

基于粒子群优化的最小二乘支持向量机税收预测模型研究

张淑娟¹ 邓秀勤¹ 刘波²

(广东工业大学应用数学学院 广州 510006)¹ (广东工业大学自动化学院 广州 510006)²

摘要 针对税收收入预测存在着非线性、不稳定性和多经济因素影响的复杂性,提出用最小二乘支持向量回归机的方法对广东省从化市的税收收入进行预测,并建立数学模型。由于模型中的参数 C 和 σ^2 直接影响支持向量机的预测效果,因此巧妙地融合了粒子群优化算法的思想,采用粒子群算法对参数进行寻优来确保预测模型的精确性和稳定性。仿真实验结果表明,相对于各参比模型,用粒子群算法对参数进行寻优的最小二乘支持向量回归机的预测精度有了显著提高,从而说明了该模型的有效性和实用性。

关键词 最小二乘支持向量机,粒子群优化,税收预测

中图分类号 TP301 **文献标识码** A

Research on Tax Forecasting Model Based on PSO and Least Squares Support Vector Machine

ZHANG Shu-juan¹ DENG Xiu-qin¹ LIU Bo²

(School of Applied Mathematics, Guangdong University of Technology, Guangzhou 510006, China)¹

(School of Automation, Guangdong University of Technology, Guangzhou 510006, China)²

Abstract Aiming at the tax revenue forecast for the existence of nonlinearity, instability and economic factors that affect multiple complexities, this paper offered to use the method of least squares support vector regression machine to predict the tax revenue of Guangdong conghua, and established the mathematical model. As the model parameters addi-rectly affect the quality of support vector machine, so the author ingeniously incorporated the idea of particle swarm optimization algorithm, and PSO for parameters optimization was used to ensure the accuracy and stability of the forecasting model. The simulation experimental results show that with respect to each reference model, using the PSO for parameters optimization of least squares support vector regression machine accuracy has improved significantly, illustrates the validity and practicability of the model.

Keywords Least squares support vector regression, Particle swarm optimization, Tax forecasting

1 引言

税收预测是以政府宏观调控政策为指导,以充分掌握影响税收收入变化的因素和税收历史资料为基础,运用数理统计和逻辑思维方法等手段,经过推理判断,对未来税收收入的前景做出趋势分析的一种工作方法。税收预测对加强组织收入工作,更好地完成税收任务,为领导的科学决策和管理提供服务等,都具有重大意义^[1]。税收数据是一种时间序列数据,受国家经济和政治的影响,具有非线性和不稳定的特点,对税收预测的准确度是选择算法的关键。目前,预测税收的方法有多元回归预测模型^[2]、时间序列模型^[3]、因素法、弹性系数预测模型、人工神经网络方法^[4]等。

支持向量机(Support Vector Machine, SVM)是由 Vapnik^[5]提出的,是统计学习理论的重要内容,它基于 VC 维理论和结构风险最小化原理,能在很大程度上减小泛化误差的上界,保证得到的是全局最优解而不是局部最优解,从而在机器学习和模式识别等领域得到了广泛的应用^[6-9]。支持向量回归(Support Vector Regression, SVR)通过加入 Vapnik's 的不敏感损失函数的支持向量机来解决非线性回归估计问

题,已经在很多预测领域得到了应用,如控制系统^[10]、生物信息^[11]、旅游需求^[12]、电力市场的价格^[13]等。Van Gestel 等^[14]用最小二乘支持向量回归(LS-SVR)结合贝叶斯证据框架做经济时间序列预测。在预测 DAX30 指数时,最小二乘支持向量机比时间序列分析方法(ARIMA)和非参数回归模型都取得了更好的效果。Pai 等^[15]首次提出了支持向量回归(SVR)的季节性分解法,实验证明季节性的 SVR 的预测效果优于传统的 SVR 模型。Li 等^[16]提出了一种一致的 LS-SVR 方法来校正最近的红外线光谱,通过对比 LS-SVR 和偏最小二乘回归(PLS)模型,发现 LS-SVR 模型更稳定,预测精度更高。这些研究结果表明,LS-SVR 是一种非常有效的预测方法。

