

多分类器组合的情感分析方法研究

王超名 陈旭东

(北京交通大学 软件学院 软件工程)

摘要: 近年来,随着 O2O 的迅速发展,外卖领域产生了大量的文本评论。在大数据时代,这些包含用户主观情感倾向的文本蕴含着大量有用的信息,人工分析文本的情感倾向效率极低,所以,利用计算机对文本进行情感分析具有巨大的应用价值。本文针对外卖领域的文本评论,对情感极性分析的常用方法进行了系统研究。本文首先分析了基于 k-近邻、朴素贝叶斯、最大熵模型、支持向量机的情感极性分析算法,发现基于最大熵模型和支持向量机的算法准确率最高。接着,本文在观察各个分类器分类结果的基础上,发现单个分类器都有各自的缺点,于是本文尝试将多个分类器组合起来以期克服其缺点。本文探究了多分类器组合的两种方式 and 组合后的两种决策策略,在标准康奈尔影评数据集上,最好的组合策略能将准确率提升 6.5%。另外,在本文构建的外卖数据集上,该方法同样有效。

关键词: 多分类器融合; 情感分析

Sentiment Analysis Research on Multiple Classifiers Combination

Chaoming Wang, Xudong Chen

Abstract: Recent years, with the rapid development of online to offline (O2O) business model,

Keywords: Multiple Classifier Combination; Sentiment Analysis

基金项目: 北京交通大学大学生科研训练项目资助 (150185017)

作者简介: 王超名 (1995—), 男, 本科 2014 级, 主要研究方向为情感分析; 陈旭东 (), 男, 副教授, 主要研究方向网络技术与管理信息系统, 情感分析。

1 引言

近年来,随着 O2O 的迅速发展,外卖领域产生了大量的文本评论。在大数据时代,这些包含用户主观情感倾向的文本蕴含着大量有用的信息,然而人工分析文本的情感倾向效率极低,所以,利用计算机对文本进行情感分析具有巨大的应用价值。

情感分析在粒度上可以有三种划分: 词语粒度、句子粒度和篇章粒度。¹而每种粒度下,又有其相应的几个子任务。²

1)、篇章粒度。篇章粒度上的情感分析主要考虑整个篇章的情感极性。Lee et al. 将传统机器学习方法应用于电影评论的情感分析,最后得到从篇章中提取出 unigrams 后 SVM 的分类效果最好。³这三种机器学习方法都是有监督的方法。而 Peter D. Turney 提出了一种无监督的方法进行篇章情感极性的自动分类。⁴篇章粒度有几个子任务,都有相应的研究尝试解决。比如,主客观分类⁵、观点有效性检测⁶、垃圾观点检测⁷。

2)、句子粒度。句子粒度的情感分析首先是判断句子是否主观或客观,然后将主观语句进行极性分类。Weibe et al. 提出主观性分析在句子中的有效性及其在自然语言处理的应用。⁸熊德兰等人提出了基于 HowNet 的语义距离和语法距离相结合的句子褒贬倾向性计算方法利用夹角余弦法对语义倾向进行了改进。⁹句子粒度的子任务有: 观点检索¹⁰、观点问答¹¹、

观点摘要¹²。

3)、词语粒度。词语粒度的情感分析主要是提取出评价对象，评价观点，实体，属性，特征及其情感倾向。¹³¹⁴¹⁵

本文的工作是尝试对外卖领域的文本，进行句子粒度的情感分析，归纳总结基于机器学习的有监督的情感极性分析方法。最后提出多种分类器并联+串联的混合融合策略，使用置信平均决策或投票决策，来决定句子的情感极性。结果发现，准确率有极大提升。

本文内容组织如下：第二部分是情感分析的预处理技术，第三部分是基于情感词典的情感极性分析方法，第四部分是基于机器学习（k-近邻算法、朴素贝叶斯、最大熵模型、支持向量机）的情感分析方法，第五部分是本文提出的多分类器融合的情感分析方法。

