

一类矩阵方程最小二乘问题及其在跨语言情感分类中的应用

陈秋娟, 袁仕芳*

五邑大学数学与计算科学学院, 广东 江门

收稿日期: 2025年7月2日; 录用日期: 2025年8月5日; 发布日期: 2025年8月29日

摘要

本文利用矩阵拉直算子、Moore-Penrose广义逆得到矩阵方程 $AXB = C$ 极小范数最小二乘解的表达式, 进一步得到矩阵方程 $AXB = C$ 对称极小范数最小二乘解的表达式, 最后讨论矩阵方程 $WX = Y$ 在跨语言情感分类中的应用。

关键词

矩阵方程, 最小二乘解, Moore-Penrose广义逆, 跨语言情感分类

The Least Squares Problem for a Kind of Matrix Equation and Its Applications in Cross-Lingual Sentiment Classification

Qiujuan Chen, Shifang Yuan*

School of Mathematics and Computation Science, Wuyi University, Jiangmen Guangdong

Received: Jul. 2nd, 2025; accepted: Aug. 5th, 2025; published: Aug. 29th, 2025

Abstract

This paper derives the expression of the linear least squares solution for the matrix equation $AXB = C$ by using the matrix straightening operator and the Moore-Penrose generalized inverse, obtains the expression of the symmetric least squares solution for the matrix equation $AXB = C$ with the least norm. Finally, the applications of the matrix equation $WX = Y$ in cross-lingual sentiment classification are discussed.

*通讯作者。

文章引用: 陈秋娟, 袁仕芳. 一类矩阵方程最小二乘问题及其在跨语言情感分类中的应用[J]. 理论数学, 2025, 15(8): 111-122. DOI: 10.12677/pm.2025.158226

Keywords

Matrix Equation, Least Squares Solution, Moore-Penrose Generalized Inverse, Cross-Lingual Sentiment Classification

Copyright © 2025 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

随着电商平台的蓬勃发展,人们在使用电商平台的时候产生了大量的评论数据。这些评论数据包含了用户对该商品带有情感信息的评价,我们可以根据这些评论数据去评估这件商品的实际情况。这类评价数据对于研究文本情感分类有很大的帮助。

不同国家和地区的发展水平不平衡,导致了一些国家和地区的情感资源匮乏,从而产生了某些语言的情感资源缺少大量标注的问题。互联网用户所使用的语言是多样的,以及某些地区语言情感资源匮乏的原因,跨语言情感分类问题应运而生。跨语言情感分类顾名思义为利用源语言语料去训练分类器,利用训练好的分类器去对目标语言语料进行分类,从而实现利用源语言具有的丰富情感资源的特性帮助情感资源较少的目标语言做分类的目的。

近年来,跨语言情感分类技术和方法推陈出新,最小二乘法是很多算法的理论基础之一,并且最小二乘法还称为数理统计学之灵魂[1]。研究发现,源语言和目标语言之间的映射关系是解决跨语言情感分类问题的关键。Mikolov [2]和彭晓娅[3]通过线性变换,将英语和西班牙语的词向量对齐,形成一个跨语言的共享特征空间。并且源语言词向量和目标语言中对应的词向量可建立简单的线性对应关系:对源语言中任意给定的单词向量 x , 目标语言中对应的单词向量 y 存在一种简单线性关系 $y = Wx$ 。它对应的最小二乘问题是求 $W = W^*$, 使得

$$\min \sum_{i=1}^n \|Wx_i - y_i\|_2^2 = \sum_{i=1}^n \|W^*x_i - y_i\|_2^2, \quad (1)$$

其中 x_i 指源语言的第 i 个词向量, y_i 指目标语言中对应的第 i 个词向量。

实际上该最小二乘问题还可以化为如下矩阵形式的最小二乘问题:求 $W = W^*$, 使得

$$\min \|WX - Y\|_F^2 = \|W^*X - Y\|_F^2, \quad (2)$$

其中矩阵 $X = [x_1, x_2, \dots, x_n]$ 是由 n 个源语言训练集词向量组成的矩阵, 矩阵 $Y = [y_1, y_2, \dots, y_n]$ 是由 n 个对应目标语言训练集词向量组成的矩阵, $\|\cdot\|_F$ 是矩阵的 Frobenius 范数。待求矩阵 W 即是矩阵方程 $WX = Y$ 的最小二乘解[4]。

Xing [5]发现 Mikolov [6]的回归映射方法存在一些问题,他们提出了对词向量进行归一化处理来解决距离定义计算不一致问题。故 Xing 将 $WX = Y$ 方程中的映射矩阵 W 定义为正交映射矩阵。并且 Smith [7]在理论上证明正交约束的可行性。与此同时,不少学者研究跨语言情感分类模型并提出了分类模型方法。其中, Faruqui [8]利用典型相关分析(CCA)方法实现。Artetxe [9]训练正交变换用迭代的方法实现。Hauer [10]利用自助法(Bootstrapping)迭代实现, Conneau [11]利用生成对抗网络实现。