粒子群算法是一种后发式的智能优化算法,具有比较强的全局协同搜索能力和全局优化的特性,并且算法的收敛速度快,更容易找到全局最优值。支持向量机的参数选择直接影响模型应用的效果。针对支持向量机和税收预测的特点,为了得到更优的参数和更好的预测效果,本文巧妙地融合了粒子群优化算法的思想,提出了通过控制预测误差的取值,采用粒子群优化算法对最小二乘支持向量回归模型中的参数 C 和 σ^2 进行近似寻优。粒子群优化算法中的粒子状态更新利

本文受国家自然科学基金项目(61472090, 61203280),广东省自然科学基金项目(S2013050014133)资助。

张淑娟(1991—),女,硕士生,主要研究方向为数据挖掘, E-mail: 1135451677@qq.com; 邓秀勤(1966—),女,硕士,教授,主要研究方向为数据挖掘、智能计算, E-mail: xiuqindeng@163.com(通信作者); 刘波(1978—),男,博士,教授,博士生导师,主要研究方向为数据挖掘、机器学习、模式识别。

用了粒子当前位置、经验位置和邻居经验位置 3 个方面的信息来调整自身状态。将粒子群优化算法的这种信息交换模式应用到参数的优化过程中,粒子更新受到自身经验和群体经验的双重影响,从而使得整个迭代过程更具方向性。然后,采用最小二乘支持向量回归对税收收入进行预测。最后,通过算例分析,对比传统的支持向量机(SVM)、网格参数寻优的支持向量机(GRID-SVM)、遗传算法参数寻优的支持向量机(GA-SVM)模型,验证了本文所构建的模型具有更优的预测效果。

2 最小二乘支持向量机

SVR 的基本思想是:定义 $x \in R^n, y \in R$, 其中 R^n 表示输入空间,通过非线性变换 $\varphi(\cdot)$ 将 x 输入空间映射到一个高维特征空间,并在这个特征空间中用线性函数拟合样本数据,同时得到很好的泛化能力。

在特征空间中,线性函数定义为:

$$y = f(x) = w^T \varphi(x_i) + b \quad (1)$$

其中, w 为权重, b 为偏置项, $\varphi(\cdot)$ 为非线性映射函数。

其目标函数为:

$$\text{Min } J_1(w, b, e) = \frac{1}{2} \|w\|^2 + \frac{1}{2} C \sum_{i=1}^N e_i^2 \quad (2)$$

$$\text{s. t. } y_i = w^T \varphi(x_i) + b + e_i, i = 1, 2, \dots, N \quad (3)$$

引入 Lagrange 乘子,得到的 Lagrange 函数为:

$$L_1(w, b, e, \alpha) = J_1(w, b, e) + \sum_{i=1}^N \alpha_i (y_i - w^T \varphi(x_i) - b - e_i) \quad (4)$$

其中, α 是拉格朗日乘子向量。

根据 KKT 条件^[17]得到:

$$\begin{cases} \frac{\partial L_1}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^N \alpha_i \varphi(x_i) \\ \frac{\partial L_1}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L_1}{\partial e_i} = 0 \Rightarrow e_i = \frac{1}{C} \alpha_i, i = 1, \dots, N \\ \frac{\partial L_1}{\partial \alpha_i} = 0 \Rightarrow y_i = w^T \varphi(x_i) + b + e_i, i = 1, \dots, N \end{cases} \quad (5)$$

在特征空间 $\varphi(x_i)$ 和 $\varphi(x_j)$ 中, $\varphi(x_i)^T$ 和 $\varphi(x_j)$ 的内积满足 Mercer 条件^[18], 可作为核函数, $k(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, 其中 $k(x_i, x_j)$ 是核函数。常见的核函数有多项式核函数、线性核函数、高斯径向基核函数等, 本文使用泛化能力较好的高斯径向基核函数, 其表达式如下:

$$k(x_i, x_j) = \exp(-\|x_i, x_j\|^2 / \sigma^2) \quad (6)$$

则 LS-SVR 的矩阵形式为:

$$\begin{bmatrix} A & E \\ E^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (7)$$