2 基于机器学习的情感分析

2.1 k-近邻算法

k-近邻算法简单直观。给定一个训练数据集，对新的输入实例，在训练数据集中找到与该实例最近邻的 k 个案例，这 k 个案例的多数属于某个类，就把该输入实例分为这个类。¹⁶

基于 K-近邻算法的情感分析的一般步骤为：

- 1、计算已知类别的数据集中的每个句子与当前句子之间的距离；
- 2、按照距离递增次序进行排序；
- 3、选取与当前句子距离最小的 k 个句子；
- 4、确定这 k 个句子所在类别的出现频率；
- 5、返回前 k 个句子出现频率最高的类别作为当前点的预测分类。

2.2 朴素贝叶斯

假设句子 X 包含词汇 (a_1, a_2, \dots, a_m) 。Y 为情感极性类别的集合 (y_1, y_2) 。则 X 的情感极性为 $\operatorname{argmax}(P(y_i|x)) = \operatorname{argmax}(P(y_i|x_1, x_2, \dots, x_m))$ 。由贝叶斯公式

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)}$$

$P(x)$ 是相同的，可直接忽略。所以，只需要找到使得 $P(x|y_i)P(y_i)$ 最大的 y_i 。

朴素贝叶斯 (Naïve Bayes) 被称为 naïve 是因为假设各个词之间是相互独立的，于是：

$$P(x|y_i)P(y_i) = P(a_1|y_i)P(a_2|y_i) \dots P(a_m|y_i)P(y_i) = P(y_i) \prod_{j=1}^m P(a_j|y_i)$$

Naïve Bayes 的条件独立假设看上去很傻很天真，但结果却很强大。因为有些独立假设在各个分类之间的分布都是均匀的，所以对于似然的相对大小不产生影响；即便不是如此，也有很大的可能性各个独立假设所产生的消极影响或积极影响互相抵消，最终导致结果受到的影响不大。¹⁷

2.3 最大熵

最大熵原理指出，当我们需要对一个随机事件的概率分布进行预测时，我们的预测应当满足全部已知的条件，而对未知的情况不要做任何主观假设。在这种情况下，概率分布最均匀，预测的风险最小。因为这时概率分布的信息熵最大，所以人们称这种模型叫“最大熵模型”。以下分别介绍了从数学上如何 1)、满足全部已知条件，2)、对未知条件尽可能均匀。

2.3.1 满足所有已知条件

我们假设 p 为最大熵模型的学习后的所有条件概率， $p(y|x)$ 为 p 中的一个元素。训练集内包含大量的样本， $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ 。我们定义 $\tilde{p}(x, y)$ 为样本 (x, y) 出现的概率。从训练集中我们可以得到很多有用的特征，我们定义一个特征函数：

$$f(x, y) = \begin{cases} 1 & \text{满足特征条件} \\ 0 & \text{不满足特征条件} \end{cases}$$

特征函数在满足成为一个特征时取值 1，否则取值 0。则特征函数的经验概率为：

$$\tilde{p}(f) \equiv \sum_{x, y} \tilde{p}(x, y) f(x, y)$$

在最大熵模型中，特征函数的期望概率为：

$$p(f) \equiv \sum_{x, y} \tilde{p}(x) p(y|x) f(x, y)$$

最大熵模型的训练，就是满足这样一个约束条件——模型的期望概率与训练集中得到的经验概率相等：

$$p(f) = \tilde{p}(f)$$

所有这样的约束组成一个集合 C 。

2.3.2 对未知条件尽可能均匀

现在，最大熵模型已经满足了所有约束条件了。那对于未知条件，我们要尽可能均匀 (uniform)。那怎样才能达到“尽可能均匀”呢？一个衡量均匀分布的数学度量为熵：

$$H(p) \equiv - \sum_{x, y} \tilde{p}(x) p(y|x) \log(p(y|x))$$

当熵最大时，则对位置条件就尽可能均匀了。

最大熵模型的训练，就是要找到一个模型 $p^* \in C$ ，并且具有最大熵 $H(p)$ 。即，

$$p^* = \operatorname{argmax}_{p \in C} H(p) = \operatorname{argmax}_{p \in C} \left(- \sum_{x, y} \tilde{p}(x) p(y|x) \log(p(y|x)) \right)$$