较之矩阵方程 $WX = Y$ 更一般的方程是

$$AXB = C \quad (3)$$

其中 $A \in R^{m \times n}, B \in R^{n \times s}, C \in R^{m \times s}, X \in R^{n \times n}$ 。袁仕芳[12]利用矩阵的 Kronecker 积和 Moore-Penrose 广义逆研究矩阵方程 $AXB + CYD = E$ 的对称极小范数最小二乘解。周立平[13]利用矩阵的奇异值分解, 研究矩阵方程 $AX = B$ 在加权范数下的对称最小二乘解。李永勤[14]利用矩阵对的商奇异值分解(QSVD), 研究矩阵方程 $AXB = C$ 的对称最小二乘解。Peng Z. Y. [15]通过迭代法求矩阵方程 $AXB = C$ 的对称最小二乘解。Peng Z. Y. [16]提出两种新的矩阵形式迭代方法来求矩阵方程 $AXB + CYD = E$ 的对称最小二乘解。Deng Y. B. [17]利用矩阵的商奇异值分解研究矩阵方程 $BXA^T = T$ 的对称最小二乘解。

本文在上述研究工作的基础上, 利用矩阵拉直算子、Moore-Penrose 广义逆得到矩阵方程 $AXB = C$ 极小范数最小二乘解的表达式, 进一步得到矩阵方程 $AXB = C$ 对称极小范数最小二乘解的表达式, 最后讨论矩阵方程 $WX = Y$ 在跨语言情感分类中的应用。

为了方便起见, 本文用 $R^{m \times n}$ 表示 $m \times n$ 阶实矩阵的集合, $SR^{n \times n}$ 表示全体 n 阶对称矩阵的集合, $OR^{n \times n}$ 表示 $n \times n$ 阶正交矩阵的集合, A^T 和 A^+ 表示 A 的转置和 Moore-Penrose 广义逆。在实矩阵 $A = (a_{ij})$ 中, $a_{ij} = a_{ji}$, 称 A 为实对称矩阵。对 $A, B \in R^{m \times n}$, 定义 A 和 B 的内积为 $\langle A, B \rangle = \text{tr}(B^T A)$ 。由此导出的矩阵 A 的 Frobenius 范数为 $\|A\| = \sqrt{\langle A, A \rangle}$ 。

定义 1.1. 设 $A \in (a_{ij})_{m \times n}$, 记 $a_i = (a_{1i}, a_{2i}, a_{3i}, \dots, a_{mi})$ ($i = 1, 2, 3, \dots, n$), 令

$$\text{vec}(A) = (a_1, a_2, a_3, \dots, a_n)^T. \quad (4)$$

定义 1.2. 设 $A = (a_{ij})_{n \times n}$, 记,

$$a_1 = (a_{11}, a_{21}, a_{31}, \dots, a_{n1}), a_2 = (a_{22}, a_{32}, a_{42}, \dots, a_{n2}), \dots, a_{n-1} = (a_{(n-1)(n-1)}, a_{n(n-1)}), a_n = a_{nn}.$$

定义

$$\text{vec}_s(A) = (a_1, a_2, a_3, \dots, a_n)^T \in R^{\frac{n(n+1)}{2}}. \quad (5)$$

定义 1.3. [18] 记 $x = (x_1, x_2, \dots, x_k)^T \in R^k$ 和 $y = (y_1, y_2, \dots, y_k)^T \in R^k$ 和 $A = (A_1, A_2, \dots, A_k)$, $A_i \in R^{m \times n}$, 定义

$$(1) A \odot x = x_1 A_1 + x_2 A_2 + \dots + x_k A_k \in R^{m \times n}; \quad (6)$$

$$(2) A \odot (ax + by) = a(A \odot x) + b(A \odot y); \quad (7)$$

$$(3) \text{假设 } M_1 \in R^{m \times s}, M_2 \in R^{m \times t}, M_3 \in R^{n \times s}, M_4 \in R^{n \times t}, x_1, x_2 \in R, \quad (8)$$

$$\begin{bmatrix} M_1 & M_2 \\ M_3 & M_4 \end{bmatrix} \odot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (M_1, M_2) \odot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ (M_3, M_4) \odot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{bmatrix}; \quad (9)$$

$$\begin{bmatrix} M_1 & 0 \\ 0 & M_4 \end{bmatrix} \odot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} M_1 x_1 \\ M_4 x_2 \end{bmatrix}. \quad (10)$$

本文主要研究如下两个问题:

问题 I 给定矩阵 $A \in R^{m \times n}$, $B \in R^{n \times s}$, $C \in R^{m \times s}$, 记

$$R_L = \left\{ X \mid X \in R^{n \times n}, \|AXB - C\|_F^2 = \min_{X_0 \in R^{n \times n}} \|AX_0 B - C\|_F^2 \right\}, \quad (11)$$

求 $X_R \in R_L$, 使得

$$\|X_R\| = \min_{X_0 \in R_L} \|X\|. \quad (12)$$

问题 II 给定矩阵 $A \in R^{m \times n}$, $B \in R^{n \times s}$, $C \in R^{m \times s}$, 记

$$S_L = \left\{ X \mid X \in SR^{n \times n}, \|AXB - C\|_F^2 = \min_{X_0 \in SR^{n \times n}} \|AX_0B - C\|_F^2 \right\}, \quad (13)$$

求 $X_S \in S_L$, 使得

$$\|X_S\| = \min_{X_0 \in S_L} \|X\|. \quad (14)$$

本文在第二节介绍若干个引理, 在第三节利用相关结果推导出问题 I 和问题 II 解的表达式。最后在第四节讨论矩阵方程 $WX = Y$ 在跨语言情感分类中的应用。

