=1}^m \alpha_i^2} \frac{\sum_{i=1}^m \alpha_i \beta_i}{\sqrt{\sum_{i=1}^m \alpha_i^2} \sqrt{\sum_{i=1}^m \beta_i^2}} = \frac{\sum_{i=1}^m \alpha_i \beta_i}{\sqrt{\sum_{i=1}^m \beta_i^2}} = \frac{\alpha \beta}{|\beta|} \quad (2)$$

称为向量  $\alpha$  在向量  $\beta$  上的投影，其中  $|\alpha| = \sqrt{\sum_{i=1}^m \alpha_i^2}$  和  $|\beta| = \sqrt{\sum_{i=1}^m \beta_i^2}$  是向量  $\alpha$  和  $\beta$  的模。显然， $\text{Pr}_{j_\beta}(\alpha)$

的值越大, 向量  $\alpha$  和向量  $\beta$  越接近。投影过程如图 5 所示:

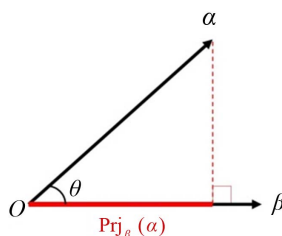


Figure 5. The projection of  $\alpha$  vector onto  $\beta$  vector

图 5. 向量  $\alpha$  在向量  $\beta$  上的投影

定义 3: 设  $A = (a_{ij})_{m \times n}$  和  $B = (b_{ij})_{m \times n}$  是两个矩阵, 则

$$\text{Prj}_B(A) = \frac{\sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij}}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n b_{ij}^2}} \quad (3)$$

称为矩阵  $A$  在矩阵  $B$  上的投影。显然,  $\text{Prj}_B(A)$  越大, 矩阵  $A$  和矩阵  $B$  越接近。

## 4. 案例仿真

### 4.1. 实验指导过程

教学团队成员选取自媒体平台上的网络热点事件报道文章及其评论信息为研究对象, 利用 transformer 对文章本身及 2 万条评论信息进行特征提取并将特征值做成 txt 文档, 作为教学数据集发放给学生。要求学生分组讨论, 利用理论课上学习到的知识对 KMP 算法代码进行修改, 形成改进后的可以完成模糊匹配的模糊 KMP 算法。完成对所有小组代码检查工作后, 由教师下发高维聚类自适应算法代码, 让学生们将自己开发的模糊 KMP 算法嵌入到高位聚类算法中, 并对数据集进行处理得到最终结果。各组展开实验结果横向比对, 由指导老师进行点评, 同时带领同学们进行实验总结。

### 4.2. 部分仿真结果

利用 2.1 与 2.2 中提到的算法对数据集进行处理, 得出的三个评论类的综合评论矩阵和报道作者信息向量如下所示:

$$A_1 = \left[ \begin{array}{cccccc} (0.28, 0.46, 0.28) & (0.27, 0.44, 0.27) & (0.30, 0.47, 0.31) & (0.34, 0.47, 0.33) & (0.33, 0.42, 0.28) \\ (0.30, 0.45, 0.29) & (0.31, 0.49, 0.31) & (0.27, 0.48, 0.29) & (0.33, 0.44, 0.33) & (0.30, 0.46, 0.31) \end{array} \right]$$

$$A_2 = \left[ \begin{array}{cccccc} (0.10, 0.27, 0.61) & (0.08, 0.28, 0.59) & (0.14, 0.24, 0.60) & (0.09, 0.23, 0.64) & (0.10, 0.25, 0.59) \\ (0.12, 0.27, 0.64) & (0.14, 0.26, 0.61) & (0.09, 0.25, 0.57) & (0.09, 0.23, 0.61) & (0.10, 0.22, 0.64) \end{array} \right]$$