2.4 支持向量机

支持向量机的基本思想是：求解能够正确划分训练数据集并且几何间隔最大的分离超平面。对于线性可分的训练数据集，线性可分离超平面有无穷多个（等价于感知机），但是几何间隔最大的分离超平面却只有一个。间隔最大的直观解释是：对训练集找到几何间隔最大的超平面意味着以充分大的确信度对训练数据集进行分类。¹⁹ 以下分别介绍了从数学上如何

1)、划分数据集，2)、最大化间隔。

2.4.1 划分数据集

一个线性分类器的学习目标就是要在 n 维的数据空间中找到一个分类超平面，其方程可以表示为：

$$W^T x + b = 0$$

于是，我们得到一个分类函数：

$$f(x) = W^T x + b$$

为了便于计算距离，我们将分类标签设为 1 和 -1。

2.4.2 最大化间隔

这个分类超平面要满足这样的条件：找到离分割超平面最近的点，确保它们离分割超平面的距离尽可能远。¹⁸

令 d 为离分割超平面最近的点 x 到分类超平面的距离：

$$d = \frac{W^T x + b}{\|W\|} = \frac{f(x)}{\|W\|}$$

d 小于 0 表示点在超平面负侧，标签 $y = -1$ ； d 大于 0 表示点在超平面正侧，标签 $y = 1$ 。而我们需要不是 d ，而是 d 的绝对值，因此乘上对应的类别 y 得：

$$\tilde{d} = y * d = \frac{f(x)}{\|W\|}$$

我们的目标便是： $\max \tilde{d}$ 。方便推导和优化的目的，我们可以令最近的那些点的 $f(x) = 1$ 。于是上述目标函数 $\max \tilde{d}$ 转化为：

$$\max \frac{1}{\|W\|}, s.t. y_i * (W^T + b) \geq 1, i = 1, \dots, n$$

这个问题等价于：

$$\min \frac{1}{2} \|W\|^2, s.t. y_i * (W^T + b) \geq 1, i = 1, \dots, n$$

2.5 实验结果

对于外卖领域的文本评论，本文采用正负各 3500 条作为训练数据，正负各 1000 条作为测试数据，特征数为 4000。其中基于支持向量机的情感分析方法利用了 Python 机器学习包 `scikit-learn` 中的 SVC。最终实验数据统计如下：

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
K-NN (k=1)	99.84	60.8	75.57	71.82	99.9	83.56	80.35
K-NN (k=3)	93.02	58.6	71.9	69.78	95.6	80.68	77.1
Bayes	91.11	69.7	78.98	75.47	93.2	83.4	81.45
MaxEnt	91.91	80.7	85.94	82.80	92.9	87.56	86.8

SVM	91.92	81.9	86.62	83.68	92.8	88.0	87.35
-----	-------	-------------	--------------	--------------	------	-------------	--------------

对于标准的康奈尔影评数据集，本文采用正负各前 500 篇作为训练数据，正负各 200 篇作为测试数据，特征数为 4000。其中基于支持向量机的情感分析方法利用了 Python 机器学习包 scikit-learn 中的 SVC。实验数据统计如下：

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
k-NN (k=1)	100	65.5	79.15	74.35	100	85.29	82.75
k-NN (k=3)	77.17	71	73.96	73.15	79	75.96	75
Bayes	99.24	65	78.55	73.98	99.5	84.86	82.25
MaxEnt	90.5	82	86.09	83.56	91.5	87.35	86.75
SVM	99.34	75.5	85.80	80.24	99.5	88.84	87.5

