

浙师大学生的英文能力能帮助《人工智能》课程学习吗？

蓝 瑋

(浙江师范大学计算机科学系, 浙江 金华 321004)

摘要: 计算机科教界对英文的重视程度与日俱增, 一个具体表现是使用权威英文原版教材进行教学。一个有趣的问题是, 如果使用英文原版教科书, 中国学生的英文能力能够帮助吸收计算机科学知识吗? 为回答此问题, 我对 2018 年春季我在浙江师范大学所讲授的《人工智能》课程的大三学生做了一学期的跟踪研究, 收集了他们的《人工智能》期末考试成绩、以及英文成绩、《线性代数》成绩、《概率论与数理统计》成绩。本文的主要结论是, 在浙江师范大学, 全国大学英语等级考试成绩与采用英文原版教材的《人工智能》课程期末成绩无显著关系, 英语成绩无法预测《人工智能》成绩。然而, 《概率论与数理统计》成绩则与《人工智能》课程期末成绩正相关。

关键词: 英文原版教材; 计算机科学教育; 全国大学英语等级考试; 人工智能; 概率论与数理统计; 语言与认知

引言

浙江高校对英文的重视程度与日俱增^[1], 如浙江大学计算机学院早在 2000 年就开设实验班对部分计算机专业课程如《数据结构》、《操作系统》、《计算机体系结构》采用英文原版教材。由于(计算机)科学在英美最为发达, 所以学界有这样一种模糊认识, 似乎英文好就能够科学好, 所以为了科学好我们要使用英文, 你看, 英美很发达, (因为)他们使用英文。然而这种认识鲜有可靠数据支持。我们不禁要思考, 科学原创的驱动力是什么? 使用英文对科学原创真有助益吗? 是否还有其它决定因素? 回答这样笼统问题的确困难, 因为需要大量材料, 且人有思维惯性, 只相信他们愿意相信的东西。

可是我们可以试着去回答比较具体问题, 如, 在采用英文原版教材的情况下, 某个大学某个班级非英语为母语的学生的英语能力能显著影响该门科学课程的成绩吗? 如果英语(阅读理解)能力与科学课程成绩有显著的正相关性, 即, 总体来看, 英文成绩越好, 科学成绩越高, 英文成绩越差, 科学成绩越差, 那么虽然不能完全说明使用英文对科学原创有益, 但是至少也能说明使用英文对于科学原创的基础 — 科学教育, 是有效的。

材料与方法

2018 年春季学期我在浙江师范大学(浙师大)讲授《人工智能》, 对象是当时数理信息学院的大三学生, 我采用了英文原版教材^[2], 教材可在作者网站下载。我希望检验如下的假设是否成立: 在采取权威英文原版教材的情况下, 中国学生的英语能力可以帮助《人工智能》课程的学习。简单地说, 就是英文成绩好对学习这门课程有帮助。

为此, 我记录了教学班 67 名学生的《人工智能》考试成绩(Final Total), 并通过问卷调查获得他们下面几门课程的成绩: 大学英语四级(CET4)、六级(CET6), 《线性代数》(Linear Algebra), 《概率论与数理统计》(Probability & Statistics)。最终, 我获得了 31 位学生的上述成绩(图 1 与附件 1), 用于统计分析(附件 2)。