2. 几个引理

引理 2.1. [19] 设 $A \in R^{m \times n}$, $b \in R^n$, 则不相容线性方程组 $Ax = b$ 的最小二乘解为

$$x = A^+b + (I - A^+A)y, \quad (15)$$

其中 $y \in R^n$ 为任意的。

引理 2.2. [18] (i) 对 $X \in R^{n \times n}$, 则

$$X \in R^{n \times n} \Leftrightarrow X = K_n \odot \text{vec}(X) \quad (16)$$

$$K_n = [E_{11}, E_{12}, \dots, E_{1n}, E_{21}, E_{22}, \dots, E_{2n}, \dots, E_{n1}, E_{n2}, \dots, E_{nn}], \quad (17)$$

其中 $K_n \in R^{n \times n^3}$ 。

(ii) 对 $X \in R^{n \times n}$, 则

$$X \in SR^{n \times n} \Leftrightarrow X = K_s \odot \text{vec}_s(X) \quad (18)$$

$$K_s = [E_{11}, E_{21} + E_{12}, \dots, E_{n1} + E_{1n}, E_{22}, E_{32} + E_{23}, \dots, E_{n2} + E_{2n}, \dots, E_{(n-1)(n-1)}, E_{n(n-1)} + E_{(n-1)n}, E_{nn}]. \quad (19)$$

其中 $K_s \in R^{n \times n^2(n+1)/2}$ 。

引理 2.3. [18] 给定实矩阵 $A_1, A_2, A_3, \dots, A_l \in R^{m \times n}$ 和 $B \in R^{m \times n}$, 求 $k = (k_1, k_2, k_3, \dots, k_l)^T \in R^l$, 使得

$$k_1A_1 + k_2A_2 + k_3A_3 + \dots + k_lA_l = B. \quad (20)$$

如果矩阵方程(20)是相容的, 则满足矩阵方程的向量 k 的集合是以下相容方程的解集合,

$$\begin{bmatrix} \langle A_1, A_1 \rangle & \langle A_1, A_2 \rangle & \cdots & \langle A_1, A_l \rangle \\ \langle A_2, A_1 \rangle & \langle A_2, A_2 \rangle & \cdots & \langle A_2, A_l \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle A_l, A_1 \rangle & \langle A_l, A_2 \rangle & \cdots & \langle A_l, A_l \rangle \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_l \end{bmatrix} = \begin{bmatrix} \langle A_1, B \rangle \\ \langle A_2, B \rangle \\ \vdots \\ \langle A_l, B \rangle \end{bmatrix}. \quad (21)$$

如果矩阵方程(20)不相容, 则有矩阵 $A_1, A_2, A_3, \dots, A_l \in R^{m \times n}$ 和 $B \in R^{m \times n}$, 求 $k = (k_1, k_2, k_3, \dots, k_l)^T \in R^l$, 使得

$$\|k_1A_1 + k_2A_2 + k_3A_3 + \dots + k_lA_l - B\| = \min, \quad (22)$$

则该问题的解集为相容方程(21)的解集。

3. 主要结论

设矩阵 $A = (A_1, A_2, \dots, A_n) \in R^{m \times n}$, 其中 $A_i \in R^m$ 是矩阵 A 的第 i 列向量, $B = (B_1^T, B_2^T, \dots, B_n^T)^T \in R^{n \times s}$,

其中 $B_j \in R^s$ 是矩阵 B 的第 j 行向量, $C \in R^{m \times s}$. 令 $E_{ij} = (e_{st}) \in R^{n \times n}$, 其中

$$e_{st} = \begin{cases} 1, & (s, t) = (i, j), \\ 0, & \text{其它}, \end{cases}$$

$i, j = 1:n$. 可得

$$AE_{ij}B = A_i B_j.$$

令

$$T = \begin{bmatrix} \langle T_{11}, T_{11} \rangle & \langle T_{11}, T_{12} \rangle & \cdots & \langle T_{11}, T_{1n} \rangle & \cdots & \langle T_{11}, T_{nn} \rangle \\ \langle T_{12}, T_{11} \rangle & \langle T_{12}, T_{12} \rangle & \cdots & \langle T_{12}, T_{1n} \rangle & \cdots & \langle T_{12}, T_{nn} \rangle \\ \vdots & \vdots & & \vdots & & \vdots \\ \langle T_{nn}, T_{11} \rangle & \langle T_{nn}, T_{12} \rangle & \cdots & \langle T_{nn}, T_{1n} \rangle & \cdots & \langle T_{nn}, T_{nn} \rangle \end{bmatrix}, \quad e = \begin{bmatrix} \langle T_{11}, C \rangle \\ \langle T_{12}, C \rangle \\ \vdots \\ \langle T_{nn}, C \rangle \end{bmatrix}, \quad (23)$$

其中 $T_{ij} = A_i B_j$. 显然 $T \in R^{n^2 \times n^2}$, $e \in R^{n^2}$.

