基于多元线性回归的螺纹钢价格分析及预测模型

陈海鹏[1] 卢旭旺[2] 申铉京[1] 杨英卓[2]
(吉林大学计算机科学与技术学院 长春 130012)[1] (吉林大学软件学院 长春 130012)[2]

摘 要 通过分析期货黑色系品种螺纹钢产业链上下游的关系,提出了一种基于多元线性回归分析的螺纹钢价格分析及预测模型。首先,收集影响螺纹钢价格的主要因素数据,包括焦炭期货结算价、焦煤期货结算价、铁矿石期货结算价、热卷期货结算价与人民币兑美元汇率中间价;然后,通过散点图与趋势线对这些影响因素进行分析以确定影响因素,借助 SPSS 与 NCSS 软件利用收集到的数据构建基于最小二乘法的多元线性回归模型,并通过岭回归分析消除自变量间的共线性,得到修正后的模型;最后,运用此模型对未来一个月交易日的螺纹钢价格进行较为精准的预测。实验表明,该模型拟合度较高,具有一定的实用性。

关键词 多元线性回归,螺纹钢价格,最小二乘法,岭回归
中图法分类号 TP391 **文献标识码** A

Analysis and Prediction on Rebar Price Based on Multiple Linear Regression Model

CHEN Hai-peng[1] LU Xu-wang[2] SHEN Xuan-jing[1] YANG Ying-zhuo[2]
(College of Computer Science & Technology, Jilin University, Changchun 130012, China)[1]
(College of Software, Jilin University, Changchun 130012, China)[2]

Abstract A kind of rebar price analysis as well as prediction model based on multiple linear regression analysis was proposed by means of analyzing the upstream and downstream relationship of rebar industrial chain in futures black line variety. Firstly, the data of major factors influencing rebar price is collected, including coke futures settlement price, coking coal futures settlement price, iron ore futures settlement price, hot rolled futures settlement price, and central parity rate of RMB to USD. Later, these influencing factors are analyzed through scatter diagram and rend line to determine influencing factors. The multiple linear regression model based on least square method is constructed by virtue of SPSS and NCSS, and the collected data. Meanwhile, the collinearity among independent variables are moved through ridge regression to obtain revised model. At last, this model is applied to carry out accurate prediction of rebar price on trade day in the next month. The experiment indicates that the fitting degree of this model is higher with certain practicability.

Keywords Multiple linear regression, Rebar price, Least-square method, Ridge regression

1 引言

螺纹钢是我国产量最大的钢材品种之一,被广泛用于房屋、桥梁、道路等土建工程建设。大到高速公路、铁路、桥梁、涵洞、隧道、防洪、水坝等公用设施,小到房屋建筑的基础、梁、柱、墙、板,螺纹钢都是不可或缺的结构材料。自钢材期货、铁矿石期货、焦炭期货、焦煤期货上市后,我国钢铁产品更具金融属性。

上海期货交易所(SFE)的螺纹钢期货成交量已经连续 4 年成为全球最大的钢铁类期货品种[1]。国内外经济实践都表明,大宗商品是实体经济的晴雨表,"黑色系"是我国大宗商品中使用最多的。因此,螺纹钢价格的变动将直接影响到建筑行业与我国钢材的进出口状况。此外,螺纹钢还与基础建设投资有着密切的关系。通过分析螺纹钢价格变化的影响因素,可以加深对螺纹钢与宏观经济之间的了解。

目前,国内关于螺纹钢期货价格的影响因素的实证研究还很少,大部分是关于钢材期货价格与现货价格或与宏观经济之间的分析。文献[2]利用计量经济学方法对上海期货交易所螺纹钢期货价格和天津现货螺纹钢价格两者间的关系进行了实证研究,研究结果显示,两者之间存在长期的均衡关系,并且螺纹钢期货价格和现货价格存在双向引导关系,但现货市场在钢材价格形成中处于主导地位。文献[3]对上海螺纹钢期货价格与现货价格序列进行了单位根检验、协整检验、Granger 因果检验和方差分解,且实证结果表明螺纹钢期货与现货价格保持着长期的均衡关系,而且期货价格与现货价格能有效地相互引导。文献[4]运用经验模态分解方法(EMD),分别将螺纹钢期货价格和现货价格时间序列分解成若干 IMF 分量和趋势项。通过对分解后的不同分量进行统计和计量分析得出,螺纹钢期货市场已经基本具备了价格发现的功能。文献[5]用 VAR 模型对我国螺纹钢期货和现货价格建立动态关系,利用 ADF 单根检验、协整检验、因果检验等方法来研究两者的关系和相互引导作用,结果表明两者间保持长期均衡,且期货价格对现货价格有显著的单向引导。文献[6]选取制造业 PMI、国房景气指数、SHIBOR 及人民币汇率 4 个宏观