3 多分类器融合的情感分析

3.1 实验观察

整体上看，最大熵、SVM 的准确率最高。而为了更直观、详细地观察到每一个分类器对每一个测试数据的分类结果，本文做了如下统计工作：

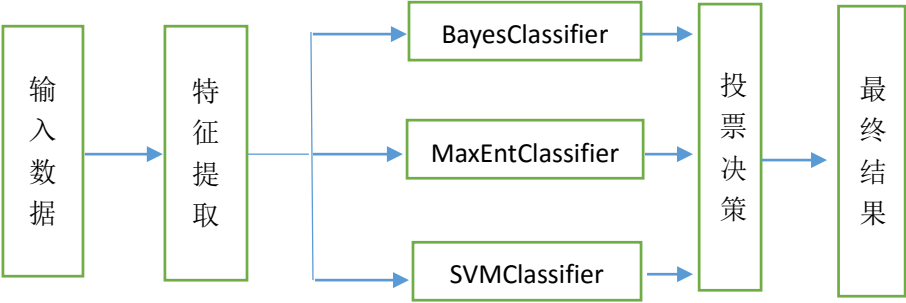
	数据 2	数据 11	数据 202	数据 488	数据 506	数据 510	数据 1019	数据 1022	数据 1029	数据 1099	数据 1113	数据 1203
Origin	1	1	1	1	1	1	0	0	0	0	0	0
K-NN	1	0	1	0	1	0	0	0	0	0	0	0
Bayes	0	0	1	1	0	1	0	0	1	1	0	0
MaxEnt	1	1	1	0	0	1	1	1	0	1	1	0
SVM	1	1	0	1	0	1	1	0	0	1	0	1

本文选取了部分外卖数据的分类结果用于展示。“Origin”一行表示测试数据原始标签，“k-NN”一行表示基于 K-近邻算法的情感分析结果，“Bayes”、“MaxEnt”、“SVM”三行如是。可以发现，对于第 2、1029 条数据，Bayes 分类错误，但另外三个分类器分类正确；对于第 202、1230 条数据，SVM 分类错误，另外三个分类器分类正确；对于第 1022 条数据，最大熵分类错误，但是其他三个分类器又分类正确。于是，本文提出多分类器并联组合的投票决策策略。哪个分类票数多，则结果就属于那个分类。

3.2 并联组合：投票决策

但投票决策可能会出现一些问题，如第 11、488 条数据等，正负分类结果各有两个分类器，投票将得不出结果。因此，应该选用奇数个分类器用于投票决策。最大熵和 SVM 性能最好，入围理所当然。在 Bayes 和 K-NN 中，K-NN 内存占用大，计算复杂度高，空间复杂度也高，所以，本文淘汰 k-NN，选用 Bayes+最大熵+SVM 进行并联组合。其并联组合流

程为：



若 C_i 为分类为 i 的分类器个数, 则多分类器并联组合的投票决策策略下的情感分类为:

$$i = \operatorname{argmax}(C_i).$$

对于外卖领域的文本评论, 本文采用正负各 3500 条作为训练数据, 正负各 1000 条作为测试数据, 特征数为 4000。最后实验数据如下:

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
Fusion1	92.54	80.6	86.16	82.82	93.5	87.83	87.05
Bayes	91.11	69.7	78.98	75.47	93.2	83.4	81.45
MaxEnt	91.91	80.7	85.94	82.80	92.9	87.56	86.8
SVM	91.92	81.9	86.62	83.68	92.8	88.0	87.35

对于标准的康奈尔影评数据集, 本文采用正负各前 500 篇作为训练数据, 正负各 200 篇作为测试数据, 特征数为 4000。最后实验结果如下:

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
Fusion1	99.39	81	89.26	83.97	99.5	91.08	90.25
Bayes	99.24	65	78.55	73.98	99.5	84.86	82.25
MaxEnt	90.5	82	86.09	83.56	91.5	87.35	86.75
SVM	99.34	75.5	85.80	80.24	99.5	88.84	87.5

多分类器并联组合的投票决策在外卖数据集上无效, 而在康奈尔影评数据集上性能却提升显著。要寻找原因, 只能重新回到细节微观上看每个分类器对每个数据的分类效果及投票决策效果。以下是外卖数据集上摘取的部分分类结果:

	数据 2	数据 9	数据 32	数据 39	数据 53	数据 417	数据 583	数据 598
Origin	1	1	1	1	1	1	1	1
Bayes	0	0	1	0	0	0	1	1
MaxEnt	1	0	1	1	0	1	0	1
SVM	1	0	0	0	1	0	0	0
Fusion1	1	0	1	0	0	0	0	1

可以看出, 对于第 2、32、598 条数据而言, “3 局 2 胜” 的投票决策的结果是正确的, 而对于第 39、53、417、583 条数据, “3 局 2 胜” 的投票决策的结果却是错误的。于是, 本文做了一些反思。

首先, 现实生活中的确有很多投票决策的实例, 但是, 投票决策是建立在投的每一票都是等价、公平的基础之上的。若每个分类器对于正负分类的投票质量都是一致, 则少数服从多数理所当然, 但是实际情况是, 每个分类器判断一个句子为 “正” 或 “负” 时, 准确率是不同的, 甚至是有显著差异。如, 由上述外卖语料的实验结果可得, SVM 投 “正” 票时, 其

准确率为 91.92%，而投“负”票时，准确率却只有 83.68%。于是，本文尝试提出多分类器并联组合的置信平均策略。

3.3 并联组合：置信平均

本文尝试引入置信度 C 的概念。置信度 C_{i-j} 表示分类器 j 对于输入数据分类为 i 的置信度。对于外卖语料而言，三种分类器中，最大熵判断一条语句为正向时，其正确的可能性为 91.91%，所以其正向置信度 C 为 0.9191；判断一条语句为负向时，其正确的可能性为 82.80%，所以负向置信度 C 为 0.828。SVM 判断一条语句为正向时，其正确的可能性为 91.92%，所以其正向置信度 C 为 0.9154；判断为负向时，其正确的可能性为 83.68%，所以其负向置信度 C 为 0.8368。贝叶斯分类器也如是。另外，令 N_i 表示分类为 i 的分类器数目。其中 i 取 0 和 1。则，多分类器并联融合下的情感分类为：

$$i = \begin{cases} \operatorname{argmax} \left(\frac{\sum_j C_{i-j}}{N_i} \right) & \text{若 } N_i \neq 0 \\ 0 & \text{若 } N_1 = 0 \\ 1 & \text{若 } N_0 = 0 \end{cases}$$

对于外卖领域的文本评论的实验数据如下：

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
Fusion2	89.03	87.7	88.36	87.88	89.2	88.54	88.45
Fusion1	92.54	80.6	86.16	82.82	93.5	87.83	87.05
Bayes	91.11	69.7	78.98	75.47	93.2	83.4	81.45
MaxEnt	91.91	80.7	85.94	82.80	92.9	87.56	86.8
SVM	91.92	81.9	86.62	83.68	92.8	88.0	87.35

对于标准的康奈尔影评数据集的实验结果如下：

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
Fusion2	91.22	93.5	92.35	93.33	91	92.15	92.25
Fusion1	99.39	81	89.26	83.97	99.5	91.08	90.25

实验结果表明，此方法有效。但为了进一步弄清其工作原理，我们需要再次在微观细节上观察每个分类器对每个数据的分类效果及投票决策效果。以下是外卖数据集上摘取的部分分类结果：

	数据 2	数据 32	数据 39	数据 53	数据 417	数据 583	数据 1029	数据 1037	数据 1054	数据 1456
Origin	1	1	1	1	1	1	0	0	0	0
Bayes	0	1	0	0	0	1	1	1	1	1
MaxEnt	1	1	1	0	1	0	0	1	0	0
SVM	1	0	0	1	0	0	0	0	0	0
Fusion1	1	1	0	0	0	0	0	1	0	0
Fusion2	1	1	1	1	1	1	1	1	1	1

可以看出，并联组合投票决策下判断正确的第 2、32 条数据，在并联组合置信平均策略下也判断正确；而对于并联组合投票决策下判断错误的第 39、53、417、583 条数据，即便 3 票内有 2 票投票错误，并联组合置信平均策略也可以判断出其正确类别。但是，对于数据