其中, $A = K + V, K = (k_{ij})_{N \times N}, V = \text{diag}(1/C, 1/C, \dots, 1/C)$; E 为元素全为 1 的 $N \times 1$ 矩阵。因此可得到 LS-SVR 的回归模型为:

$$f(x) = \sum_{i=1}^N \alpha_i k(x, x_i) + b \quad (8)$$

可以看出, 这个模型中有正则化参数 C 和径向基核函数中的 σ^2 两个参数, 比标准的 SVR 少了一个参数。

3 基于粒子群寻优的最小二乘支持向量机算法

粒子群算法最早是由美国电气工程师 Eberhart 和社会

心理学家 Kennedy 在 1995 年基于群鸟觅食提出的。在群鸟觅食模型中, 每个个体可以被看成一个粒子, 将 (C, σ^2) 看作粒子的位置值, 本文中把 (C, σ^2) 记为 $x = (x_1, x_2)$, 并将其看成群体中的粒子, 其中第 i 个粒子的位置表示为 $x_i = (x_{i1}, x_{i2})$, 粒子 i 的速度记为 $v_i = (v_{i1}, v_{i2})$, 粒子个体经历过的最好位置记为 $p_i = (p_{i1}, p_{i2})$, 整个群体所有粒子经历过的最好位置记为 $p_g = (p_{g1}, p_{g2})$ 。粒子群算法采用如下方程对粒子所在的位置不断进行更新:

$$v_{ij}^{(t+1)} = \omega v_{ij}^{(t)} + c_1 \gamma_1 (p_{ij}^{(t)} - x_{ij}^{(t)}) + c_2 \gamma_2 (p_{gij}^{(t)} - x_{ij}^{(t)}) \quad (9)$$

$$x_{ij}^{(t+1)} = x_{ij}^{(t)} + v_{ij}^{(t+1)}, j = 1, 2 \quad (10)$$

其中, ω 是非负数, 称为惯性因子; 加速常数 c_1 和 c_2 是非负常数; γ_1 和 γ_2 是在 $[0, 1]$ 范围内变换的随机数。

最小二乘支持向量回归机的模型的好坏与正则化参数 C 和径向基核函数中的 σ^2 两个参数有很大关系。 C 的作用是权衡学习机器置信范围和经验风险, 如果 C 太大, 目标仅使经验风险最小化; 反之, 当 C 的取值过小时, 则对经验误差的惩罚小, 从而使经验风险值变大。 σ 控制高斯函数的宽度以及训练数据的分布范围, σ 越大, 经验风险越大; σ 越小, 结构风险越大, 进而造成过拟合的现象。因而用支持向量机做税收预测的关键是, 如何在一定区域内搜索 LS-SVR 的参数, 找到最优的 C 和 σ^2 , 从而得到更好的预测模型。鉴于粒子群优化算法中的粒子状态更新, 利用粒子当前位置、经验位置和邻居经验位置 3 方面的信息来调整自身状态, 将粒子群优化算法的这种信息交换模式应用到了参数的优化过程中。粒子更新受到自身经验和群体经验的双重影响, 使整个迭代过程更具方向性, 从而提高了粒子群算法的全局搜索能力和收敛速度。本文用粒子群算法对最小二乘支持向量机的参数 C 和 σ^2 进行迭代寻优, 将数据分为训练集和测试集, 训练集用来训练支持向量回归机, 寻找最优参数 C 和 σ^2 , 从而可以得到最优的 LS-SVR 模型, 然后将测试数据集代入最优的 LS-SVR 模型即可进行税收预测。而成功应用粒子群优化算法求解最小二乘支持向量机的参数的关键是: 如何构造合适的适应度函数。鉴于粒子群算法中的每一个粒子代表支持向量回归机的一组参数, 用粒子群算法中的适应度函数来改变粒子的适应度值。本文选取均方误差(MSE)作为适应度函数, 即为:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i)^2 \quad (11)$$

其中, n 是训练样本个数, $\bar{y}_i$ 是预测值, y_i 为真实值。这样, 粒子的适应度值 MSE 为非负整数, 其值越小, 对应的可行解的适应度就越高。当适应度值 MSE 满足给定的精度要求时, 对应的解就是最优解, 即最小二乘支持向量机(LS-SVR)的最优参数。

基于粒子群寻优的最小二乘支持向量机算法(PSO-LS-SVM)的步骤如下:

- 1) 输入样本数据, 并对样本数据进行预处理;
- 2) 初始化粒子群的速度和位置、惯性因子、加速常数、最大迭代次数, 并将每个粒子的最优位置记为 p_i , 将整个群体所有粒子的全局最优位置记为 p_g ;
- 3) 根据粒子的适应度函数获得的粒子的适应度值调整粒子的个体最优位置 p_i 和全局最优位置 p_g ;
- 4) 根据式(9)和式(10)对粒子的位置和速度进行更新, 从而得到一组新的 LS-SVR 的参数, 进而建立优化的 LS-SVR 模型;

5)当达到最大迭代次数或粒子群目前为止搜索到的最优位置满足目标函数的最小容许误差时,终止迭代,输出结果,否则返回步骤 3)继续计算,直到达到上述终止条件为止。

4 模型应用

4.1 数据来源

本文选择 2005—2014 年广东省从化市的税收收入 Y (万

元)作为实验根据。根据影响因素的大小、资料的可比性及预测模型的要求等,选择以下 10 个指标(各指标单位都为万元):第一产业增加值 X_1 、第二产业增加值 X_2 、第三产业增加值 X_3 、农业总产值 X_4 、工业总产值 X_5 、固定资产投资 X_6 、社会消费品零售总额 X_7 、商品销售总额 X_8 、外贸进出口总值 X_9 、城乡居民储蓄存款余额 X_{10} 。根据统计年鉴得到各指标的数据,将数据分成两部分,2005—2011 年的数据作为训练数据,2012—2014 年的数据作为测试数据,如表 1 所列。

表 1 2005—2014 年从化市的税收及指标因子数据

年份	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
2005	17700	100700	476300	341000	189900	1778200	317900	316900	551000	137500	612900
2006	25048	130089	529247	420820	211355	2090117	399056	352295	587004	165931	690290
2007	44424	147750	635980	492511	238447	2578265	461488	402332	695712	214362	739493
2008	113981	158616	703931	591417	272523	2868266	593267	471445	605993	219471	893731
2009	134089	167563	710201	632013	273769	2967423	729329	545608	737345	178866	1059361
2010	157419	180209	854345	818394	306990	3536433	947763	663177	1466143	239082	1244444
2011	216432	197277	1047656	993263	338935	4214250	1112986	783243	1911534	310377	1448542
2012	210144	219350	1126353	1114416	367652	4961849	1278603	901756	2362620	294258	1624146
2013	215897	225525	1325516	1290418	383331	6149686	1535793	1050546	3184148	281178	1849054
2014	163382	240197	1474186	1456364	403092	7040365	1838640	1221785	4196095	270000	1948626

4.2 税收及指标数据的归一化处理

为了消除支持向量机预测指标数据量纲的不同对预测性能的影响,用式(12)对所有数据指标进行标准化处理,将所有数据归一化到 $[0, 1]$ 区间。

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (12)$$

其中, $\bar{x}_i$ 是归一化后的数据, x_i 是指标列数据, $x_{\max}$ 和 $x_{\min}$ 分别表示列数据中的最大值和最小值。

4.3 结果分析

在粒子群算法寻优中,设置 C 和 σ^2 在区间 $(0, 100)$ 内,迭代次数设为 100,种群数为 20。在 MATLAB2012b 上进行实验,结果显示最好的 C 为 1.027,最好的 σ 为 0.1。

为了评价预测方法的有效性,本文选择以下两种误差指标。

1) 平均绝对百分比误差

$$MAPE = \frac{1}{n} \sum_{i=1}^n |(y_i - \hat{y}_i) / y_i| \quad (13)$$

2) 均方百分比误差

$$MSPE = \frac{1}{n} \sqrt{\sum_{i=1}^n [(y_i - \hat{y}_i) / y_i]^2} \quad (14)$$

其中,用本文的预测方法得到的预测结果如图 1 所示。

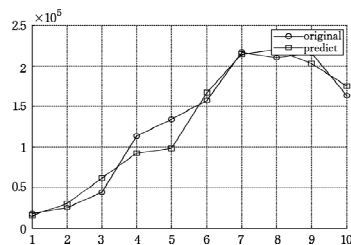


图 1 预测值与实际值的对比

为了验证本文所提模型的优劣,利用传统的支持向量机(SVM)、网格参数寻优的支持向量机(GRID-SVM)、遗传算法参数寻优的支持向量机(GA-SVM)和 BP 神经网络(BPNN)模型进行对比,预测分析结果见表 2。从表 2 可以看出,本文采用粒子群算法寻优的最小二乘支持向量机的预测模型,其预测的平均绝对百分比误差为 5.82%,均方百分比误差为 3.42%,预测结果比参比模型的 MAPE 和 MSPE 的误差值都小,表明采用本文的预测方法来预测税收收入的精度