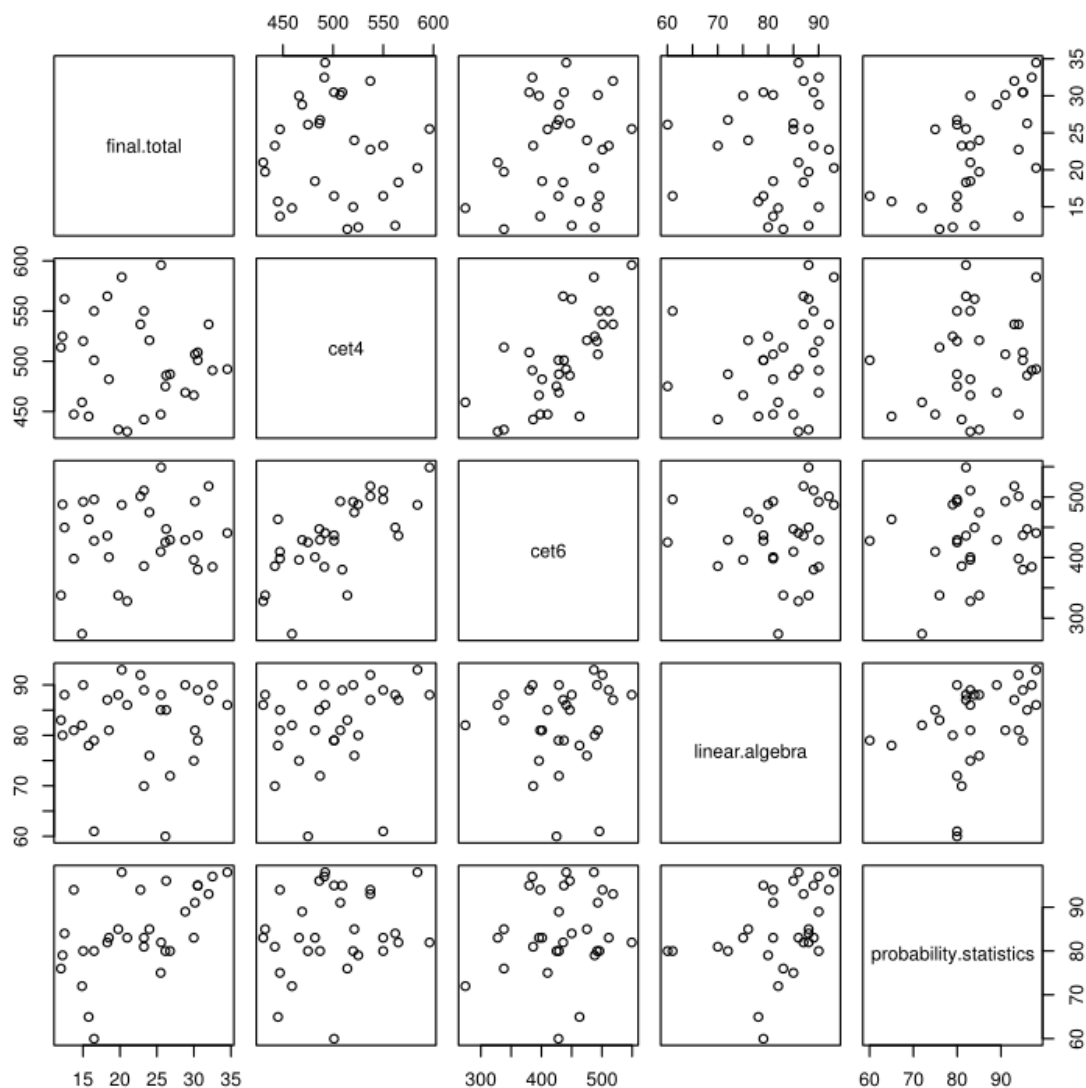


图 1: 5 个变量的散点图矩阵 (scatterplot matrix)。在对角线上, 左上角的 final.total 为响应变量, 其余 4 个变量, cet4, cet6, linear.algebra, probability.statistic 为预测变量。仔细观察可以发现, cet4 与 cet6(第 2 行第 3 列), final.total 与 probability.statistics(第 1 行第 5 列)有线性关系。有趣的是, cet4 与 final.total 没有线性关系(第 1 行第 2 列), cet6 与 final.total 同样也没有线性关系(第 1 行第 3 列)。

结果

响应变量 Final Total 与预测变量 CET4, CET6, Linear Algebra, Probability & Statistics 的关系

为得出 Final Total (《人工智能》课程期末考试成绩) 与其余 4 个预测变量的关系, 考虑以下多元线性回归模型:

$$F = \beta_0 + \beta_1 \text{ CET4} + \beta_2 \text{ CET6} + \beta_3 \text{ L} + \beta_4 \text{ S} + \epsilon,$$

其中响应变量 F 代表 Final Total, 预测变量 L 代表 Linear Algebra, 预测变量 S 代表 Probability & Statistics。估计参数我们得到 $F \approx 5.28 - 0.055 \text{ CET4} + 0.032 \text{ CET6} - 0.077 \text{ L} + 0.44 \text{ S}$ 。

检验如下假设:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,$$

Ha: 上面至少有一个 β_j 不为0。

零假设 H_0 是说 CET4, CET6, Linear Algebra, Probability & Statistics 这四个预测变量全部与响应变量 Final Total 无关。

检验结果如下:

final.total ~ cet4 + cet6 + linear.algebra + probability.statistics				
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.27864	14.46217	0.365	0.718066
cet4	-0.05530	0.03270	-1.691	0.102771
cet6	0.03192	0.02294	1.391	0.175972
linear.algebra	-0.07683	0.13900	-0.553	0.585162
probability.statistics	0.44375	0.11848	3.745	0.000905 ***
Residual standard error: 5.553 on 26 degrees of freedom				
Multiple R-squared: 0.4087, Adjusted R-squared: 0.3177				
F-statistic: 4.492 on 4 and 26 DF, p-value: 0.006822				

因为整体 p-value 是 0.006822, 所以拒绝 H_0 。即四个预测变量中至少有一个与 Final Total 有关系。(p-value 代表假如零假设是对的, 我们在服从零假设的分布上观察到统计量 *F-statistic* 大于等于 4.492 的概率。p-value 越小, 则零假设 H_0 越不可能成立。一般情况下, 如果 p-value 小于 0.05, 就拒绝零假设。)就每个预测变量来看, Probability & Statistics 与《人工智能》的期末成绩 Final Total 有显著的正关系(系数是正数 0.44 且 p-value = 0.0009)。这可以解释得通, 因为本课程与统计(statistics)有关, 教科书的书名就含有 statistical 这个单词^[2], 且期末考试内容与概率相关的内容比重大。固定其它三个因素(CET4, CET6 与 Linear Algebra), probability.statistics 每增加一个单位, 则 Final Total 增加 0.44 个单位。CET4、CET6 与 Final Total 均没有显著线性关系(p-value 均大于 0.05, 且前面的系数均接近 0)。同样, Linear Algebra 也没有, 这也许与本次《人工智能》期末考试很少用到线性代数知识有关。

既然《概率论与数理统计》(Probability & Statistics)与《人工智能》的期末成绩 Final Total 有显著的正关系, 那么就检查去掉预测变量 probability.statistics 的影响。为此, 对比整体模型 final.total ~ cet4 + cet6 + linear.algebra + probability.statistics 与去除了 probability.statistics 这个预测变量的部分模型 final.total ~ cet4 + cet6 + linear.algebra 的 RSS (Residual Sum of Squares), 来检验零假设 H_0 : probability.statistics 前的系数 $\beta_4=0$ 是否成立。

检验结果如下:

Analysis of Variance Table					
Model 1: final.total ~ cet4 + cet6 + linear.algebra					
Model 2: final.total ~ cet4 + cet6 + linear.algebra + probability.statistics					
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	27 1234.40				
2	26 801.81	1	432.59	14.027	0.0009051 ***

