

基于混淆交叉支撑向量机树的 自动面部表情分类方法

徐琴珍¹⁾ 章品正²⁾ 裴文江¹⁾ 杨绿溪¹⁾ 何振亚¹⁾

¹⁾(东南大学信息科学与工程学院, 南京 210096) ²⁾(东南大学计算机科学与工程学院, 南京 210096)

摘要 面部表情自动分类是情感信息处理研究中的重要内容,为了提高表情识别的准确率以及鲁棒性,提出了一种基于混淆交叉支撑向量机树的面部表情自动分类方法。该方法依据伪 Zernike 矩特征,以混淆交叉支撑向量机树对矩特征进行学习,实现面部表情的自动分类。混淆交叉支撑向量机树的结构使模型能够根据教师信号将面部表情识别问题分解,在不同的层次上以相对较低的复杂度解决子问题;在训练阶段,对当前中间节点划分的两个子样本集进行混淆交叉,增强了模型在面部表情识别上的整体泛化性能以及鲁棒性。实验对 Cohn-Kanade 面部表情数据库中的 6 类基本表情进行自动分类,准确率达到 96.31%;与同样基于该数据库的识别方法相比,该方法在识别正确率和鲁棒性上具有较大的优势。

关键词 面部表情自动识别 混淆交叉 支撑向量机树 伪 Zernike 矩

中图法分类号:TP391 文献标识码:A 文章编号:1006-8961(2008)07-1329-06

An Automatic Facial Expression Recognition Approach Based on Confusion-crossed Support Vector Machine Tree

XU Qin-zhen¹⁾, ZHANG Pin-zheng²⁾, PEI Wen-jiang¹⁾, YANG Lü-xi¹⁾, HE Zhen-ya¹⁾

¹⁾(Southeast University, School of Information Science and Engineering, Nanjing 210096)

²⁾(Southeast University, School of Computer Science and Engineering, Nanjing 210096)

Abstract Automatic facial expression recognition is the kernel part of emotional information processing. This study is dedicated to develop an automatic facial expression recognition approach based on confusion-crossed support vector machine tree (CSVMT) to improve recognition accuracy and robustness. Pseudo-Zernike moment features were extracted to train a CSVMT for automatic recognition. The structure of CSVMT enables the model to divide the facial recognition problem into sub-problems according to the teacher signals, so that it can solve the sub-problems in decreased complexity in different tree levels. In the training phase, those sub-samples assigned to two internal sibling nodes perform decreasing confusion cross, thus, the generalization ability of CSVMT for recognition of facial expression is enhanced. The experiments are conducted on Cohn-Kanade facial expression database. Competitive recognition accuracy 96.31% is achieved. The compared results on Cohn-Kanade facial expression database also show that the proposed approach appeared higher recognition accuracy and robustness than other approaches.

Keywords automatic facial recognition, confusion cross, support vector machine tree, Pseudo-Zernike moment

基金项目:国家自然科学基金项目(60702029,60672093,60672095);江苏省自然科学基金项目(BK2006061)

收稿日期:2006-08-20; **改回日期:**2007-03-21

第一作者简介:徐琴珍(1977~),女。讲师。现为东南大学信息科学与工程学院信号处理专业博士研究生。研究方向为图像处理、混合学习模型。E-mail:summer@seu.edu.cn

1 引言

人类的面部表情包含了个人情感和精神状态方面的重要信息,随着计算机技术与模式识别技术的发展,面部表情自动识别在人机交互、情感分析、心理学研究、电视电影特技制作、公安刑侦等领域受到了广泛的关注,有广阔的应用前景^[1]。Ekman 和 Friesen 定义了 6 大类基本面部表情:吃惊、悲伤、恐惧、愤怒、厌恶和快乐(如图 1 所示),Mase 和 Pentland 在此基础上展开了早期面部表情自动识别研究^[2]。后期的工程研究基本延用了这个心理学假设,以该 6 类表情为基础展开的面部表情自动识别工作包括:Littlewort 等人运用 AdaBoost 进行特征