定理 3.1. 令 K_n 如式(17)所示, T, e 如式(23)所示, 则问题 I 的解集合

$$R_L = \left\{ X \mid X = K_n \odot \left[T^+ e + (I - T^+ T) y \right], \forall y \in R^{n^2} \right\}. \quad (24)$$

问题 I 存在唯一的极小范数最小二乘解

$$X = K_n \odot (T^+ e) \quad (25)$$

证明 由引理 2.2 可得

$$\begin{aligned} \|AXB - C\|_F^2 &= \|A[K_n \odot \text{vec}(X)]B - C\|_F^2 \\ &= \|\left[(AK_n) \odot \text{vec}(X)\right]B - C\|_F^2 \\ &= \|(AE_{11}B, AE_{12}B, \cdots, AE_{1n}B, \cdots, AE_{n1}B, AE_{n2}B, \cdots, AE_{nn}B) \odot \text{vec}(X) - C\|_F^2 \\ &= \|(T_{11}, T_{12}, \cdots, T_{1n}, T_{21}, T_{22}, \cdots, T_{2n}, \cdots, T_{n1}, T_{n2}, \cdots, T_{nn}) \odot \text{vec}(X) - C\|_F^2. \end{aligned}$$

由引理 2.1 和 2.3 可得

$$\text{vec}(X) = T^+ e + (I - T^+ T)y.$$

进一步可得式(25)。

由定理 3.1 和参考文献[18], 我们现在求问题 II 的解。令

$$F = \begin{bmatrix} \langle F_{11}, F_{11} \rangle & \cdots & \langle F_{11}, F_{n1} \rangle & \langle F_{11}, F_{22} \rangle & \cdots & \langle F_{11}, F_{n2} \rangle & \cdots & \langle F_{11}, F_{nn} \rangle \\ \langle F_{21}, F_{11} \rangle & \cdots & \langle F_{21}, F_{n1} \rangle & \langle F_{21}, F_{22} \rangle & \cdots & \langle F_{21}, F_{n2} \rangle & \cdots & \langle F_{21}, F_{nn} \rangle \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ \langle F_{n1}, F_{11} \rangle & \cdots & \langle F_{n1}, F_{n1} \rangle & \langle F_{n1}, F_{22} \rangle & \cdots & \langle F_{n1}, F_{n2} \rangle & \cdots & \langle F_{n1}, F_{nn} \rangle \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ \langle F_{n(n-1)}, F_{11} \rangle & \cdots & \langle F_{n(n-1)}, F_{n1} \rangle & \langle F_{n(n-1)}, F_{22} \rangle & \cdots & \langle F_{n(n-1)}, F_{n2} \rangle & \cdots & \langle F_{n(n-1)}, F_{nn} \rangle \\ \langle F_{nn}, F_{11} \rangle & \cdots & \langle F_{nn}, F_{n1} \rangle & \langle F_{nn}, F_{22} \rangle & \cdots & \langle F_{nn}, F_{n2} \rangle & \cdots & \langle F_{nn}, F_{nn} \rangle \end{bmatrix}, \quad (26)$$

$$f = \left[\langle F_{11}, C \rangle, \langle F_{21}, C \rangle, \cdots, \langle F_{n1}, C \rangle, \cdots, \langle F_{n(n-1)}, C \rangle, \langle F_{nn}, C \rangle \right]^T \quad (27)$$

其中

$$F_{ij} = \begin{cases} A_i B_j, & i = j, \\ A_i B_j + A_j B_i, & i > j. \end{cases}$$

显然 $F \in R^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}}$, $f \in R^{\frac{n(n+1)}{2}}$ 。

定理 3.2. 令 K_s 如式(19)所示, F, f 如式(26) (27)所示, 则问题 II 的解集合

$$S_L = \left\{ X \mid X = K_s \odot [F^+ f + (I - F^+ F) y], \forall y \in R^{\frac{n(n+1)}{2}} \right\}. \quad (28)$$

问题 II 存在唯一的极小范数最小二乘解

$$X = K_s \odot (F^+ f). \quad (29)$$

考虑到矩阵方程 $WX = Y$ 是矩阵方程 $AXB = C$ 的特殊形式, 我们分别利用定理 3.1 和定理 3.2 得到它的极小范数最小二乘解和对称极小范数最小二乘解的相关结论。设矩阵 $X = (X_1, X_2, \dots, X_n) \in R^{n \times s}$, 其中 $X_i \in R^n$ 是矩阵 X 的第 i 列向量, $Y = (Y_1^T, Y_2^T, \dots, Y_n^T)^T \in R^{n \times s}$, 其中 $Y_j \in R^s$ 是矩阵 Y 的第 j 行向量。令

$$H = \begin{bmatrix} \langle H_{11}, H_{11} \rangle & \langle H_{11}, H_{12} \rangle & \cdots & \langle H_{11}, H_{1n} \rangle & \cdots & \langle H_{11}, H_{nn} \rangle \\ \langle H_{12}, H_{11} \rangle & \langle H_{12}, H_{12} \rangle & \cdots & \langle H_{12}, H_{1n} \rangle & \cdots & \langle H_{12}, H_{nn} \rangle \\ \vdots & \vdots & & \vdots & & \vdots \\ \langle H_{nn}, H_{11} \rangle & \langle H_{nn}, H_{12} \rangle & \cdots & \langle H_{nn}, H_{1n} \rangle & \cdots & \langle H_{nn}, H_{nn} \rangle \end{bmatrix}, \quad g = \begin{bmatrix} \langle H_{11}, Y \rangle \\ \langle H_{12}, Y \rangle \\ \vdots \\ \langle H_{nn}, Y \rangle \end{bmatrix}, \quad (30)$$