F -statistic=14.027 对应的 p -value 是 0.0009051, 拒绝 H_0 。所以 probability.statistics 对 Final Total 有显著贡献。注意到这里 F -statistic=14.027 所对应的 p -value 与上面多元线性回归 t -statistic=3.745 对应的 p -value 是一样的。 F -statistic=14.027 反映了加入 probability.statistics 这个变量后对部分模型产生的影响 (partial effect)^[2]。 F 越大, 影响越显著。

检查去掉预测变量 cet4 与 cet6 的影响。对比整体模型 $\text{final.total} \sim \text{cet4} + \text{cet6} + \text{linear.algebra} + \text{probability.statistics}$ 与去除了 cet4 与 cet6 这两个预测变量的部分模型 $\text{final.total} \sim \text{linear.algebra} + \text{probability.statistics}$ 的 RSS (Residual Sum of Squares), 来检验零假设 H_0 : CET4 前的系数 $\beta_1 = \text{CET6}$ 前的系数 $\beta_2 = 0$ 是否成立。

检验结果如下:

Analysis of Variance Table						
Model 1: final.total ~ linear.algebra + probability.statistics						
Model 2: final.total ~ cet4 + cet6 + linear.algebra + probability.statistics						
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	28	892.69				
2	26	801.81	2	90.881	1.4735	0.2476

F -statistic=1.4735, 接近 1, 所对应的 p -value 是 0.2476, 无法拒绝 H_0 。即, 证据表明, CET4 与 CET6 这两个预测变量合起来对 Final Total 并无显著贡献。CET4、CET6 成绩与《人工智能》期末考试成绩无关。

变量之间的两两关系

参考图 1, 每次分析只考虑两个变量, 将其中一个变量作为响应变量 Y , 另外一个变量作为预测变量 X , 做一元线性回归, $Y = \beta_0 + \beta_1 X + \epsilon$, 检验如下假设:

H_0 : Y 与 X 无关系。

H_a : Y 与 X 有关系。

让 Y 代表 Final Total。让 X 分别代表 CET4, CET6, Linear Algebra 与 Probability & Statistics。Final Total 与 CET4 之间没有线性关系, p -value=0.719。Final Total 与 CET6 之间没有线性关系, p -value=0.448。Final Total 与 Linear Algebra 之间没有线性关系, p -value=0.725。Final Total 与 Probability & Statistics 之间有十分显著线性关系, X 前的系数是 0.4036, p -value=0.00099, 这与多元线性回归得到的结论一致。

两种数学成绩之间应具有相关性。让 Y 代表 Linear Algebra, X 代表 Probability & Statistics。Linear Algebra 与 Probability & Statistics 有显著线性关系 (p -value=0.035, X 前的系数 0.33)。这说明我获得的数学成绩是合理的。

CET4 成绩与 CET6 成绩也应该有相关性。让 Y 代表 CET6, X 代表 CET4。分析显示, CET4 与 CET6 的线性关系异常显著 (p -value=1.53e-05, X 前的系数是 0.97)。这容易解释, 因为 CET6 只是在 CET4 上增加了难度, 可是考察的知识基本一致。这说明我获得的英语数据也是合理的。

值得注意的是, 上《人工智能》课的中国学生的数学能力与他们的英语能力并没有显著线性关系。对 CET4 与 probability.statistics 做一元线性回归分析, p -value=0.297。对 CET4 与 linear.algebra 做一元线性回归分析, p -value=0.176。对 CET6 与 probability.statistics 做一元

线性回归分析, $p\text{-value}=0.356$ 。对 CET6 与 linear.algebra 做一元线性回归分析, $p\text{-value}=0.753$ 。如此看来, 数学与英语需要不相关的学习能力。

图 2 给出变量之间的两两相关性系数热力图, 颜色越深代表相关性系数越大。相关性系数矩阵在附件 3 中。可以看到, CET4 与 CET6, Final Total 与 Probability & Statistics, Linear Algebra 与 Probability & Statistics 有显著相关性。