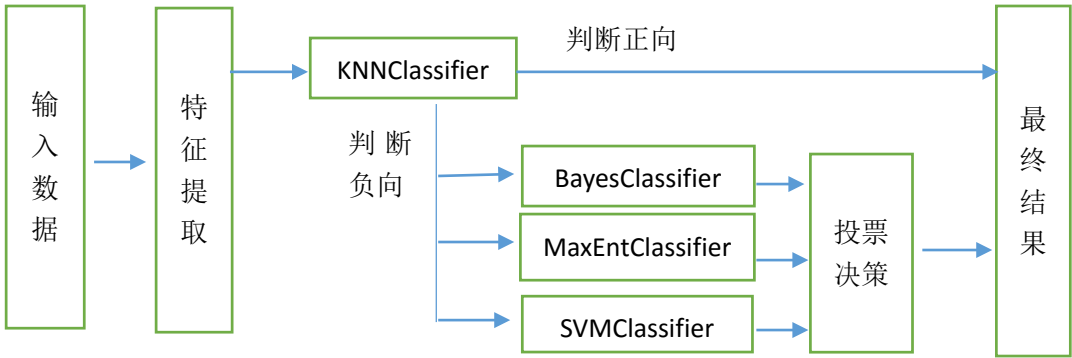
1029、1037、1054、1456，原本“3局2胜”投票策略分类正确，然而在并联组合置信平均策略下，却判断错误。进一步反思，可以发现，其实每个分类器的正向置信度都高于负向置信度，所以，3个分类器中只要有1个分类器投票为“正”，则整个语句分类结果必为正向。也就是说，置信平均会抛弃投票决策的优势，而不是改进投票决策。那么，怎样才能兼得置信平均与投票决策各自的优势呢？于是，本文尝试引入多分类器并联组合下的置信平均+投票策略。

3.4 串并联组合：投票决策

由5.3节的分析可得，置信平均策略无法正确分类的数据大部分都是负向评论，因为每个分类器的正向置信度都高于负向置信度，一个正向评论只要3个分类器中有1个分类器投票为“正”，则整个语句分类就可投票正确，分类为正向。则，我们的关注点落在负向分类。若我们能得到一个高负向准确率的分​​类器，则可以先用此分类器进行投票，若此分类器投票为负向，则直接判断该语句为负向；若此分类器投票为正向，则利用8.3节多分类器并联组合下的置信平均策略进行分类。也就是，将高负向准确率的分​​类器与并联组合置信平均分​​类器进行串联组合，即串并联组合下执行置信平均策略。

然而，本文所有分类器都不具备高负向准确率，却有一个k-NN分类器具备接近100%的高正向准确率。但是一个高正向准确率的分​​类器与置信平均策略的作用其实是相同的，而不是互补。置信平均策略的本质，是在“2票负向1票正向”的投票情况下，将错误的负向投票决策结果，利用置信度，更正为正向投票结果；然而，若数据原本分类的确是负向，则置信投票决策就错误校正了正确的“3局2胜”投票策略的分类结果。

因此，为了充分利用高正向准确率分​​类器，本文尝试将正向判断良好工作的高正向准确率的分​​类器与负向判断能良好工作的并联组合投票决策进行串联，优势互补。于是，此种多分类器混合组合策略流程如下：



对于外卖领域的文本评论，最后实验数据如下：

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
Fusion4	92.65	84.5	88.39	85.75	93.3	89.37	88.9
Fusion3	92.77	82.1	87.11	83.95	93.6	88.5	87.85
Fusion2	89.03	87.7	88.36	87.88	89.2	88.54	88.45
Fusion1	92.54	80.6	86.16	82.82	93.5	87.83	87.05

对于标准的康奈尔影评数据集，最后实验结果如下：

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	

Fusion4	99.44	88.5	93.65	89.64	99.5	94.31	94
Fusion3	99.42	85.5	91.94	87.28	99.5	92.99	92.5
Fusion2	91.22	93.5	92.35	93.33	91	92.15	92.25
Fusion1	99.39	81	89.26	83.97	99.5	91.08	90.25