高,预测值与实际值的拟合程度最高,效果最好。

表 2 各模型预测结果

年份	实际值	SVM	GRID-SVM	GA-SVM	BPNN	PSO-SVM
2012	210144	247670	242240	229480	280236	219440
2013	215897	277390	276040	239780	244342	203000
2014	163382	288850	306550	235840	272383	174940
MAPE		0.3775	0.4291	0.2154	0.3774	0.0582
MSPE		0.2794	0.3113	0.1554	0.2524	0.0342

结束语 本文采用粒子群算法寻优的最小二乘支持向量机对税收进行预测。首先,用粒子群优化算法找到最小二乘支持向量回归(LS-SVR)模型中的最优参数 C 和 σ^2 ,然后用最小二乘支持向量机进行税收预测。在建立税收预测模型时,通过粒子群优化算法寻找模型最优参数 C 和 σ^2 的过程,将预测结果与“训练数据”的实际结果进行比较,不断调整预测模型,直到模型的预测结果达到一个预期的准确率,提高了(LS-SVR)模型的预测能力。最后通过对比实验验证了本文所构建的模型具有更优的预测精度和预测效果。但是,对核函数的选择和对参数的设置依然是今后的研究方向。

参考文献

- [1] 中国国际税收研究会课题组. 当代税源管理理论[M]. 北京:人民出版社,2004.
- [2] 赵云河. 逐步回归法在税收预测中的应用研究[J]. Journal of Yunan Finance and Economics University, 2006, 22(5): 88-89.
- [3] 张梦瑶, 崔晋川. 基于时间序列法的国税月度收入预测模型研究[J]. 系统科学与数学, 2008(11): 77-84.
- [4] 张绍秋, 胡跃明. 基于 BP 神经网络的税收预测模型[J]. 华南理工大学学报(自然科学版), 2006, 34(6): 55-58.
- [5] VAPNIK V. The Nature of Statistical Learning Theory[M]. Springer-Verlag, New York, 1995.
- [6] LIU D H, QIAN H, DAI G, et al. An iterative SVM approach to feature selection and classification in high-dimensional datasets[J]. Pattern Recognition, 2013, 46(9): 2531-2537.
- [7] LI Y F, TSANG I W, KWOK J T, et al. Convex and scalable weakly labeled SVMs[J]. Journal of Machine Learning Research, 2013, 14(1): 2151-2188.
- [8] ALHAM N K, LI M Z, LIU Y, et al. A Map Reduce-based distributed SVM ensemble for scalable image classification and an-

- notation[J]. Computers & Mathematics with Applications, 2013, 66(10):1920-1934.
- [9] SALEM Y B, NASRI S. Automatic recognition of woven fabrics based on texture and using SVM[J]. Signal, Image and Video Processing, 2010, 4(4):429-434.
- [10] SHIN J, KIM H J, KIM Y. Adaptive support vector regression for UAV flight control[J]. Neural Network, 2011, 24(1):109-120.
- [11] AUEPANWIRIYAKUL S, SUMONPHAN E, THEERA-UM-PON N, et al. Automatic Nevirapine concentration interpretation system using support vector regression[J]. Computer Methods and Programs in Biomedicine, 2011, 101(3):271-281.
- [12] LEVIS A A, PAPAGEORGIOU L G. Customer demand forecasting via support vector regression analysis[J]. Chemical Engineering Research and Design, 2003(83):1009-1018.
- [13] GAO A C, BOMPARD E, NAPOLI R, et al. Price forecast in the competitive electricity market by support vector machine[J]. Physica A: Statistical Mechanics and its Applications, 2007, 382(1):98-113.
- [14] VAN GESTEL T, SUYKENS J A K, BAESTAENS D E, et al. Financial time series prediction using least squares support vector machines within the evidence framework[J]. IEEE Transactions on Neural Networks, 2001, 12(4):809-821.
- [15] PAI P F, LIN K P, LIN C S, et al. Time series forecasting by a seasonal support vector regression model[J]. Expert Systems with Applications, 2010, 37(6):4261-4265.
- [16] LI X, HU B, DU R. Predicting the parts weight in plastic injection molding using least squares support vector regression[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Review, 2008, 38(6):827-833.
- [17] KUHN H W, TUCKER A W. Nonlinear Programming[C]// Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1951:481-492.
- [18] MERCER J. Function of positive and negative type and their connection with the theory of integral equations[J]. Philosophical Transactions of the Royal Society, 1909, 209:415-446.