本文受国家青年科学基金项目(61305046,61602203),吉林省自然科学基金项目(20140101193JC)资助。
陈海鹏(1978—),男,博士,副教授,CCF 会员,主要研究方向为图像处理与模式识别、信息安全;卢旭旺(1991—),男,硕士生,主要研究方向为大数据、云计算,E-mail:luxuwangallen@foxmail.com(通信作者);申铉京(1958—),男,博士,教授,主要研究方向为多媒体技术、数字图像处理、智能控制、嵌入式系统;杨英卓(1991—),男,硕士生,主要研究方向为云计算。

经济指标和国内螺纹钢期货价格，构建误差修正模型(VEC)，来考查各宏观经济因素对钢材期货价格的短期影响和长期影响，并针对钢铁企业及期货投资者提出相关建议。

本文从期货黑色系品种螺纹钢产业链中上、下游本身存在的密切关系出发，同时引入影响大宗商品行情的人民币兑美元汇率中间价这一宏观经济指标，并结合多元线性回归分析模型在解决受多因素综合影响的实际经济问题方面的优势，建立了基于多元线性回归的螺纹钢价格分析及预测模型，用于对影响螺纹钢价格的因素进行分析并实现了螺纹钢价格预测功能。相比于传统的关于螺纹钢期货价格的影响因素分析方法，该方法具有考虑全面、操作简单方便、预测结果准确、模型解释能力强的特点，并且可以准确地描述各个影响因素之间的相关程度和回归拟合程度，具有一定的实用效果。

2 多元线性回归

回归分析分为单变量回归分析与多元回归分析[7]。在实际中，对因变量的影响往往有两个或两个以上的自变量。例如，影响产品单位成本的变量不仅有产量，还包括原材料价格、劳动力价格、劳动效率及废品率等因素。在多元回归分析中，如果因变量和多个自变量的关系为线性，就属于多元线性回归。多元线性回归广泛应用在解决两个或两个以上的解释变量来解释因变量的实际场景中。例如，文献[8]通过分析下水道系统建筑合同的成本和基础设施数据，利用水管材料、水管直径、挖掘深度、入孔深度、流速等因素建立了用于评估建造污水系统成本的函数，使得建造成本显著降低。文献[9]通过建立基于主成分分析的评价模型，从近 30 个葡萄酒指标中选择 8 个主成分，然后建立多元线性回归模型，通过分析葡萄酒与葡萄酒物理化学指标的关系，得出了葡萄酒的物理化学指数不适合评价葡萄酒的质量这一结论。文献[10]利用多元线性回归模型的理论知识，针对近年来愈演愈烈的能源问题，引入 1980 年到 2011 年的统计数据来建立能源需求模型，并运用多元线性回归分析的方法对影响国家最终能源需求的各种因素进行分析，为国家能源的利用提出一些可行性建议。

多元线性回归分析的一般模型为：

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p+\varepsilon \quad (1)$$

$$\varepsilon\sim N(0,\sigma^2)$$

其中，$x_1,x_2,\cdots,x_p$ 是 $p(p\geqslant 2)$ 个自变量(解释变量)，y 是因变量，$\beta_0,\beta_1,\beta_2,\cdots,\beta_p$ 是 $p+1$ 个未知参数，β_0 称为回归常数，$\beta_0,\beta_1,\beta_2,\cdots,\beta_p$ 称为回归系数，$\varepsilon\sim N(0,\sigma^2)$ 为随机误差。