其中 $H_{ij} = E_{ij} X$. 显然 $H \in R^{n^2 \times n^2}$, $g \in R^{n^2}$ 。

推论 3.1. K_n 如式(17)所示, H, g 如式(30)所示, 则矩阵方程 $WX = Y$ 最小二乘问题

$$\min_{W \in R^{n \times n}} \|WX - Y\| \quad (31)$$

解集合 R_E 表示为

$$R_E = \left\{ W \mid W = K_n \odot [H^+ g + (I - H^+ H) y], \forall y \in R^{n^2} \right\}, \quad (32)$$

它的极小范数最小二乘解

$$W = K_n \odot (H^+ g). \quad (33)$$

下面求矩阵方程 $WX = Y$ 的对称极小范数最小二乘解, 令

$$M = \begin{bmatrix} \langle M_{11}, M_{11} \rangle & \cdots & \langle M_{11}, M_{n1} \rangle & \langle M_{11}, M_{22} \rangle & \cdots & \langle M_{11}, M_{n2} \rangle & \cdots & \langle M_{11}, M_{nn} \rangle \\ \langle M_{21}, M_{11} \rangle & \cdots & \langle M_{21}, M_{n1} \rangle & \langle M_{21}, M_{22} \rangle & \cdots & \langle M_{21}, M_{n2} \rangle & \cdots & \langle M_{21}, M_{nn} \rangle \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ \langle M_{n1}, M_{11} \rangle & \cdots & \langle M_{n1}, M_{n1} \rangle & \langle M_{n1}, M_{22} \rangle & \cdots & \langle M_{n1}, M_{n2} \rangle & \cdots & \langle M_{n1}, M_{nn} \rangle \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ \langle M_{n(n-1)}, M_{11} \rangle & \cdots & \langle M_{n(n-1)}, M_{n1} \rangle & \langle M_{n(n-1)}, M_{22} \rangle & \cdots & \langle M_{n(n-1)}, M_{n2} \rangle & \cdots & \langle M_{n(n-1)}, M_{nn} \rangle \\ \langle M_{nn}, M_{11} \rangle & \cdots & \langle M_{nn}, M_{n1} \rangle & \langle M_{nn}, M_{22} \rangle & \cdots & \langle M_{nn}, M_{n2} \rangle & \cdots & \langle M_{nn}, M_{nn} \rangle \end{bmatrix}, \quad (34)$$

$$h = [\langle M_{11}, Y \rangle, \langle M_{21}, Y \rangle, \dots, \langle M_{n1}, Y \rangle, \dots, \langle M_{n(n-1)}, Y \rangle, \langle M_{nn}, Y \rangle]^T \quad (35)$$

其中

$$M_{ij} = \begin{cases} E_{ij} X, & i = j, \\ (E_{ij} + E_{ji}) X, & i > j. \end{cases}$$

显然 $M \in R^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}}$, $h \in R^{\frac{n(n+1)}{2}}$ 。

推论 3.2. K_s 如式(19)所示, M, h 如式(34)和(35)所示, 则矩阵方程 $WX = Y$ 最小二乘问题

$$\min_{W \in SR^{n \times n}} \|WX - Y\| \quad (36)$$

解集合 S_E 表示为

$$S_E = \left\{ W | K_s \odot [M^+ h + (I - M^+ M) y], \forall y \in R^{\frac{n(n+1)}{2}} \right\}, \quad (37)$$

它的极小范数最小二乘解

$$W = K_s \odot (M^+ h). \quad (38)$$

4. 应用

4.1. 实验设置

本文以英文为源语言, 以中文为目标语言。操作系统是 Windows11 64 位, 处理器是 Inter Core i7-13620H, 内存为 16 GB。采用的编程语言为 Python, 编辑器为 PyCharm。

4.2. 数据集

本文使用来自 NLP&CC2013 提供的跨语言情感分类评测任务数据集, 该数据集取自亚马逊电商平台的中英文书籍、DVD 以及音乐商品评论数据。该数据集每一类别中含有 2000 条正面评价和 2000 条负面评价, 在实验中按照一定比例随机抽取文本数据作为训练集和测试集。数据量如下表 1 所示:

Table 1. Comment on the text dataset

表 1. 评论文本数据集

数据集		未标注数据集(篇)	标注数据集(篇)
英语	BOOK	——	4000
	DVD	——	4000
	MUSIC	——	4000
中文	BOOK	47071	4000
	DVD	17814	4000
	MUSIC	29677	4000

4.3. 实验过程

我们下面描述矩阵方程 $WX = Y$ 的最小二乘解在跨语言情感分类中的应用。

算例 I 极小范数最小二乘解在跨语言情感分类中的应用

第一步, 数据预处理

提取文本数据, 进行文本预处理操作, 对所有文本做去特殊符号, 去数字等预处理。对英文文本做大写字母转小写字母, nltk 分词, 对中文文本做 jieba 分词。利用 Google Translate 将提取的文本评论翻译成一对应的中英文对应文本评论。

第二步, 获取中文和英文评论文本的词向量

选取 fastText 提供的开源预训练词向量, 预训练词向量在 Wikipedia 数据集上使用 fastText 训练得到的, 词向量维度为 300, 使用默认参数的 skip-gram 模型获得。加载预训练词向量之后, 将文本转化为词