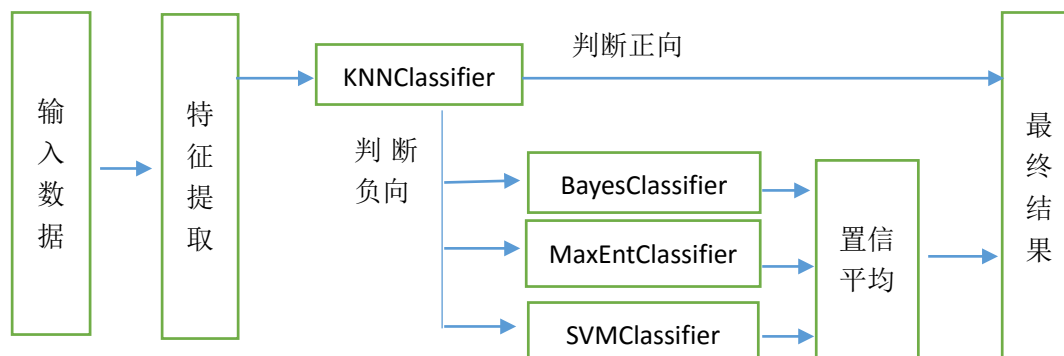
实验结果表明，此方法有效，性能提升显著。但为了弄清此方法是否是在保证并联组合的分类效果的基础上提升了分类精度，我们还需要再次在微观细节上观察每个分类器对每个数据的分类效果及投票决策效果。以下是 5.3 节摘取的同样的外卖数据的分类结果，并添加上 k-NN 的分类结果，汇总如下：

	数据 2	数据 32	数据 39	数据 53	数据 417	数据 583	数据 1029	数据 1037	数据 1054	数据 1456
Origin	1	1	1	1	1	1	0	0	0	0
k-NN	1	1	1	0	0	1	0	0	0	0
Bayes	0	1	0	0	0	1	1	1	1	1
MaxEnt	1	1	1	0	1	0	0	1	0	0
SVM	1	0	0	1	0	0	0	0	0	0
Fusion2	1	1	1	1	1	1	1	1	1	1
Fusion4	1	1	1	0	0	1	0	1	0	0

可以看出，并联组合置信平均策略下错误分类的第 1029、1054、1456 条数据，在串并联投票决策下被分类正确；在并联组合置信平均策略下正确分类的第 2、32、39 条数据，在串并联投票决策下也能被正确分类。然而，第 53 条数据却被错误分类，这也说明此种组合策略的局限性，丧失了 5.3 节置信平均策略的优势。

3.5 串并联组合：置信平均

k-NN 分类器虽然具有高正向准确率，但是其正向召回率却很低，只有 60.8%。也就是说，虽然 k-NN 判断为正向的结果都是正确的，但是接近 40% 本应该判断为正向的语句，却被判断为负向。由 5.5 节的分析可得，多分类器并联组合的置信平均策略也对正向评论具有良好的判断能力。因此，本文尝试将 k-NN 判断为负向的语句利用置信平均策略进行再次判断，以期弥补 k-NN 低正向召回率的缺陷。即串并联的置信平均策略，此种多分类器混合组合策略流程如下：



对于外卖领域的文本评论，最后实验数据如下：

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
Fusion5	89.17	89.7	89.43	89.64	86.64	89.37	89.4

Fusion4	92.65	84.5	88.39	85.75	93.3	89.37	88.9
Fusion3	92.77	82.1	87.11	83.95	93.6	88.5	87.85
Fusion2	89.03	87.7	88.36	87.88	89.2	88.54	88.45
Fusion1	92.54	80.6	86.16	82.82	93.5	87.83	87.05

对于标准的康奈尔影评数据集，最后实验结果如下：

Method	Positive			Negative			total
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	
Fusion5	91.47	96.5	93.92	96.30	91	93.57	93.75
Fusion4	99.44	88.5	93.65	89.64	99.5	94.31	94
Fusion3	99.42	85.5	91.94	87.28	99.5	92.99	92.5
Fusion2	91.22	93.5	92.35	93.33	91	92.15	92.25
Fusion1	99.39	81	89.26	83.97	99.5	91.08	90.25