提取,采用多种组合二分类 SVM 的方法对 6 类基本表情进行分类,最好的识别效果是 92.90%^[3];Aleksic 和 Katsaggelos 采用多重流隐马尔可夫(MS-HMM)方法实现表情分类^[1],正确率达到 93.66%;Chibelushi 和 Bourel 采用 Kanade-Tucas-Tomasi 特征点跟踪算法计算特征点间的距离与夹角进行特征抽取,而后实现基于模糊集的分类方法^[4],识别正确率为 79.00%;Cohen 等人采用非刚体模型进行运动矢量的抽取并以此构成用于表情分析的特征向量,并且分别通过 HMM 或学习 Bayesian 网进行表情识别,识别正确率分别为 74.00% 和 81.00%^[5]。这些面部表情识别的方法获得了一定的成功,并给后来的研究者很多有益的启发,同时,在识别的准确率和算法的鲁棒性上还有待进一步增强。



图 1 6 种基本面部表情示例

Fig. 1 Examples of the six basic facial expressions

近年来,支撑向量机作为机器学习一个新兴的重要分支被广泛地应用于模式识别和函数拟合问题中,其优越性在理论和实验方面都得到了深入的研究和验证,在面部表情自动识别的应用中获得了优良的识别效果^[3]。本文提出的混淆交叉支撑向量机树(CSVMT)结合了树型结构包含的分而治之地处理问题的思想与支撑向量机的优良泛化性能^[6]。在训练阶段,对 CSVMT 当前中间节点划分的两个子样本集进行混淆交叉,使对局部决策有较大影响

的样例参与更深层次的分类型超曲面的构建,进一步增强了模型的整体泛化性能和鲁棒性,降低了支撑向量机参数对模型泛化性能的影响,对面面部表情识别有着独特的优势和应用潜力。本文主要研究 6 类基本面部表情的自动分类,结合基于伪 Zernike 矩的面部表情特征抽取,针对提高面部表情识别的准确率以及在非均衡样本集上面面部表情识别的鲁棒性问题,提出了一种基于 CSVMT 的面部表情的自动分类方法。

2 特征抽取

为了提高面部表情自动识别系统的性能,研究人员在形状特征提取和图像灰度特征提取两个方面进行了积极的探索。前者主要通过面部弹性模型来定位面部表情的特征并通过提取模型的变形信息得到表情特征,需要鲁棒性较强的脸部特征的定位与跟踪方法;后者针对图像灰度进行特征提取(包括纹理统计参数特征、变换系数特征等),它们相对容易获取并且具有很好的稳定性^[1]。基于矩方法的特征提取属于后者,它具有较好的抗噪声性能^[7]。在统计学中,矩用于表征随机变量的分布情况,可以通过将图像看作是一个 2 维或 3 维的密度分布函数,将矩引入图像分析领域。基于矩的图像分析就是借助对矩值的分析得到物体的特征表述。自从 20 世纪 60 年代矩方法应用到图像分析与模式识别以来,矩方法已成为图像处理和计算机视觉领域的一类重要方法。在研究中对几种基于矩的特征提取方法(Zernike 矩,伪 Zernike 矩,Fourier-Mellin 矩,Krawtchouk 矩和 Tchebichef 矩)进行了对比分析,发现伪 Zernike 矩特征在表情识别上具有较好的表现^[7],本文采用基于伪 Zernike 矩的特征抽取方法构建特征集。

伪 Zernike 矩是由一类定义在单位圆上的正交多项式计算得到的,将图像以外接圆方式映射到单位圆内,图像内的像素位置由极坐标 (ρ, θ) 表示,定义 2 维图像 $f(\rho, \theta)$ 的伪 Zernike 矩为^[8]

$$C_{nm} = \frac{n+1}{\pi} \int_0^1 \int_0^{2\pi} f(\rho, \theta) W_{nm}^*(\rho, \theta) \rho d\rho d\theta \quad (1)$$

式中, m, n 为伪 Zernike 多项式的阶数,满足 $|m| \leq n$ 。当 m 和 n 较小时,伪 Zernike 矩的值代表了图像的整体信息;当 m 和 n 较大时,矩值表示图像的细节信息。 $W_{nm}^*(\rho, \theta)$ 是伪 Zernike 基函数的共轭形式,该基函数的定义形式为

$$W_{nm}(\rho, \theta) = S_{nm}(\rho) e^{jm\theta} \quad (2)$$

式中, $S_{nm}(\rho)$ 为伪 Zernike 多项式,其定义如下:

$$S_{nm}(\rho) = \sum_{s=0}^{n-1-|m|} \frac{(-1)^s (2n-s+1)! \rho^{n-s}}{s! (n-|m|-s)! (n-|m|-s+1)!} \quad (3)$$