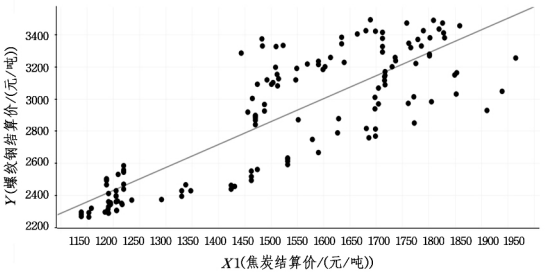
3 模型的建立

3.1 样本的选取

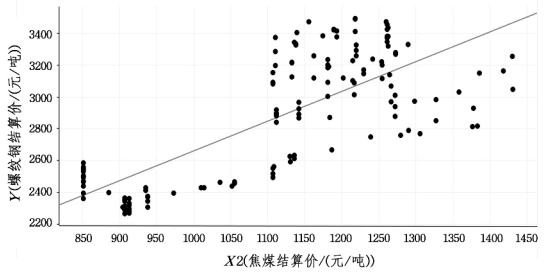
由于期货的交易时长是一年，因此选择交易代码为 1708 的黑色系品种的商品期货从 2016 年 8 月 16 日交易首日至 2017 年 3 月 16 日这 7 个月交易日的交易数据以及对应交易日的人民币对美元汇率中间价数据作为测试集。其中包括了大连期货交易所的焦炭(J)结算价、焦煤(JM)结算价、铁矿石(I)结算价、上海期货交易所的螺纹钢(RB)结算价、热卷(HC)结算价以及国家外汇管理局人民币汇率中间价数据。本文选择以上几项因素作为影响因素是由于有产业链的关系，假设我们想要预测螺纹钢的价格，那么影响螺纹钢价格的因素可以会涉及到原材料、能源资源和同类材料等。比如，如果铁矿石价格上涨，螺纹钢就应该要涨价了。同时，人民币兑美元汇率中间价的选取，旨在分析宏观经济指标对大宗商品期货影响的显著性。

3.2 自变量的选取

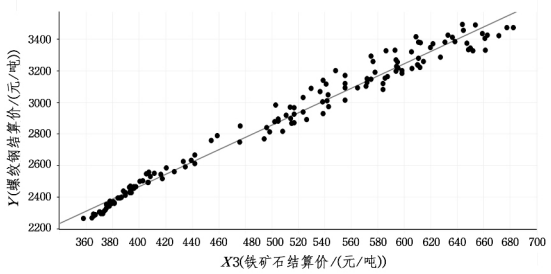
自变量的选取可以通过画散点图与趋势线相结合的方式来实现[11]。在这里，假设 $X1$ 为焦炭(j1708)合约的日结算价数据，$X2$ 为焦煤(jm1708)合约的日结算价数据，$X3$ 为铁矿石(i1708)合约的日结算价数据，$X4$ 为热卷(hc1708)合约的日结算价数据，$X5$ 为对应螺纹钢交易日人民币对美元汇率中间价数据，$X1-X5$ 为自变量；Y 为螺纹钢(rb1708)合约的日结算价数据，为因变量。图 1 分别给出了螺纹钢结算价与各影响因素的散点图及趋势线。