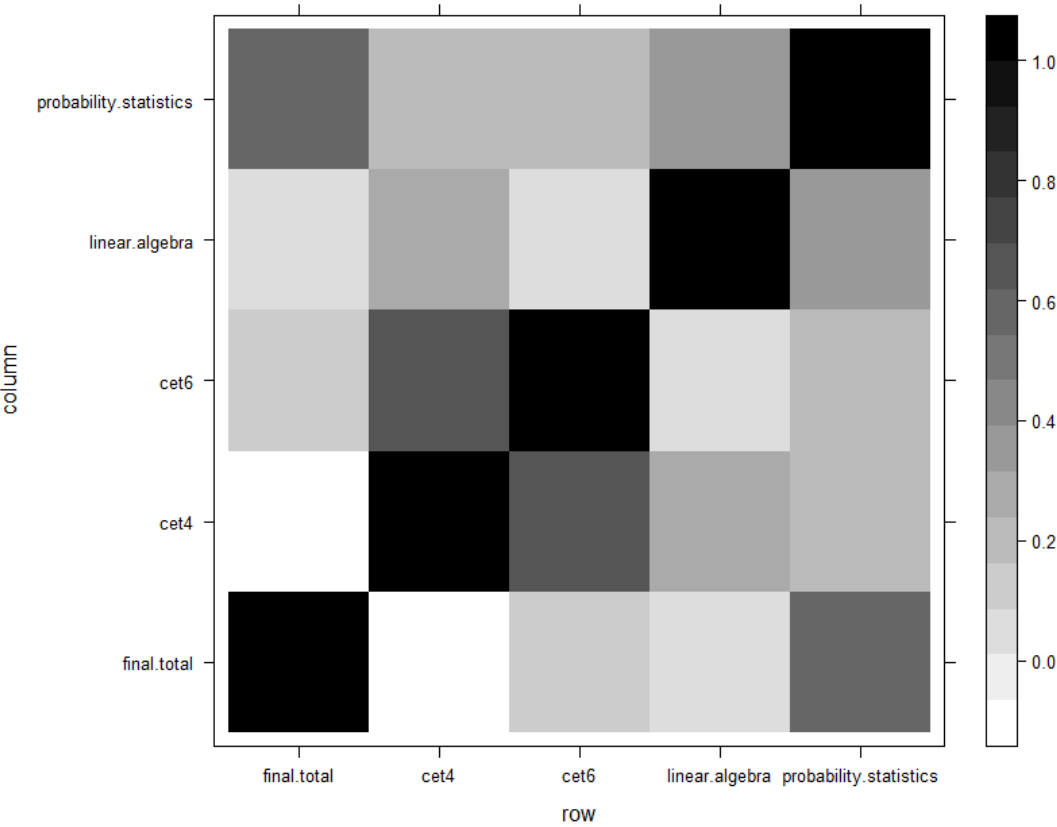


图 2: 本研究中 5 个变量的两两相关性强弱。选取相关性系数大于等于 0.3 的两个变量, 按照从大到小排序, 分别是 $\text{corr}(\text{cet4}, \text{cet6}) = 0.69$, $\text{corr}(\text{probability.statistics}, \text{final.total}) = 0.56$, $\text{corr}(\text{probability.statistics}, \text{linear.algebra}) = 0.38$ 。

结论与讨论

从我 2018 年春季在浙师大所教的《人工智能》班级来看, 全国大学英语等级考试成绩与采用英文原版教材的《人工智能》期末考试成绩无关。这个结论也被 2019 年春季《人工智能》班级的数据所证实, 见附件 4。与《人工智能》期末考试成绩关系最为密切的是《概率论与数理统计》成绩。有趣的是,《人工智能》成绩与《线性代数》成绩也没有显著关系, 虽然《概率论与数理统计》成绩与《线性代数》成绩有关系。前者可能是因为此次《人工智能》期末考试很少涉及线性代数知识。

英语成绩好坏无法预测《人工智能》成绩好坏, 即便本课程采用的是对英文程度好的学生有利的英文教材。而《概率论与数理统计》成绩则可以。根据上述研究结果, 若要在人工智能领域有所成就, 我建议浙师大理工科专业把教学重点放在《概率与数理统计》基础知识上。

启蒙时期英国实用主义哲学家约翰·洛克指出在拉丁文上花过多精力是没有必要的

(§147), 特别是如果以后英国孩子用不到拉丁文简直是浪费时间(§164)^[3]。尽管本研究否定了英文(阅读理解)能力对人工智能这样的计算机科学课程学习起到正面或负面作用,但是无法忽视当前英文在世界范围所起的学术交流作用。因此,为了提高国家时间利用率,我建议优化外语学习策略:注意挑选一部分对外语有天赋的中国人来起这个交流作用,同时着重改进智能翻译系统,比如有道翻译官、有道翻译王等。

本次研究还未解决的问题是,上述结论是否只是在浙江师范大学适用,还是对浙江其它高校也适用。是否只对《人工智能》课程适用,还是对其它计算机课程也适用。将来值得研究的问题有:(1)如果采用全英文授课,结论是否会不同。(2)全英文授课模式下,以中文为母语的老师的讲授与以英文为母语的老师的讲授的效果有何差别。

附件

附件 1：2018 年春季学期人工智能教学班成绩数据 AIClassData。学生学号保留，姓名隐去。其中 final.total 总分是 40 分。

student.no	student.name	final.total	q3a	q1a	q1b	q1c	q1d	q1e	q1f	q1g	q1h	q1i	q1j	q2	q3	q4	q5	cet4	cet6	linear.algebra	probability.statistics
201530800126	confidential	26.6	2	1	1	1	1	1	1	1	1	1	-1	6	6.5	1	3.85	450	438	NA	84
201531900101	confidential	22	2	1	1	1	1	1	-1	1	1	-1	1	3.5	4.5	1	4.5	NA	NA	NA	NA
201531900102	confidential	26.125	2	1	1	1	1	1	1	0	1	0	1	5	5	1	6.625	475	425	60	80
201531900103	confidential	13	0.5	1	1	1	0	1	0	1	0	1	1	2.5	2	1	2	NA	NA	NA	NA
201531900104	confidential	23.75	0.5	1	1	1	0	1	1	1	1	1	1	8.5	3	1	2.75	NA	NA	NA	NA
201531900105	confidential	22.75	2	1	-1	1	0	1	0	0	1	1	1	5.5	4.5	1	7	537	501	92	94
201531900106	confidential	21	0	1	1	1	1	0	1	1	0	1	1	4.5	3.5	1	5	430	328	86	83
201531900107	confidential	30	2	1	0	0	-1	1	1	-1	1	1	1	8	7.5	1	8	NA	NA	NA	NA
201531900108	confidential	19.75	1	0	1	1	1	-1	-1	-1	1	1	1	2.75	3.5	1	6.25	NA	NA	NA	NA
201531900109	confidential	26.75	2	1	1	1	-1	1	1	1	0	1	1	5	7.5	1	5.5	487	429	72	80
201531900110	confidential	24.25	2	1	-1	1	1	1	1	1	1	1	1	2.5	6	1	5.5	440	NA	91	97
201531900111	confidential	26.25	1	1	1	1	0	-1	1	-1	1	1	1	5.5	5.5	1	7.25	486	447	85	96
201531900112	confidential	16.5	2	0	1	0	0	1	0	0	1	1	1	3	4	1	6	550	496	61	80
201531900113	confidential	25.55	2	1	1	1	0	1	1	1	0	1	1	4.75	7.5	1	5.3	596	549	88	82
201531900114	confidential	21.25	2	1	1	1	0	1	0	0	1	1	1	4	5.5	1	5.25	NA	NA	NA	NA
201531900116	confidential	30	2	-1	1	1	1	1	1	1	1	1	1	8.5	5.5	1	5.75	466	396	75	83
201531900117	confidential	19.75	0.5	-1	-1	1	0	1	1	0	0	1	1	4.5	4.5	1	6.5	NA	NA	NA	NA
201531900118	confidential	15	0.5	-1	1	1	0	1	-1	-1	-1	1	-1	4	2	1	3.25	433	NA	73	62
201531900119	confidential	15.25	0.5	1	0	-1	0	-1	1	-1	1	1	-1	5.25	3	1	2	NA	NA	NA	NA
201531900120	confidential	12.25	0.5	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	4	2	1	2	525	488	80	79
201531900121	confidential	17.25	2	0	1	-1	0	1	-1	1	-1	1	1	3.5	2.5	1	5.5	418	0	69	80
201531900122	confidential	16.25	0.5	1	1	-1	1	1	0	-1	-1	1	-1	3.25	3	1	3.5	NA	NA	NA	NA
201531900123	confidential	25	2	-1	-1	1	-1	-1	0	-1	1	1	-1	5.5	7.5	1	7	NA	NA	NA	NA