为了提高效率,在对伪 Zernike 矩进行离散化计算时,采用 Chong 等人设计的伪 Zernike 多项式的递推关系式快速计算各阶矩值^[8],本文抽取 $n \leq 11$ 情

况下的 66 个不同的伪 Zernike 矩值作为面部表情特征。

3 混淆交叉支撑向量机树

CSVMT 模型的构建出于两个因素考虑:首先,树型结构的性质使模型能够根据教师信号(学习集中每个样本对应的表情类别)将问题分解,并在不同的层次上以相对降低的复杂度解决子问题。CSVMT 模型的整体结构为每个中间节点上嵌入了 SVM 模块的二叉树,在每两个同属一个父节点的中间节点上进行按层次递减的子样本集混淆交叉。对于二分类问题,类别标签可以直接作为二叉树节点分裂时的教师信号。对于类别数大于 2 的多分类问题,需要为各中间节点上的样本子集构建教师信号,可以通过简化的启发式算法实现两分类到多分类的扩展^[6]。

根节点或上一层的中间节点的划分效果直接影响其左右子树上样例的分配,若划分曲面太复杂,容易导致过拟合。为降低这种风险,上层节点比其子树上的中间节点的交叉程度充分。随着树型层次的加深,混淆交叉因子的收敛性使样例交叉程度逐层递减。设 s_j 和 s_{j+1} 为父节点 s_p 的两个子节点, S_p 为分配到节点 s_p 上的训练样本子集。考虑输入模式 $\mathbf{x}$ 对于 s_p 节点上的支撑向量机模块的空间变量:

$$\gamma_p(\mathbf{x}) = \sum_{i=1}^l \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \quad (4)$$

式中, $\mathbf{x}_i$ 和 $\alpha_i, i=1, 2, \dots, l$ 分别为支撑向量和拉格朗日乘子, y_i 为 $\mathbf{x}_i$ 对应的类别, $\mathbf{x}$ 为当前样本点。以样例混淆交叉因子作为训练样本在子节点上的分配依据,取代简单的符号函数操作 $\text{sgn}(\gamma)$, 定义样例混淆交叉因子为

$$C_{\gamma_p, m} = \rho_0 \exp(-\lambda \cdot m) \bar{\gamma}_p \quad (5)$$

式中, m 为中间节点 s_j 和 s_{j+1} 所在的层数, $\rho_0 \in (0, 1)$ 为初始混淆交叉因子,

$$\bar{\gamma}_p = \left(\frac{1}{|S_p|} \right) \sum_{x_i \in S_p} |\gamma_p(x_i)|$$

λ 控制交叉因子随着树的层数增加而收敛的速度。混淆交叉的过程也就是以混淆因子为依据,对子节点上训练样例进行分配的过程,即 s_j 和 s_{j+1} 节点上的样例分配分别为

$$S_j = \{ \mathbf{x} | \mathbf{x} \in S_p, \gamma_p(\mathbf{x}) \geq -C_{\gamma_p, m} \} \quad (6)$$

$$S_{j+1} = \{x | x \in S_p, \gamma_p(x) \leq C_{\gamma_p, m}\} \tag{7}$$

这样,靠近决策曲面从而易于被误分的混淆样例: $S_j \cap S_{j+1} = \{x | x \in S_p, |\gamma_p(x)| \leq C_{\gamma_p, m}\}$ 被同时保存在 s_p 的下层中间节点 s_j 和 s_{j+1} 的训练样例集中,使对决策曲面有较大影响的训练样例能够在树型结构更深层次的中间节点的训练过程中参与更精细的分类超曲面的构建,从而增强支撑向量机树的泛化性能和鲁棒性。