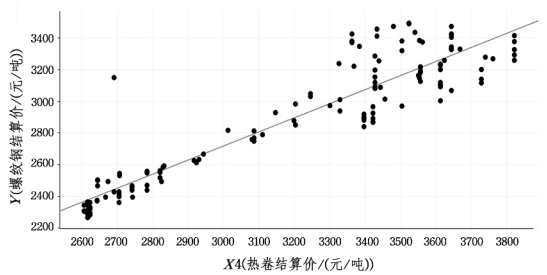
(a)螺纹钢结算价与焦炭结算价



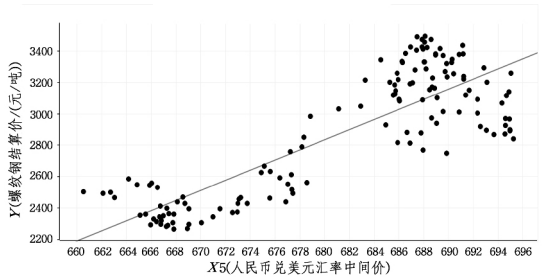
(b)螺纹钢结算价与焦煤结算价



(c)螺纹钢结算价与铁矿石结算价



(d)螺纹钢结算价与热卷结算价



(e)螺纹钢结算价与人民币兑美元汇率中间价

图 1 螺纹钢结算价和与各影响因素散点图

通过分析图 1 可知，焦炭结算价、焦煤结算价、铁矿石结算价、热卷结算价和人民币对美元汇率中间价与螺纹钢的结算价之间均呈现出正向线性相关关系。其中，铁矿石结算价和热卷结算价与螺纹钢结算价之间的正向线性相关关系尤其显著，焦炭结算价和焦煤结算价与螺纹钢结算价之间的线性相关关系次之，而人民币对美元汇率中间价与螺纹钢结算价之间的线性相关性一般。在此，先保留以上所有影响因素，通过下一步观察各趋势线的信息来综合决定影响因素的引入。

表 1 列出了各趋势线的详细信息。趋势线是数据趋势的图形表示形式，利用最小二乘法原理得出，这些趋势线的可靠性用 R^2 来描述，当趋势线的 R^2 值为 1 或者接近 1 时，趋势线最可靠。

表 1　各趋势线信息

列	行	R^2	标准误差	P 值(显著性)
Y	$X1$	0.795185	217.831	<0.0001
Y	$X2$	0.743682	260.815	<0.0001
Y	$X3$	0.977925	58.6204	<0.0001
Y	$X4$	0.860225	147.507	<0.0001
Y	$X5$	0.696727	217.278	<0.0001

通过分析表 1 各个影响因素的趋势线的详细信息可知，各趋势线 R^2(拟合度)都很好，P 值(显著性)均十分显著，因此将这 5 个因素都引入模型作为输入变量。

3.3　利用最小二乘法建立多重回归模型

如表 1 所列，各趋势线信息显著性均满足 $P\leqslant 0.05$ 时有意义。利用 spss24 软件对以上 5 个变量建立多元线性回归模型，得到的结果如表 2、表 3 所列。

表 2　拟合优度检验

模型摘要[b]				
模型 1	R	R^2	调整后 R^2	标准估计的误差
	0.994	0.989	0.988	42.656488

a. 预测变量：(常量)，$X5$，$X4$，$X3$，$X2$，$X1$

b. 因变量：Y

表 2 拟合优度检验列出了模型的 R 值、R^2 值、调整的 R^2 值以及标准估计的误差。R^2 值越大，所反映的两变量的共变量比率越高，模型与数据的拟合程度越好。

表 3　回归系数的最小二乘法估计结果

(a)未标准化系数与标准化系数

系数[a]			
	未标准化系数		标准化系数
	B	标准误差	β
常量	1355.956	598.829	
$X1$	0.443	0.065	0.259
$X2$	−0.351	0.092	−0.140
$X3$	2.887	0.101	0.735
$X4$	0.183	0.027	0.191
$X5$	−1.176	0.989	−0.030

a. 因变量：Y

(b)t 值、显著性、容差、VIF

系数[a]				
	t	显著性	容差	VIF
常量	2.264	0.025		
$X1$	6.830	0.000	0.059	16.919
$X2$	−3.793	0.000	0.062	16.191
$X3$	28.721	0.000	0.129	7.731
$X4$	6.825	0.000	0.108	9.242
$X5$	−1.189	0.237	0.129	7.752

a. 因变量：Y

观察表 2 可以看出，利用 5 个变量建立的多元线性回归模型的拟合度 $R^2=98.4\%$，说明利用最小二乘法得出的多元线性回归模型的拟合效果较好。

从表 3 得到模型各自变量的未标准化以及标准化系数、t 检测值、容差以及 VIF(方差膨胀因子)信息，结合以上信息，得到最小二乘法下的回归模型：

$$Y=1355.956+0.443X1-0.351X2+2.887X3+0.183X4-1.176X5 \quad (2)$$

4　多重共线性的诊断

4.1　共线性分析

多重共线性由 R. Frisch 于 1934 年提出[12]，是指线性回归模型中的解释变量之间由于存在精确相关关系或高度相关关系而使模型估计失真或难以估计准确。因此，多重共线性是在进行线性回归时存在的问题。存在共线性时，普通最小二乘估计仍是线性无偏估计，不再是有效估计，同时各种变量的显著性检验也会失效，参数估计值的方差增大，模型的预测也会没有意义。观察各个变量的 t 检验结果发现变量的 t 检验结果并不显著，方差膨胀因子 VIF 值均大于 2，说明解释变量之间的共线性很强，存在相互解释的关系，因此需进一步诊断模型中是否存在多重共线性。如表 4 所列，各个条件指数都很大，表明变量间存在多重共线性。接下来，需要解决自变量之间的共线性问题。