向量, 运用 TF-IDF 算法计算文档词向量。忽略未出现在预训练词向量列表中的词, 对相应的词向量计算加权词向量。

第三步, 特征提取

在实验中仅使用文本嵌入特征, 为训练映射矩阵, 引入 MUSE 提供的中英双语词典。利用该中英双语词典以及一一对应的中英文文本数据, 根据双语词典, 找到数据集样本中对应的中英文词, 将训练集中的中文文本数据与英文文本数据平行样本进行对齐, 将对应的文本转化为加权词向量, 将对齐的中文和英文加权词向量分别组合为训练集词向量矩阵。

第四步, 求映射矩阵

使用最小二乘法训练线性映射矩阵, 使用十折交叉验证方法, 将数据集分成十份, 前九份作为训练集, 最后一份作为测试集。分别选取 3000 条中文和英文文本数据作为训练集和测试集样本数据, 为验证不同的词向量维度对实验的影响, 在实验中对词向量进行 pca 降维, 将词向量维度降维至 100 维和 150 维。

经过第一步至第三步的步骤之后, 得英文训练集组成的词向量矩阵 X , 得中文训练集组成的词向量矩阵 Y , 设映射矩阵为 W , $X \in R^{n \times s}$, $Y \in R^{n \times s}$, 求 $W = W^* \in R^{n \times n}$ 使得

$$\min \|WX - Y\|_F^2 = \|W^*X - Y\|_F^2.$$

第五步, 分类

对分类器进行训练, 采用的分类器为逻辑回归(LR), 支持向量机(SVM), 朴素贝叶斯(NB)。对源语言中某个任意的单词的向量 x , 目标语言中对应单词的向量 y 存在一种简单线性关系: $y = Wx$, 故可以通过映射矩阵变换到共享空间中。得到共享空间表示后, 应用第五步训练出来的分类器, 对目标语言数据进行分类。

算例 II 对称极小范数最小二乘解在跨语言情感分类中的应用

第一步及第二步同算例 I

第三步, 提取特征

仅使用文本嵌入特征, 对文本计算 TF-IDF 值得到加权文档词向量。为训练映射矩阵, 使用 kmeans 聚类算法, 通过计算样本特征间的距离, 将样本划分到 k 个聚类(簇)中, kmeans 聚类算法原理如下,

$$\arg \min_s \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|, \quad (39)$$

其中 S_i 表示被划分到簇 i 中的样本集合, μ_i 表示簇 i 的簇, x 表示样本。

在该实验中同样采用十折交叉验证的方法, 并且为验证词向量维度对实验结果的影响大小, 对原始词向量维度 pca 降维至 50 维和 100 维。在实验中, 由 kmeans 聚类算法分别对源语言训练集特征矩阵 X 和目标语言训练集特征矩阵 Y 分别做聚类, 分别将源语言和目标语言训练集划分为 50 个簇和 100 个簇, 每个簇代表特征空间中相似的文本样本群体。计算每个簇的簇中心词, 接着对于每个簇, 计算簇内所有词到簇中心的距离, 以欧几里得距离作为衡量标准, 选择距离最近的词作为代表性词。分别提取出 50 个和 100 个源语言和目标语言代表性词, 作为训练映射矩阵的源语言和目标语言词向量。对词向量做 pca 降维, 将词向量维度由 300 维降维 50 维和 100 维。其中词向量降维维度与聚类簇数保持一致。

第四步, 求映射矩阵

利用 Moore-Penrose 逆和拉直算子求矩阵方程 $WX = Y$ 的对称极小范数最小二乘解, 该解即为所求的映射矩阵。分别选取 3000 条中文和英文文本数据作为训练集和测试集样本数据, 分别做十折交叉验证,

经过第一步至第三步的步骤之后, 得到英文训练集组成的词向量矩阵 X , 得到中文训练集组成的词向量矩阵 Y 。设映射矩阵为 W , $X \in R^{n \times s}$, $Y \in R^{n \times s}$, 求 $W = W^* \in SR^{n \times n}$, 使得

$$\min \|WX - Y\|_F^2 = \|W^*X - Y\|_F^2. \quad (40)$$

存在唯一的解 $W \in S_E$,

$$W = K_S \odot (M^+h). \quad (41)$$

计算得到 W 之后, $W = (w_1, w_2, w_3, \dots, w_n)^T$, 列拉直为对称矩阵 W , 得到最终的映射矩阵。该映射矩阵的维度为(50, 50)和(100, 100)。

第五步, 分类

对分类器进行训练, 采用的分类器为逻辑回归(LR), 支持向量机(SVM)。通过对称映射矩阵变换, 得到共享空间表示后, 应用第五步训练出来的分类器, 对目标语言数据进行分类。