为了直观地说明 CSVMT 的泛化性能,在未对支撑向量机的两项参数进行优化的情况下,以常用于衡量机器学习模型性能的 2 维双螺旋复杂模式分类问题为例,将 CSVMT 与 SVM,SVMT(没有混淆交叉的支撑向量机树)构建的分类超曲面进行对比,如图 2 所示。单个的 SVM 尽管由于本身具有的较强的泛化性能而产生了一个相对简单的分类超曲面,但是由于两个参数是任取的并且没有经过优化,使得 SVM 在训练样本识别的准确率上显然要低于树型结构模型所得到的结果;对比树型结构模型的学习结果可知,两种模型在识别准确率上都达到了比较理想的结果,带混淆交叉的 CSVMT 与 SVMT 相比,能够在只增加两个简单的中间节点的情况下获得更接近于理想状态的分类超曲面,在模型整体的泛化性能上有所提高。

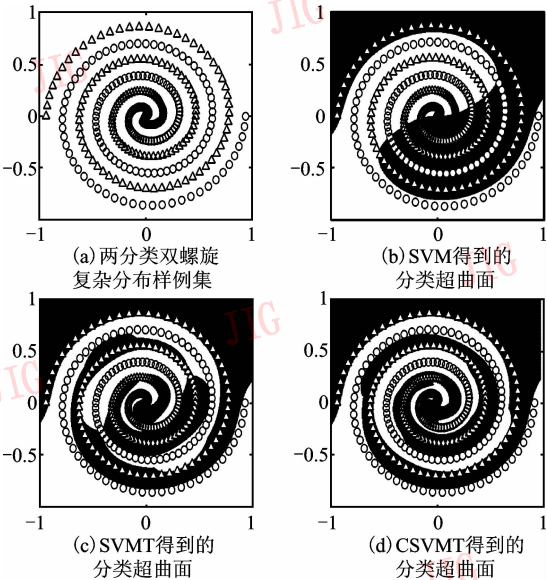


图 2 各类学习方法构建的分类超曲面对比结果
Fig. 2 The compared decision hyperplanes formed by different learning approaches

另外,支撑向量机的参数选择对模型的性能有很大的影响,其中最重要的一个参数为核函数。有

关核函数选择的理论依据还不够充分,近年来,在支撑向量机研究中,以高斯核、拉普拉斯核以及多项式核为最常用的几种核函数,其中高斯核是用于分类的 SVM 的首选^[9],这里在 CSVMT 模型中嵌入到中间节点上的支撑向量机采用高斯核函数:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right) \tag{8}$$

该支撑向量机的分类性能与核的宽度 σ 以及支撑向量机的惩罚项参数有密切的关系,参数值选择不当会导致过拟合或欠拟合。针对此情况的研究包括:基于规则的元学习方法实现支撑向量机核宽度的调整^[10];根据高斯核支撑向量机的对称性质提出启发式选择两个参数的策略^[11];采用交叉验证或构建一个验证子集,以在训练样本上预测的模型泛化性能为衡量准则,多次迭代实现支撑向量机两个参数值的调整^[12]。CSVMT 模型能够以树型结构上的混淆交叉降低参数的对模型性能的影响,在更宽泛的参数选择范围内获得良好的泛化性能鲁棒性,注意到该性能是在样本训练过程中实现的,无需额外的调整参数计算过程。同样以双螺旋复杂分布问题中支撑向量机的两项参数对模型分类性能的影响为例:在固定惩罚项参数 $C = 500$ 的情况下,SVM 高斯核宽度 $\sigma > 0.6$ 时,识别精确率出现明显下降,而 CSVMT 则在 $\sigma > 1.5$ 时模型识别性能才出现明显的影响,由此可知 CSVMT 模型的核宽度取值范围远大于单个支撑向量机,从而在一定程度上降低了核宽度对分类性能的影响;在固定核宽度 $\sigma = 0.5$ 情况下,惩罚项参数 $C < 2^{10}$ 对 SVM 模型识别性能有非常明显的影响,而对 CSVMT 模型的性能影响很小。由此可知 CSVMT 模型能够降低支撑向量机两项参数对支撑向量机的分类性能的影响。

可以将面部表情自动识别问题看成高维空间中的复杂模式分类问题,基于 CSVMT 的面部表情自动识别过程也就是在伪 Zernike 矩值构成的特征集上训练 CSVMT 模型,从而实现高维空间中复杂模式分类的过程。