表 4　共线性诊断

共线性诊断[a]							
特征值	条件指标	常量	$X1$	$X2$	$X3$	$X4$	$X5$
5.960	1.000	0.00	0.00	0.00	0.00	0.00	0.00
0.027	14.787	0.00	0.00	0.00	0.06	0.00	0.00
0.010	24.770	0.00	0.03	0.04	0.12	0.02	0.00
0.002	55.080	0.00	0.02	0.02	0.54	0.58	0.00
0.001	89.362	0.00	0.86	0.68	0.28	0.09	0.00
1.588E−5	612.662	1.00	0.08	0.25	0.01	0.31	1.00

a. 因变量：Y

4.2　利用岭估计法建立线性模型

Hoerl 于 1962 年首先提出了岭回归分析(Ridge Regression Analysis)，1970 年后又与 Kennard 合作系统地对其做了发展[13]。它是一种改良的最小二乘估计方法，用于解决线性回归分析中自变量存在共线性的问题。当自变量之间存在多重共线性即 $|X^{\mathrm{T}}X|\approx 0$ 时，设想给 $X^{\mathrm{T}}X$ 加一个正常数矩阵 kI($k>0$，I 为单位矩阵)，则得到岭回归的估计量为：

$$\hat{\beta}(k)=(X^{\mathrm{T}}X+kI)^{-1}X^{\mathrm{T}}Y \quad (3)$$

显然，当 $k=0$ 时，岭回归的估计量即为最小二乘估计；当 $k\to\infty$ 时，岭回归的估计量趋于 0，因此 k 不宜过大[14]。

用 NCSS11 进行岭回归分析时，岭回归不同值计算的回归系数结果见表 5。岭估计法在不同 k 值时计算的回归系数的岭迹图见图 2。表 5 所列为进行岭回归分析时不同 k 值所对应的各系数值的变化。

鉴于 $k>0$，设置 k 的范围为[0，1]，每次增加步长为 0.001。由图 2 可以看出，当 k 从 0 变化到 1 时，各个自变量的回归系数有很大的变化幅度。当 k 值增大至 0.1 到 0.3 时，得到的各回归系数的岭估计值基本趋于稳定，用最小二乘法估计得出的不合理符号也变得合理，且没有不符合经济意义的回归系数。随着 k 值的增大，VIF 值越来越小，自变量间的共线性得到消除。选取 k 值为 0.2 时的回归系数、标准误、VIF 值列于表 6 中。