$$A_3 = \left[ \begin{array}{cccccc} (0.65, 0.18, 0.22) & (0.62, 0.24, 0.27) & (0.68, 0.23, 0.22) & (0.65, 0.23, 0.27) & (0.67, 0.23, 0.23) \\ (0.64, 0.20, 0.24) & (0.66, 0.19, 0.25) & (0.66, 0.20, 0.21) & (0.67, 0.18, 0.28) & (0.68, 0.23, 0.23) \end{array} \right]$$

$$A^* = \left[ \begin{array}{cccccc} (0.2074, 0.2517, 0.6054) & (0.1989, 0.2148, 0.5847) & (0.2470, 0.2373, 0.6185) \\ (0.2164, 0.2186, 0.5712) & (0.2076, 0.2524, 0.5575) & (0.2089, 0.2366, 0.5821) \\ (0.2061, 0.1939, 0.6216) & (0.1955, 0.2289, 0.5824) & (0.2339, 0.2451, 0.5739) \\ (0.1952, 0.2113, 0.5830) \end{array} \right]$$



利用公式(2)和公式(3)计算出各类评论信息与文章观点相似度如下所示:

$$S(A_1) = -0.186, S(A_2) = 0.0549, S(A_3) = -0.027$$

显然,  $A_2$ 类持基本赞同观点,  $A_1$ 类持基本反对观点,  $A_3$ 类持中立观点。

## 5. 结语

《数据结构》是计算机科学与技术、软件工程、数据与大数据科学等专业的一门极其重要的基础课,课程所涉及的知识点是人工智能、大数据计算等前沿技术的理论基础,在工程实际问题中应用非常广泛。传统授课过程中讲解的案例较为陈旧,与现实中的新兴问题存在脱节现象。本文基于 OBE 教学理念,结合教学团队科研成果,对融入课程思政元素、紧跟国家发展战略需求的案例教学方式探索,提出了一个“串的模式匹配算法”的教学案例。

## 基金项目

- 1) 江苏理工学院 2022 年校教学改革与研究项目“OBE 视域下课程思政素材库与案例库的建设研究——以数据结构课程为例”。项目号: 11610312306;
- 2) 江苏省高等教育教改研究项目“新工科背景下计算机公共基础课程教学改革与探索”。项目号: 2021JSJG658;
- 3) 江苏省高等教育学会专项课题“基于产教融合的计算机类本科专业综合实训教学探索与研究”。项目号: 2020JDKT135;
- 4) 江苏理工院校级一流课程、课程思政示范课、产教融合课程——数据结构。

## 参考文献

- [1] 于真. 基于 OBE 理念的数据结构与算法课程教学改革研究[J]. 科技风, 2023(19): 123-125.
- [2] 王霞, 徐永华, 胡彩平. 基于 CBE 的地方本科院校数据结构教学改革[J]. 计算机教育, 2023(6): 116-120.
- [3] 徐航, 谢莹, 严涛, 等. 基于混合式教学的数据结构课程改革与探索[J]. 高教学刊, 2023, 9(16): 148-151+157.
- [4] 司国东, 彭利民, 张义青. 认知负荷视角下的数据结构实验题库设计[J]. 计算机教育, 2023(4): 173-176.
- [5] Qiu, J.D. and Li, L. (2017) A New Approach for Multiple Attribute Group Decision Making with Interval-Valued Intuitionistic Fuzzy Information. *Applied Soft Computing*, **61**, 111-121. <https://doi.org/10.1016/j.asoc.2017.07.008>
- [6] Qiu, J.D. and Li, L. (2018) Evaluation Criterion for Different Methods of Multiple-Attribute Group Decision Making with Interval-Valued Intuitionistic Fuzzy Information. *KSI Transactions on Internet and Information Systems*, **12**, 3128-3149. <https://doi.org/10.3837/tiis.2018.07.009>
- [7] Qiu, J.D. and Li, L. (2019) A New Approach for Multiple Attribute Group Decision-Making Based on Interval Neutrosophic Sets. *Journal of Intelligent & Fuzzy Systems*, **8**, 1-14.