4 面部表情自动识别实验及分析

实验采用 Cohn-Kanade 面部表情数据库^[13],该数据库中包含年龄在 18 岁至 30 岁的不同人种、不同国家的志愿者的 6 种基本表情数据,在表情分析研究中具有广泛的影响力,国内外研究人员已经在

这个数据库上进行了一些相关工作,心理学专家已对该数据库中的数据进行了人工判定,因而表情分类的结果较好地减轻了主观因素的影响。

基于伪 Zernike 矩的特征抽取方法构建的特征集中包含 1 427 个样例,每个样例具有 66 维特征,包含 6 类基本表情,其数目依 1 至 6 类标识次序分别为:惊奇(460)、快乐(464)、悲伤(156)、愤怒(127)、厌恶(126)和恐惧(94)。训练样例和测试样例各取总样本的 1/2,取 CSVMT 模型的初始混淆交叉因子 $\rho_0 = 0.8$,衰减因子 $\lambda = 0.3$,实验结果取 8 次运行结果的平均值。表 1 列出了采用伪 Zerniker 矩特征的各种不同识别方法的表情识别正确率和总体识别正确率(均指测试识别率)。其中包括:基于支撑向量机树(SVMT)的面部表情识别方法(未执行混淆交叉)^[14];基于 Hsu 等人研究对比的 3 种典型的多分类支撑向量机(DAGSVM,1-V-1SVM,1-V-RSVM)的面部表情识别方法^[15];基于 K 近邻法则(KNN)的面部表情识别方法;基于线性判决分析的面部表情识别方法。在实验中发现基于 DAGSVM,1-V-1SVM,1-V-RSVM,SVMT 和 CSVMT 的面部表情识别方法在训练识别率上均达到了 100%。从表 1 可以看出,在训练识别率相同的情况下,首先,基于 CSVMT 的面部表情识别方法在整体测试识别正确率上优于其他对比方法,一定程度上表明该方法具有相对较高的泛化性能;其次,在样例分布不均衡的情况下,基于 CSVMT 的面部表情识别方法对大样本表情(惊奇和快乐)的识别达到了较高的正确率(97.39%和 99.86%),值得注意的是本文的识别方法对小样本表情(悲伤、愤怒、厌恶和恐惧)的识别率明显优于其他 4 种对比方法,说明本文提出的识别方法在非均衡分布样本集上的具有较强的鲁棒性。

表 1 基于 CSVMT 的面部表情自动识别对比结果
Tab. 1 The facial expression automatic recognition results of CSVMT compared with other approaches

面部表情	方法			识别率			
	DAG SVM	1-V-1 SVM	1-V-R SVM	SVMT	LDA	KNN	CSVMT
惊奇	98.86	99.42	97.68	97.39	90.82	92.46	97.39
快乐	99.65	100.00	98.85	100.00	91.77	95.61	99.86
悲伤	90.00	88.32	64.10	90.17	75.06	72.27	92.31
愤怒	89.52	86.77	76.04	92.59	70.52	75.01	94.18
厌恶	84.44	87.89	75.26	83.60	66.11	60.99	92.06
恐惧	82.97	85.82	71.01	82.98	76.02	66.45	88.65
总体	95.01	95.36	88.76	94.86	84.45	85.21	96.31

另外,还将基于 CSVMT 的面部表情识别方法的实验结果与其他在 Cohn-Kanade 面部表情数据库上并且以 6 类基本表情为研究基础的方法进行对比。与引言中总结的 4 种识别方法的识别准确率(92.90%^[3]、93.66%^[1]、79.00%^[4]、81.00%^[5])的对比结果可知,结合伪 Zernike 矩特征抽取,基于 CSVMT 的面部表情自动识别方法在识别精确率上达到 96.31%,也具有明显的优势。

从以上实验结果对比中可知,本文方法在识别正确率上优于同类算法在 Cohn-Kanade 面部表情数据库上获得的结果;在泛化性能上与其他基于多分类 SVM 的几种经典识别方法相比有一定的优势;并且在非均衡分布样本集上具有相对较强的识别鲁棒性。这主要得益于嵌入 CSVMT 中间节点的支撑向量机本身的优良泛化性能,以及在树型结构的相邻子节点间的混淆交叉进一步增强了模型在面部表情识别上的整体泛化性能和鲁棒性。