表 5 选取不同 ridge regression 回归系数的变化

k 值	$X1$	$X2$	$X3$	$X4$	$X5$
0.000000	16.9191	16.1913	7.7309	9.2423	7.7519
0.001000	15.8909	15.1932	7.4903	8.9729	7.5102
0.002000	14.9590	14.2890	7.2648	8.7159	7.2845
0.003000	14.1115	13.4673	7.0529	8.4706	7.0732
0.004000	13.3384	12.7183	6.8533	8.2361	6.8749
0.005000	12.6313	12.0335	6.6649	8.0118	6.6883
0.006000	11.9828	11.4059	6.4867	7.7972	6.5123
0.007000	11.3864	10.8292	6.3179	7.5916	6.3460
0.008000	10.8368	10.2980	6.1577	7.3946	6.1885
0.009000	10.3290	9.8076	6.0054	7.2057	6.0391
0.010000	9.8590	9.3540	5.8604	7.0244	5.8972
0.020000	6.5976	6.2177	4.7172	5.5528	4.7834
0.030000	4.8127	4.5146	3.9382	4.5230	4.0254
0.040000	3.7219	3.4822	3.3710	3.7725	3.4700
0.050000	3.0014	2.8055	2.9390	3.2073	3.0428
0.060000	2.4970	2.3355	2.5988	2.7702	2.7028
0.070000	2.1278	1.9938	2.3240	2.4246	2.4253
0.080000	1.8477	1.7362	2.0974	2.1460	2.1942
0.090000	1.6290	1.5361	1.9075	1.9178	1.9988
0.100000	1.4540	1.3767	1.7461	1.7282	1.8313
0.200000	0.6803	0.6765	0.8998	0.8160	0.9334
0.300000	0.4323	0.4473	0.5739	0.5076	0.5827
0.400000	0.3113	0.3309	0.4072	0.3588	0.4053
0.500000	0.2402	0.2599	0.3086	0.2729	0.3021
0.600000	0.1938	0.2122	0.2446	0.2178	0.2364
0.700000	0.1614	0.1780	0.2004	0.1797	0.1917
0.800000	0.1376	0.1525	0.1684	0.1521	0.1599
0.900000	0.1195	0.1328	0.1443	0.1312	0.1362
1.000000	0.1053	0.1172	0.1257	0.1150	0.1181

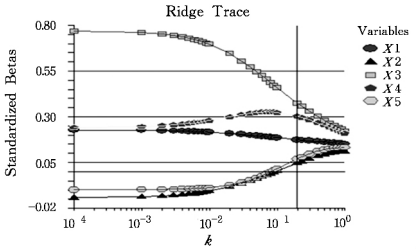


图 2 岭回归在不同 k 值时计算的回归系数的岭迹图

表 6 $k=0.200$ 时的标准化回归系数、标准误、VIF 值

	标准化回归系数	标准误	VIF
$X1$	0.1691	0.0380	0.6803
$X2$	0.0106	0.0553	0.6765
$X3$	0.4621	0.1004	0.8998
$X4$	0.2606	0.0233	0.8160
$X5$	0.0660	1.0049	0.9334

调整后拟合度 $R^2=93.3\%$，整个模型的拟合优度符合要求，效果良好。得到岭回归方程：

$$Y=-1040.198+0.290X1+0.026X2+1.815X3+0.250X4+2.548X5 \quad (4)$$

利用上面得到的回归方程对未来一个月的交易日螺纹钢价格进行预测，结果如图 3 所示。结果显示回归曲线的拟合程度较好，从而说明预测值也是可靠的。

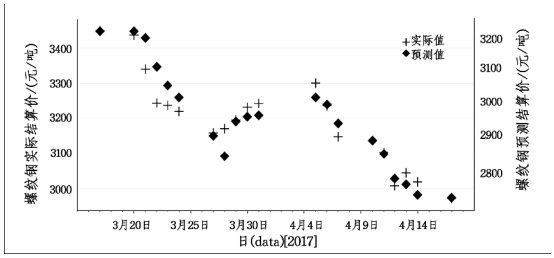


图 3 3 月 17 日－4 月 17 日交易日螺纹钢价格的预测值和真实值对比

表 7 显示 2017 年 3 月 17 日－4 月 17 日交易日螺纹钢价格的预测值和真实值统计情况。从结果可以看出，预测值与真实值之间的误差很小，预测准确率很高，预测结果很理想，说明该模型准确，有很高的实用价值。

表 7 2017 年 3 月 17 日－4 月 17 日交易日螺纹钢价格的预测值和真实值

日期	预测值	实际值	准确率
3.17	3214	3449	0.927
3.20	3214	3437	0.930
3.21	3198	3340	0.961
3.22	3123	3244	0.961
3.23	3076	3237	0.948
3.24	3046	3219	0.943
3.27	2949	3158	0.930
3.28	2898	3171	0.906
3.29	2985	3195	0.930
3.30	2997	3231	0.922
3.31	3001	3242	0.920
4.05	3047	3301	0.917
4.06	3028	3236	0.931
4.07	2980	3148	0.944
4.10	2937	3137	0.932
4.11	2904	3103	0.931
4.12	2841	3008	0.941
4.13	2826	3045	0.923
4.14	2801	3045	0.922
4.17	2793	2973	0.935