实验结果表明，此种组合策略也是有效。只是随着实验数据的不同，其性能的提升效果不尽相同。

3.6 结论

本文从外卖数据分类结果的观察出发，得出不同分类器对同一数据有不同的分类效果，单个分类器都有各自的缺点，于是尝试将多个单分类器组合起来以期克服其缺点。最后发现，组合方式的不同、同一组合方式下数据的不同、同一组合方式下决策方式的不同，都对分类效果有影响。最佳的多分类器组合方式应该是将投票决策与置信平均决策两者的优势互补，找到或者构建一个高负向准确率的分类器，与并联组合方式进行串联。在决策时，首先进行负向投票决策，然后再进行正向的置信平均决策。本文最终提出的串并联投票决策和串并联置信平均策略，在康奈尔标准影评数据集上，都有良好表现，准确率比最好的单个分类器能提升 6%以上。

4 参考文献

- ¹ Pang, Bo, and Lillian Lee. 2008. "Opinion Mining and Sentiment Analysis." Foundations and Trends® in Information Retrieval 2: 1–135.
- ² K. Vivekanandan, J. Soonu Aravindan. 2014. Aspect-based Opinion Mining: A Survey. International Journal of Computer Applications (0975 – 8887) Volume 106 – No.3, November 2014
- ³ Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. "Thumbs up? Sentiment Classification Using Machine Learning Techniques." In Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), 79–86.
- ⁴ Turney, Peter D. 2001. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02 417- 424.
- ⁵ Yu, Lei, Jia Ma, Seiji Tsuchiya, and Fuji Ren. 2008. "Opinion Mining: A Study on Semantic Orientation Analysis for Online Document." In Proceedings of the World Congress on Intelligent Control and Automation (WCICA), 4548–52.
- ⁶ Liu, Jingjing et al. 2007. "Low-Quality Product Review Detection in Opinion Summarization." In Computational Linguistics, 334–42.

-
- ⁷ Lim, Ee-Peng et al. 2010. "Detecting Product Review Spammers Using Rating Behaviors." Proceedings of the 19th ACM International Conference on Information and Knowledge Management: 939–48.
- ⁸ Wiebe, Janyce et al. 2004. "Learning Subjective Language." Computational Linguistics 30: 277–308.
- ⁹ 熊德兰, 程菊明, 田胜利. 基于 HowNet 的句子褒贬倾向性研究. 计算机工程与应用, 2008, 44(22): 143—144.
- ¹⁰ Huang, Shen et al. 2009. "Improving Product Review Search Experiences on General Search Engines." In Proceedings of the 11th International Conference on Electronic Commerce - ICEC '09, 107.
- ¹¹ Moghaddam, Samaneh, and Martin Ester. 2011. "AQA: Aspect-Based Opinion Question Answering." In Proceedings - IEEE International Conference on Data Mining, ICDM, 89–96.
- ¹² Hu, Mingqing, and Bing Liu. 2004. "Mining and Summarizing Customer Reviews." Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining KDD 04 04: 168.
- ¹³ Lu, Yue, ChengXiang Zhai, and Neel Sundaresan. 2009. "Rated Aspect Summarization of Short Comments." Proceedings of the 18th international conference on World wide web - WWW '09: 131.
- ¹⁴ Wong, Tak-Lam, Wai Lam, and Tik-Shun Wong. 2008. "An Unsupervised Framework for Extracting and Normalizing Product Attributes from Multiple Web Sites." Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval SIGIR 08: 35.
- ¹⁵ Liu, Bing, and South Morgan Street. 2005. "Opinion Observer : Analyzing and Comparing Opinions on the Web." In Proceedings of the 14th International Conference on World Wide Web, 342–51.
- ¹⁶ 统计学习方法, 李航, 清华大学出版社, 2012.3
- ¹⁷ Harry Zhang. The Optimality of Naive Bayes. 2004, American Association for Artificial Intelligence
- ¹⁸ Peter Harrington. Machine Learning In Action. 2013. Manning Publications