201531900124	confidential	17.8	0	-1	1	0	1	-1	-1	-1	1	-1	-1	4.5	2	1	6.3	NA	NA	NA	NA
201531900125	confidential	15.75	2	1	1	1	0	1	-1	0	1	1	1	2	4	1	2.5	445	463	78	65
201531900126	confidential	14.25	0	1	1	1	0	1	1	0	1	1	1	3.5	0	1	2.5	NA	NA	NA	NA
201531900127	confidential	16.5	1	0	1	1	-1	-1	-1	-1	0	1	-1	4	2.5	1	5.75	NA	NA	NA	NA
201531900128	confidential	18.5	2	1	-1	1	1	0	-1	-1	-1	0	1	4	4	1	5.5	482	401	81	83
201531900129	confidential	13	0.5	-1	1	1	-1	-1	-1	-1	-1	-1	-1	3.25	2	1	2.75	432	NA	69	82
201536000102	confidential	32.5	2	1	-1	1	1	1	1	-1	1	1	1	9	7.5	0.5	7	491	385	90	97
201536000110	confidential	23.7	2	1	1	0	0	1	1	0	1	1	1	6.5	6	0.75	4.95	NA	NA	NA	NA
201536810115	confidential	12.5	0	1	0	0	1	1	1	1	1	1	0	2.75	0.5	0.75	3	562	450	88	84
201531960109	confidential	22.75	0.5	1	1	1	0	1	1	1	1	1	1	4.75	3.5	1	5	476	NA	91	82
201536000103	confidential	32	2	1	1	1	1	1	1	1	1	1	1	7	7.5	0	7.5	537	518	87	93
201536000104	confidential	23.25	2	1	1	1	0	1	1	-1	1	1	1	6.5	4	1	4	442	386	70	81
201536000105	confidential	28.8	2	1	1	1	1	0	1	1	1	1	1	7	7.5	1	4.8	469	429	90	89
201536000106	confidential	22.8	0	1	1	1	1	1	0	0	1	1	1	6.5	3.5	1	4.8	NA	NA	NA	NA
201536000107	confidential	19.75	0.5	1	0	1	1	1	1	-1	0	1	1	3.25	6	1	3.25	405	NA	75	60
201536000108	confidential	24	1.5	1	1	1	1	1	1	0	1	1	1	6.5	2.5	1	5.5	521	475	76	85
201536000109	confidential	30.125	2	1	1	1	1	1	1	1	1	1	1	7	5	1	7.125	507	493	81	91
201536000111	confidential	15	0.5	1	1	1	1	1	0	0	0	1	1	3.25	2	0.5	3.75	398	NA	68	62
201536000112	confidential	18.3	0.5	1	1	1	1	1	1	0	0	1	1	2.5	4	1	3.8	565	436	87	82
201536000113	confidential	23.85	2	1	-1	1	0	1	-1	1	1	1	-1	6	5	1	5.6	NA	NA	NA	NA
201536000114	confidential	25.5	2	1	1	1	1	0	1	1	1	1	0	5.5	5.5	1	6.5	447	410	85	75
201536000115	confidential	24	2	1	1	1	1	0	1	1	1	1	0	5.25	6	1	4.75	NA	NA	NA	NA
201536000117	confidential	23.25	2	1	-1	1	1	1	1	-1	1	1	1	5.25	5	1	3.5	NA	NA	NA	NA
201536000118	confidential	28.5	2	1	1	1	1	-1	1	-1	1	1	1	4.5	7.5	0	8	380	0	80	85
201536000119	confidential	14.85	0.5	1	0	1	1	1	1	0	0	1	0	3.25	2	1	4.6	459	274	82	72