结束语 本文通过对影响螺纹钢价格的因素进行分析，建立了基于焦炭期货结算价、焦煤期货结算价、铁矿石期货结算价、热卷期货结算价以及人民币兑美元汇率中间价 5 个因素的多元线性回归模型来分析影响螺纹钢价格的因素并预测螺纹钢价格。由于所选取的各要素之间存在多重共线性，因此通过岭回归分析方法，参考岭回归的岭迹图，选取合适的 k 值对模型给予修正，得到了合适的岭回归模型。实验结果表明，模型拟合度良好，具有一定的理论意义与实用价值。

参 考 文 献

[1] 刘金山，王晓晓. 中外螺纹钢期现价格互动效用研究[J]. 商业研究，2015(1)：8-14.

[2] 马刚，马丽. 我国钢材期货价格、现货价格关系实证研究——基于螺纹钢期货市场与现货市场数据[J]. 中国证券期货，2010(8)：27-29.

[3] 刘任帆，陈芳舒. 我国钢材期货与现货市场的互动性探究——基于上海螺纹钢的实证研究[J]. 杭州电子科技大学学报(社会科学版)，2011(3)：13-17.

[4] 王立民，兴长宇，刘祥东，等. 基于 EMD 分解的螺纹钢期货价格发现的实证研究[J]. 科技和产业，2012，12(8)：78-82.

[5] 李静晶. 我国螺纹钢期货价格与现货价格研究[J]. 金融经济(理论版)，2016(3)：149-151.

[6] 成月. 宏观经济因素对我国螺纹钢期货价格的影响研究——基于 VEC 模型的实证分析[J]. 市场论坛，2016(6)：20-23.

[7] UYANLK G K，GÜLER N，et al. A Study on Multiple Linear Regression Analysis[J]. Procedia-Social and Behavioral Sciences，2013，106(106)：234-240.

[8] MARCHIONNI V，LOPES N，MAMOUROS L，et al. Modelling Sewer Systems Costs with Multiple Linear Regression[J]. Water Resources Management，2014，28(13)：4415-4431.

(下转第 97 页)

下提取 3 组共 900 个文本进行实验，如表 11 所列。

表 11　实验数据集

组别	文本总数	褒义文本数	贬义文本数
1	300	150	150
2	300	150	150
3	300	150	150

本文方法主要与传统语义理解、原始的语义相似度计算公式、文献[20]以及文献[21]中的基于《知网》改进的倾向性算法进行比较，比较的参数主要是查全率、查准率、F-测量值，实验数据如表 12 所列。实验表明，本文方法比传统的语义理解方法在 3 个指标上均有提高，准确性也更高。

表 12　语义理解对比实验/%

方法	评估标准	第一组	第二组	第三组	平均值
传统语义理解	查准率	69.48	67.68	74.65	70.60
	查全率	57.00	56.50	59.00	57.50
	F-测量值	62.62	61.59	65.91	63.37
原始语义理解	查准率	75.35	69.60	75.32	73.42
	查全率	72.50	67.50	73.50	71.17
	F-测量值	73.90	68.53	74.40	72.28
文献[20]改进算法	查准率	76.58	70.34	75.87	74.26
	查全率	74.21	68.70	72.45	71.79
	F-测量值	75.38	69.51	75.65	73.51
文献[21]改进算法	查准率	78.95	71.67	77.07	75.90
	查全率	76.00	69.50	75.50	73.67
	F-测量值	77.45	70.57	76.29	74.77
本文方法	查准率	79.36	72.63	79.36	76.32
	查全率	78.10	69.89	76.98	74.00
	F-测量值	78.96	70.23	77.63	74.87

结束语　本文改进了以词语相似度为基础的语义相似度计算方法，提出了一种改进的基于语义理解的文本情感分类方法来判定文本的情感倾向性。实验分析证明了该方法的有效性。否定词处理算法和文本倾向度算法没有对复杂并列句进行进一步的研究，这将是今后的重点。

参 考 文 献

[1] 中国互联网信息中心. 第 38 次中国互联网络发展状况统计报告[EB/OL]. [2016-12-30]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201608/t20160803_54392.htm.