4.4. 实验结果分析

使用带标记的英文训练集对分类器进行训练, 利用映射矩阵将英文和中文变换到共享空间中, 在共享空间中利用训练好的分类器对中文进行分类。

在前人的研究上, 映射矩阵可以被约束为正交矩阵, 故本文以正交 **procrustes** 方法作为基线模型, 与本文模型进行对比。记源语言词向量为 x_i , 目标语言词向量为 y_i , 求映射矩阵 $W = W^* \in OR^{n \times n}$, 对所有的 $x_i \in X$ 和 $y_i \in Y$, 使得

$$\min \sum_{i=1}^n \|Wx_i - y_i\|_2^2 = \sum_{i=1}^n \|W^*x_i - y_i\|_2^2,$$

该问题又可以简化为正交 **procrustes** 方法,

$$W^* = \max_{tr} (XWY^T).$$

在本文的实验中, 采用了多个分类器做分类, 在算例 I 中的实验中发现多次改变样本数以及降维维度的过程中, 采用逻辑回归分类器对文本进行跨语言情感分类可以得到较优的结果, 因此在本文的实验中主要采取逻辑回归分类器, 实验结果如图 1 所示:

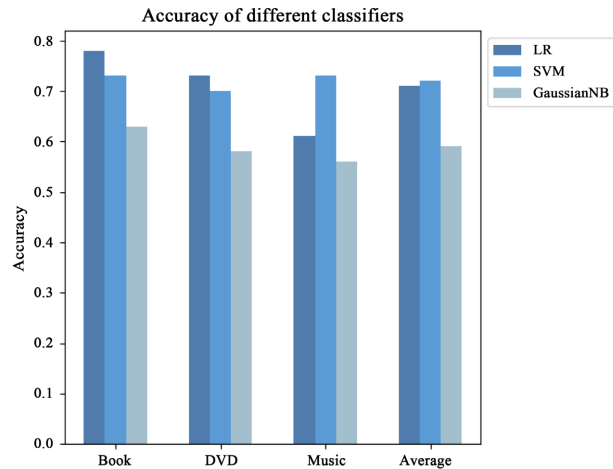


Figure 1. Classification accuracy results of different classifiers

图 1. 不同分类器分类准确率结果

算例 I 和正交 *procrustes* 方法在不同词向量维度下进行情感分类的实验结果如图 2 所示:

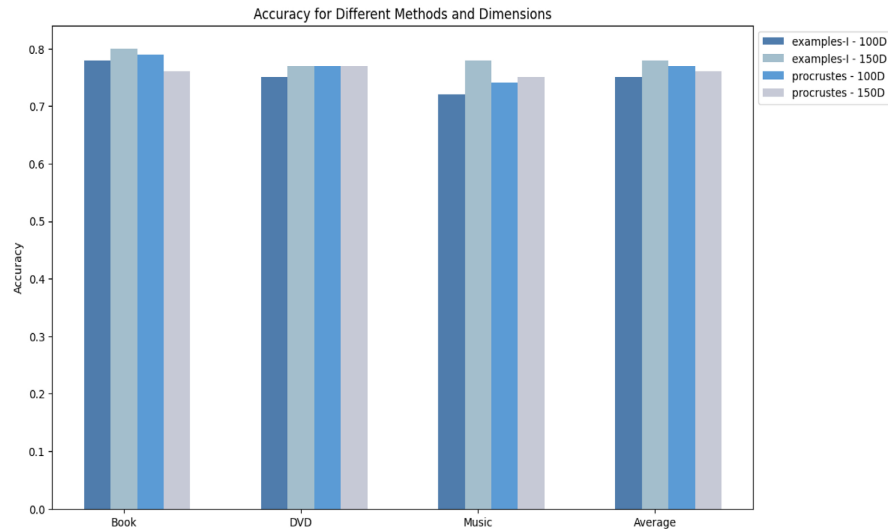


Figure 2. Classification results of examples I and orthogonal *procrustes* method
图 2. 算例 I 和正交 *procrustes* 方法分类结果

图 2 中分别为算例 I 和正交 *procrustes* 方法采用逻辑回归(LR)分类器, 在 BOOK、DVD、MUSIC 数据集上的情感分类结果。在实验中, 将词向量维度由 300 维降维至 100 维以及 150 维。由图中可以看出, 词向量维度降维至 150 维可以得到较优的分类结果。与此同时, 算例 I 在所有数据集上的表现都相对稳定, 尤其在 BOOK 数据集上表现最优。

算例 II 实验结果如图 3 所示:

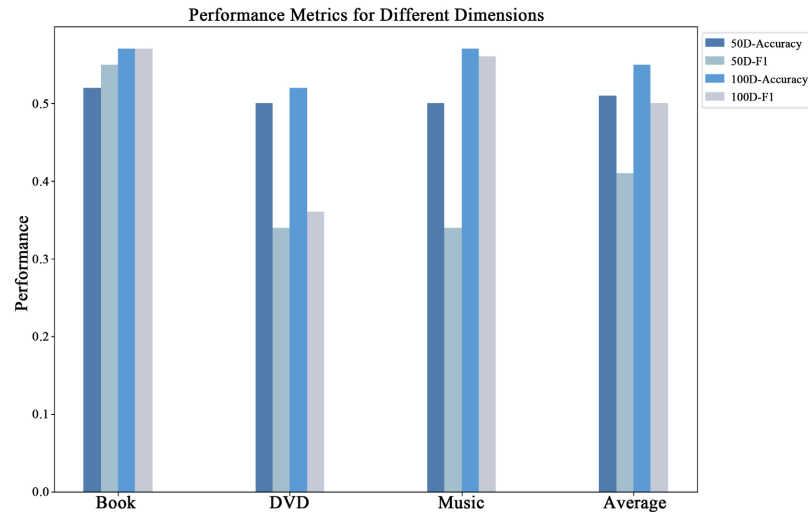


Figure 3. Classification results of examples II
图 3. 算例 II 分类结果

图 3 中分别为算例 II 在逻辑回归(LR)分类器, 在 BOOK、DVD 和 MUSIC 数据集上的情感分类结果。在实验中, 将词向量维度由 300 维降维至 50 维以及 100 维。由图中可以看出, 在所有数据集实验中, 词