5 结 论

提出了一种基于混淆交叉支撑向量机树的面部表情自动识别方法,即在伪 Zernike 矩面部表情特征抽取的基础上,以混淆交叉支撑向量机树对抽取的特征进行识别。混淆交叉支撑向量机树的结构使模型能够根据教师信号将面部表情识别问题分解,在不同的层次上以相对降低的复杂度解决子问题;树型结构的中间节点上嵌入了具有优良泛化性能的支撑向量机,在训练阶段,当前中间节点划分的两个子样本集间进行混淆交叉,使对局部决策有较大影响的样例参与更深层次的分类超曲面的构建,从而进一步增强了模型在面部表情识别上的整体泛化性能。实验结果表明,该方法在识别正确率、泛化性能以及非均衡分布样本集上的识别鲁棒性都具有一定的优势。

参考文献 (References)

1 Aleksic P S, Katsaggelos A K. Automatic facial expression recognition using facial animation parameters and multistream HMMs [J]. IEEE Transactions on Information Forensics and Security,2006, 1(1):3 ~ 11.

2 Mase K, Pentland A. Recognition of facial Expression from optical flow [J]. Institute of Electronics Information and Communication Engineers Transactions, 1991, E74(10): 3474 ~ 3483.

3 Littlewort G, Bartlett M, Fasel I, et al. Dynamics of facial expression

- extracted automatically from video [A]. In: Proceedings of 2004 IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Face Processing in Video [C]. Orlando, FL, 2004: 80 ~ 88.
- 4 Bourel F, Chibelushi C C, Low A A. Robust facial expression recognition using a state-based model of spatially-localised facial dynamics[A]. In: Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition[C], Washington, DC, USA, 2002: 106 ~ 111.
- 5 Cohen I, Sebe N, Chen L, *et al.* Facial expression recognition from video sequences: temporal and static modeling [J]. Computer Vision and Image Understanding, 2003, **91**(1): 160 ~ 187.
- 6 Xu Qin-zhen, Song Ai-guo, Pei Wen-jiang, *et al.* Tree-structured support vector machine with confusion cross for complex pattern recognition problems [A]. In: Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology [C], Suzhou, China, 2005: 195 ~ 198.
- 7 Zhang Pin-zheng. The Method and System Research on Automatic Facial Expression Classification[D]. Nanjing: Southeast University, 2005. [章品正. 面部表情自动分类的方法与系统研究[D], 南京: 东南大学, 2005.]
- 8 Chong C W, Raveendran P, Mukundan R. An efficient algorithm for fast computation of Pseudo-Zernike moments [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2003, **17**(6): 1011 ~ 1023.
- 9 Jeng J T. Hybrid approach of selecting hyperparameters of support vector machine for regression [J]. IEEE Transactions on Systems, Man and Cybernetics (Part B), 2006, **36**(3): 699 ~ 709.
- 10 Ali S, Smith K A. Kernel width selection for SVM classification: a meta-learning approach [J]. International Journal of Data Warehousing & Mining, 2005, **1**(4): 78 ~ 97.
- 11 Keerthi S S, Lin C J. Asymptotic behaviors of support vector machines with Gaussian kernel [J]. Neural Computation, 2003, **15**(7): 1667 ~ 1689.
- 12 Belousov A I, Verzakov S A, Von F J. A flexible classification approach with optimal generalization performance: Support vector machines [J]. Chemometrics and Intelligent Laboratory Systems, 2002, **64**(1): 15 ~ 25.
- 13 Kanade T, Cohn J F, Tian Y I. Comprehensive database for facial expression analysis [A]. In: Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition [C], Grenoble, France, 2000: 46 ~ 53.
- 14 Pang S P, Kim D J, Bang S Y. Face membership authentication using SVM classification tree generated by membership based LLE data partition [J]. IEEE Transactions on Neural Networks, 2005, **16**(2): 436 ~ 446.
- 15 Hsu C W, Lin C J. A comparison of methods for multiclass support vector machines [J]. IEEE Transactions on Neural Networks, 2002, **13**(2): 415 ~ 525.