201536000120	confidential	19.75	2	1	1	1	1	-1	-1	-1	0	-1	-1	6	4	1	4.5	432	338	88	85
201536000121	confidential	30.5	1	1	1	1	1	1	1	1	1	1	1	5.5	6.5	0.5	8	509	380	89	95
201536000122	confidential	13.5	0.5	0	1	-1	0	1	-1	0	-1	1	1	4	2	1	3.25	380	NA	68	71
201536000123	confidential	16.5	2	-1	1	0	-1	1	-1	-1	0	1	-1	4	3.5	1	4.75	501	428	79	60
201536000124	confidential	12	1	0	1	1	0	1	0	0	1	1	0	3.5	1	1	4	514	338	83	76
201536000126	confidential	17.25	0.5	1	1	1	0	1	1	1	0	0	1	4.75	1.5	1	4.5	NA	NA	NA	NA
201536000127	confidential	16.8	2	1	1	1	1	1	0	0	0	1	0	3.5	3.5	1	4.8	NA	NA	NA	NA
201536000128	confidential	21.75	2	1	1	1	1	1	0	0	1	1	1	5	4.5	1	4.25	NA	NA	NA	NA
201536000129	confidential	34.5	2	1	1	1	1	1	1	1	1	1	1	9.5	7.5	1	6.5	492	441	86	98
201536000130	confidential	25	2	0	1	0	1	1	1	1	1	1	1	4.5	7.5	1	5	NA	NA	NA	NA
201536000131	confidential	30.5	2	1	1	-1	1	1	1	1	1	1	1	6.75	6.5	2	6	501	437	79	95
201536210107	confidential	24	2	1	1	1	1	1	1	0	0	1	1	7	4	1	5	NA	NA	NA	NA
201539100108	confidential	23.25	0.5	1	1	1	-1	1	1	-1	-1	-1	1	4.75	3	0.5	8	550	511	89	83
201539100111	confidential	15	0.5	0	1	1	1	0	1	-1	0	1	0	4	2	1	5	520	492	90	80
201539100113	confidential	13.75	2	0	1	1	1	-1	1	-1	0	1	-1	3	5	0	1	447	398	81	94
201539300151	confidential	20.25	0.5	1	1	1	1	1	1	1	1	1	1	2.5	4.5	1	2.25	584	487	93	98
201539100105	confidential	NA	2	1	1	1	1	1	1	0	1	1	1	8.75	NA	NA	NA	505	464	93	94
201539100112	confidential	NA	2	1	1	1	1	1	1	0	1	1	1	9.5	NA	NA	NA	NA	NA	NA	NA
201539100114	confidential	NA	2	1	1	1	1	1	1	0	1	1	1	8.5	NA	NA	NA	531	478	95	92

附件 2: R 分析脚本 analyze.R。

```
# Copyright 2018 Hui Lan lanhui@zjnu.edu.cn

# Load data and make relevant columns for this study.
X <- read.table('AIClassData.txt', header=T, check.names=FALSE, sep='\t')
q1 <- apply(X[,5:14], 1, sum)
Y <- cbind(q1, X[,c(3, 4, 15, 16, 17, 18, 19, 20, 21, 22)])
Z <- Y[,c(2,8,9,10,11)]

# Make Z2, the data I used for the subsequent analysis.
index <- !is.na(Z$final.total) & !is.na(Z$cet4) & Z$cet4 > 0 & !is.na(Z$cet6) & Z$cet6 >
0 & !is.na(Z$linear.algebra) & !is.na(Z$probability.statistics)
Z2 <- Z[index,] # rows with no missing values.

# Draw a scatterplot matrix for Z2
pairs(Z2)

# The partial model takes out cet4 and cet6
full <- lm(final.total ~ cet4 + cet6 + linear.algebra + probability.statistics, data=Z2)
summary(full)

partial <- lm(final.total ~ linear.algebra + probability.statistics, data=Z2)
summary(partial)

anova(partial, full) # Refer to the following link for how to do ANOVA:
http://www.stat.columbia.edu/~martin/W2024/R6.pdf

# The partial model takes out probability.statistics
full <- lm(final.total ~ cet4 + cet6 + linear.algebra + probability.statistics, data=Z2)
summary(full)

partial <- lm(final.total ~ cet4 + cet6 + linear.algebra, data=Z2)
summary(partial)

anova(partial, full)

# Make a matrix of correlation coefficients and draw a heatmap for this matrix
C = cor(Z2, use='pairwise.complete.obs')
require(lattice)
levelplot(C)
levelplot(C, col.regions = gray(256:00/256))
```

```
# Multiple regression
full <- lm(final.total ~ cet4 + cet6 + linear.algebra + probability.statistics, data=Z2)
summary(full)

# Pairwise simple linear regressions
m <- lm(linear.algebra ~ probability.statistics, data=Z2)
summary(m)

m <- lm(cet6 ~ cet4, data=Z2)
summary(m)

m <- lm(cet4 ~ linear.algebra, data=Z2)
summary(m)

m <- lm(cet4 ~ probability.statistics, data=Z2)
summary(m)

m <- lm(cet6 ~ linear.algebra, data=Z2)
summary(m)

m <- lm(cet6 ~ probability.statistics, data=Z2)
summary(m)

m <- lm(probability.statistics ~ cet6, data=Z2)
summary(m)

m <- lm(final.total ~ linear.algebra, data=Z2)
summary(m)

m <- lm(final.total ~ probability.statistics, data=Z2)
summary(m)

m <- lm(final.total ~ cet4, data=Z2)
summary(m)

m <- lm(final.total ~ cet6, data=Z2)
summary(m)
```

附件 3：相关性系数矩阵 CorrelationMatrix

.	final.total	cet4	cet6	linear.algebra	probability.statistics
final.total	1.00	-0.07	0.14	0.07	0.56
cet4	-0.07	1.00	0.69	0.25	0.19
cet6	0.14	0.69	1.00	0.06	0.17
linear.algebra	0.07	0.25	0.06	1.00	0.38
probability.statistics	0.56	0.19	0.17	0.38	1.00