向量维度由 300 维降维至 100 维可得到最优的分类效果, 且在 BOOK 数据集上取得较优的结果。

5. 结论

在本文中, 利用矩阵拉直算子、Moore-Penrose 广义逆得到矩阵方程 $AXB = C$ 极小范数最小二乘解的表达式, 进一步得到矩阵方程 $AXB = C$ 对称极小范数最小二乘解的表达式, 最后讨论矩阵方程 $WX = Y$ 在跨语言情感分类中的应用。

基金项目

广东省高校自然科学基金重点项目(2019KZDXM025), 五邑大学港澳联合研发基金资助项目(2019WGALH20)。

参考文献

- [1] 陈希孺. 数理统计学简史[M]. 哈尔滨: 哈尔滨工业大学出版社, 2021.
- [2] Mikolov, T., Le, Q.V. and Sutskever, I. (2013) Exploiting Similarities among Languages for Machine Translation. arXiv:1309.4168.
- [3] 彭晓娅, 周栋. 跨语言词向量研究综述[J]. 中文信息学报, 2020, 34(2): 1-15+26.
- [4] 王在华. 矩阵最小二乘问题的迭代求解及其在机器翻译中的应用[J]. 数学的实践与认识, 2021, 51(7): 119-126.
- [5] Xing, C., Wang, D., Liu, C. and Lin, Y. (2015) Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Denver, 31 May-5 June 2015, 1006-1011. <https://doi.org/10.3115/v1/n15-1104>
- [6] Mikolov, T., Sutskever, I., Chen, K., et al. (2013) Distributed Representations of Words and Phrases and Their Compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, 2, 3111-3119.
- [7] Smith, S.L., Turban, D.H.P., Hamblin, S., et al. (2017) Offline Bilingual Word Vectors, Orthogonal Transformations and the Inverted SoftMax. *Proceedings of the 5th International Conference on Learning Representations*, Toulon, 24-26 April 2017, 2521-2530.
- [8] Faruqui, M. and Dyer, C. (2014) Improving Vector Space Word Representations Using Multilingual Correlation. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg, 26-30 April 2014, 462-471. <https://doi.org/10.3115/v1/e14-1049>
- [9] Artetxe, M., Labaka, G. and Agirre, E. (2017) Learning Bilingual Word Embeddings with (almost) No Bilingual Data. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, 30 July-4 August, 2017, 451-462. <https://doi.org/10.18653/v1/p17-1042>
- [10] Hauer, B., Nicolai, G. and Kondrak, G. (2017) Bootstrapping Unsupervised Bilingual Lexicon Induction. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, Valencia, 3-7 April 2017, 619-624. <https://doi.org/10.18653/v1/e17-2098>
- [11] Conneau, A., Lample, G., Ranzato, M.A., Denoyer, L. and Jégou, H. (2018) Word Translation without Parallel Data. *6th International Conference on Learning Representations (ICLR)*, Vancouver, 30 April-3 May 2018, 1-14.
- [12] 袁仕芳, 廖安平, 雷渊. 矩阵方程 $AXB + CYD = E$ 的对称极小范数最小二乘解[J]. 计算数学, 2007, 29(2): 203-216.
- [13] 周立平, 邓小辉. 矩阵方程 $AX = B$ 的加权最小二乘对称、反对称解及其最佳逼近[J]. 湖南科技学院学报, 2011, 32(8): 6-10.
- [14] 李水勤, 邓继恩. 矩阵方程 $AXB = C$ 的对称最小二乘解[J]. 平顶山学院学报, 2010, 25(2): 64-65.
- [15] Peng, Z. (2005) An Iterative Method for the Least Squares Symmetric Solution of the Linear Matrix Equation $AXB=C$. *Applied Mathematics and Computation*, 170, 711-723. <https://doi.org/10.1016/j.amc.2004.12.032>
- [16] Peng, Z. (2008) Solutions of Symmetry-Constrained Least-squares Problems. *Numerical Linear Algebra with Applications*, 15, 373-389. <https://doi.org/10.1002/nla.578>
- [17] Deng, Y., Hu, X. and Zhang, L. (2003) Least Squares Solution of $BXA^T = T$ over Symmetric, Skew-Symmetric, and Positive Semidefinite X . *SIAM Journal on Matrix Analysis and Applications*, 25, 486-494.

<https://doi.org/10.1137/s0895479802402491>

- [18] Yuan, S. and Liao, A. (2014) Least Squares Hermitian Solution of the Complex Matrix Equation $AXB + CXD = E$ with the Least Norm. *Journal of the Franklin Institute*, **351**, 4978-4997. <https://doi.org/10.1016/j.jfranklin.2014.08.003>
- [19] 戴华. 矩阵论[M]. 第三版. 北京: 科学出版社, 2022.