(上接第 96 页)

表 5 不同问题类别下基于单词和基于句子的实验

问题分类	基于单词		基于句子	
	正确数/个	准确率/%	正确数/个	准确率/%
答案匹配式问题	17	81	16	76.2
意义问题	20	44.4	23	51.1
位置问题	11	55	8	40
时间问题	5	56	3	33.3
判断问题	2	40	3	60

通过实验分析可以得出结论,基于单词的文章语义向量表示相对于基于句子的文章语义向量表示在位置问题、时间问题和答案匹配式问题上占据优势,能取得更好的结果;而基于句子的文章语义向量表示在意义问题、判断问题这类需要完整把握文章语义并进行总结的任务上的效果较好。

结束语 本文提出了双线性函数注意力 Bi-LSTM 模型,并将其应用于机器阅读理解中的备择答案预测任务,使用多层注意力转移的推理结构后,在四六级听力数据集上的准确率达到 54.65%,较好地完成了该任务;此外,对任务中的问题进行详细分析,对比了基于单词的文章语义向量表示和基于句子的文章语义向量表示,发现基于单词的文章语义向量表示能较好地把握文章中的关键字,在关键字选择推理型的任务中性能较好;而基于句子的文章语义向量表示能较好地整体把握文章的语义信息,较适用于需要通读整篇文章并作出总结的任务。

虽然本文的模型能较好地理解文章的语义,但是对于需要根据文章信息抽象总结出关于问题的答案的任务的表现仍有较大欠缺,今后的工作将考虑采用叠加记忆网络的方法^[18]来提升模型语义理解的能力。

参 考 文 献

- [1] BURGESS J C. Towards the machine comprehension of text: An essay. MSR-TR-2013-125[R]. 2013.
- [2] BORDES A, USUNIER N, CHOPRA S, et al. Large-scale simple question answering with memory networks[J]. arXiv preprint arXiv:1506.02075, 2015.
- [3] KUMAR A, IRSOY O, SU J, et al. Ask me anything: Dynamic memory networks for natural language processing[J]. arXiv preprint arXiv:1506.07285, 2015.
- [4] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12(8):2493-2537.
- [5] SUKHBATAR S, WESTON J, FERGUS R. End-to-end memory networks[C]// Advances in Neural Information Processing Systems. 2015:2440-2448.
- [6] KALCHBRENNER N, GREFFENSTETTE E, BLUNSON P. A convolutional neural network for modelling sentences[J]. arXiv preprint arXiv:1404.2188, 2014.
- [7] 尹宝才,王文通,王立春. 深度学习研究综述[J]. 北京工业大学学报, 2015, 41(1):48-59.
- [8] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization[J]. arXiv preprint arXiv:1509.00685, 2015.
- [9] CHEN D, BOLTON J, MANNING C D. A thorough examination of the cnn/daily mail reading comprehension task[J]. arXiv preprint arXiv:1606.02858, 2016.
- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [11] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [12] KADLEC R, SCHMOID M, BAJGAR O, et al. Text Understanding with the Attention Sum Reader Network[J]. arXiv preprint arXiv:1603.01547, 2016.
- [13] HERMANN K M, KOCISKY T, GREFFENSTETTE E, et al. Teaching machines to read and comprehend[C]// Advances in Neural Information Processing Systems. 2015:1693-1701.
- [14] 北京大学数学系前代数小组. 高等代数(第四版)[M]. 北京:高等教育出版社, 2013.
- [15] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv:1508.04025, 2015.
- [16] <http://nlp.stanford.edu/data/glove.6B.zip>.
- [17] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global Vectors for Word Representation[C]// EMNLP. 2014, 14:1532-1543.
- [18] WESTON J, CHOPRA S, BORDES A. Memory networks[J]. arXiv preprint arXiv:1410.3916, 2014.