附件 4:

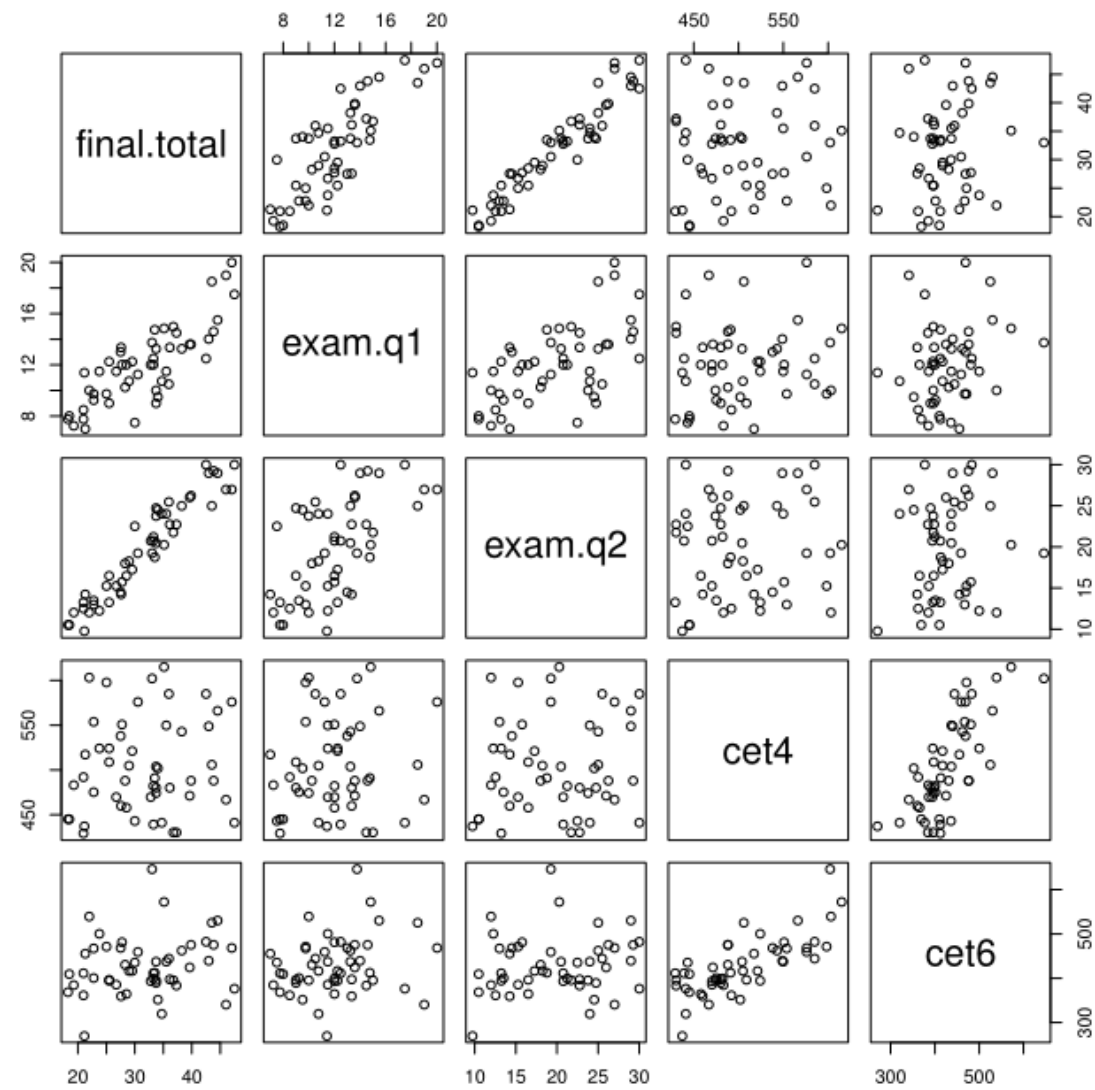
- a. 2019 年春季学期人工智能教学班的成绩数据 AIClassDataSpring2019。列 exam.q1 是期末考试第一个问题（分值 20 分），列 exam.q2 是期末考试第二个问题（分值 30 分），本次考试共两个问题，总分是 50 分。时间有限，我没有收集该教学班的《线性代数》与《概率论与数理统计》成绩。

student.no	student.name	exam.q1	exam.q2	cet4	cet6
201535500111	confidential	18.5	25	506	525
201631900101	confidential	13.25	25	543	462
201631900103	confidential	13.35	22.75	480	398
201631900104	confidential	14.85	20.25	615	572
201631900105	confidential	9.25	13.5	475	401
201631900106	confidential	10.5	25.5	585	444
201631900108	confidential	14.5	22.75	430	384
201631900109	confidential	7.75	13.25	429	412
201631900110	confidential	11.5	15.25	470	386
201631900111	confidential	7.65	16.75	360	NA
201631900112	confidential	9	24.75	480	390
201631900113	confidential	13.35	14.25	460	360
201631900114	confidential	10	23.75	474	396
201631900115	confidential	8	11.5	438	NA
201631900116	confidential	11.5	24	550	437
201631900117	confidential	9	16.5	509	397
201631900118	confidential	14	29	549	440
201631900119	confidential	10.75	18.25	505	417
201631900120	confidential	15.5	29	566	530
201631900121	confidential	13.6	26	471	425
201631900122	confidential	12	20.75	470	394
201631900123	confidential	7.25	12	483	385
201631900124	confidential	7.75	10.5	445	369
201631900125	confidential	12	16.5	458	365
201631900126	confidential	8.5	12.5	492	362
201631900127	confidential	8.9	8	430	NA
201631900128	confidential	8.25	12.5	420	NA
201631900129	confidential	5.75	10.75	437	NA
201631900130	confidential	9.75	13	554	467
201631900132	confidential	8.5	11.75	474	NA
201631900133	confidential	10.25	16.25	471	NA
201632000102	confidential	11.25	19.25	576	459
201632000112	confidential	12.5	19.25	NA	NA
201632000113	confidential	15.25	28	411	NA
201632120101	confidential	12	15	NA	NA