[2] RADA R, MILI H, BICKNELL E, et al. Development and Application of a Metric on Semantic Nets[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1989, 19(1): 17-30.

[3] RESNIK P. Using information content to evaluate semantic similarity in a taxonomy[C]// Proceeding of the 14th International Joint Conference on Artificial Intelligence. 1995: 448-453.

[4] RIGAU G, AGIRRE E. A proposal for word sense disambiguation using conceptual distance[C]// International Conference/Recent Advances in Natural Language Processing. 1995: 35-43.

[5] LEE L J. Similarity-Based Approaches to Natural Language Processing[J/OL]. arXiv: com-lg/9708011.

[6] TURNEY P D. Thumbs up or thumbs down ?Sem-antic orientation applied to unsupervised classification of reviews[C]// Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL). Philadel-phia, 2002: 417-424.

[7] PANG B, LEE L, VAITHYANATHAN S. Thumbs up ? Sentiment classification using machine learning techniques[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing(EMNLP). 2002: 79-86.

[8] 张积家. 测量语义相似性的方法[J]. 心理科学, 1992(3): 53-55, 62.

[9] 关毅，王晓龙. 基于统计的汉语词汇间语义相似度计算[C]// 语言计算与基于内容的文本处理—全国第七届计算语言学联合学术会议论文集. 2003: 221-227.

[10] 章成志. 一种基于语义体系的同义词识别研究[J]. 淮阴工学院学报, 2004, 13(1): 59-62, 67.

[11] 田久乐，赵蔚. 基于同义词词林的词语相似度计算方法[J]. 吉林大学学报(信息科学版), 2010, 28(6): 602-608.

[12] 闻彬，何婷婷，罗乐，等. 基于语义理解的文本情感分类方法研究[J]. 计算机科学, 2010, 37(6): 261-264.

[13] 马力，刘笑，宫玉龙. 基于语义的微博短文本倾向性分析研究[J]. 计算机应用研究, 2016, 33(10): 2914-2918.

[14] 徐健锋，许园，许元辰，等. 基于语义理解和机器学习的混合的中文文本情感分类算法框架[J]. 计算机科学, 2015, 42(6): 61-66.

[15] 游春晖. 基于语义情感倾向的文本相似度计算[D]. 成都：电子科技大学, 2008.

[16] 吴全娥. 汉语句子相似度计算及其在自动问答系统中的应用[D]. 重庆：西南大学, 2011.

[17] 闻彬，何婷婷，罗乐，等. 基于语义理解的文本情感分类方法研究[J]. 计算机科学, 2010, 37(6): 261-264.

[18] HATZIVASSILOGLOU V, MCKEOWN K. Predicting the semantic orientation of adjectives[C]// Proceedings of the 35th Annual Meeting of Association for Computational Linguistics (ACL-97). Madrid, Spain, 1997: 174-181.

[19] 蔺璜，郭姝慧. 程度副词的特点范围与分类[J]. 山西大学学报(哲学社会科学版), 2003, 26(2): 71-74.

[20] 党蕾，张蕾. 一种基于知网的中文句子情感倾向判别方法[J]. 计算机应用研究, 2010, 27(4): 1370-1372.

[21] 许元辰. 基于优化的语义理解与 SVM 相结合的文本情感分类研究[D]. 南昌：南昌大学, 2014.

(上接第 64 页)

[9] LIU H H, JI W L, ZHANG P, et al. The Research of Wine Quality Evaluation Based on Multiple Linear Regression[J]. Advanced Materials Research, 2013, 756-759: 2489-2493.

[10] 崔江龙. 多元回归分析在能源利用中的应用[J]. 商, 2015(49): 68.

[11] 付倩娆. 基于多元线性回归的雾霾预测方法研究[J]. 计算机科学, 2016, 43(s1): 526-528.

[12] 陈玲燕. 多重共线性下的线性回归方法综述[J]. 现代农业, 2008(5): 67-69.

[13] HOERL A E, KENNARD R W. Ridge regression: biased estimation for nonorthogonal problems[J]. Technometrics, 1970, 42(1): 80-86.

[14] 张丹平. 基于岭回归方法的我国能源消费影响因素研究[J]. 统计与决策, 2012(21): 146-148.