201632120116	confidential	13	14.5	538	469
201633030112	confidential	13.75	19.25	602	646
201635500126	confidential	19	27	467	341
201635500129	confidential	20	27	576	469
201635500226	confidential	7.5	22.5	443	436
201636000101	confidential	12.25	13.25	524	395
201636000102	confidential	12	15.75	551	481
201636000103	confidential	12.25	17.25	521	417
201636000104	confidential	8.25	21.5	441	NA
201636000105	confidential	8	10.5	445	410
201636000106	confidential	12	21.25	482	400
201636000107	confidential	11.5	12.25	524	500
201636000108	confidential	11.75	18.5	380	NA
201636000109	confidential	13.5	9	497	NA
201636000110	confidential	12.1	26	466	NA
201636000111	confidential	7.25	17.25	351	NA
201636000112	confidential	14.6	29.25	488	476
201636000113	confidential	7	14.25	517	455
201636000115	confidential	11.4	9.75	437	270
201636000116	confidential	10.25	18	488	431
201636000117	confidential	12.5	30	585	483
201636000118	confidential	15	21.75	430	396
201636000119	confidential	12.5	20.75	439	412
201636000121	confidential	16	23	424	NA
201636000122	confidential	13.25	20.5	504	437
201636000123	confidential	14.25	24	491	NA
201636000124	confidential	14.75	18.75	491	413
201636000125	confidential	12.5	28	453	NA
201636000126	confidential	10.75	25.5	504	NA
201636000127	confidential	9.5	24.5	502	352
201636000128	confidential	10.75	24	441	320
201636000129	confidential	13.6	26.25	488	476
201636000130	confidential	17.5	30	441	377
201636100119	confidential	9.75	15.25	598	471
201636810120	confidential	10	12	603	539

- b. 2019 年春季学期人工智能教学班的成绩数据总览。只保留 exam.q1, exam.q2, cet4, cet6 均有数据的 51 行。final.total 是 exam.q1 与 exam.q2 之和。final.total 与 cet4 的相关性系数 0.12（散点图见下图第一行第四个方块），final.total 与 cet6 的相关性系数是 0.20（散点图见下图第一行第五个方块）。以 final.total 为响应变量，以 cet4 与 cet6 为预测变量，做二元线性回归，得 $F \approx 23.60 - 0.008 \text{ CET4} + 0.028 \text{ CET6}$ 。回归模型的 *F-statistic* 是 0.9401, p-value 是 0.40。无法拒绝 CET4, CET6 这两个预

测变量与响应变量 F 无关的零假设。



c. 相关性系数矩阵 CorrelationMatrixSpring2019

.	final.total	exam.q1	exam.q2	cet4	cet6
final.total	1.00	0.81	0.95	0.12	0.19
exam.q1	0.81	1.00	0.60	0.13	0.22
exam.q2	0.95	0.60	1.00	0.10	0.15
cet4	0.12	0.13	0.10	1.00	0.75
cet6	0.19	0.22	0.15	0.75	1.00

致谢

我感谢以下同学参与问卷调查：安静、丁芳、杜智月、李基梅、林佳、刘旭蕾、彭雨露、杨寒霞、蔡高华、沈丹瑞、胥成、陈依茹、唐燕、洪诗诗、刘亚昕、罗荣华、彭悦、史慧、邢璐、袁红爽、陈文中、崔元泽、丁光耀、韩俞雷、李冠桦、张骏、张胜、李海霞、谢宜桦、姚芳、臧纹萱，金伦，毕奏，何芳，金玲彬，李慧敏，李佳，刘湘玥，马永芳，钱鑫玉，施佳佳，王莉，邢冉冉，袁周，曾雨佳，张梦瑶，张忆莲，张玥，周诗璇，周赞，朱军娇，蔡军碧，陈昌东，范豪杰，吕梁，任元钊，王国俊，王永峰，韦明山，谢约翰，张鑫，章敬洋，何宇焱，黄帅麒，徐玉婷，李佳慧，宋云鹏，翁嘉诚，王健博，蔡小雨，陈静，陈薇，江珊，李凌鹤，李思琪，林恩慧，林媛娜，刘佳雨，刘星雨，罗珺，史江婷，吴逸嘉，张欣，赵佳桢，赵娴，赵洋，陈俊义，郭起，黄梓祺，刘凯诚，钱晨阳，王超峰，吴金康，熊虎成，虞协俊，张哲，周鹦荐，周婧哲，张翌。

参考文献

- [1] 郑琳，王湛，林晓莹，石丛珊. (2018). 海归老师越来越多，杭城高校刮起全英文授课风. 浙江在线 9 月 13 日讯. <https://zj.zjol.com.cn/news.html?id=1029520>
- [2] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer New York Heidelberg Dordrecht London. Corrected at 8th printing 2017.
- [3] John Locke. (1690) The Works of John Locke, vol. 8 (Some Thoughts Concerning Education, Posthumous Works, Familiar Letters). London: Rivington, 1824 12th ed.