·研究报告 ·

基于组合相似度丹参样本质量稳定性评价

詹雪艳1,史新元1,孙启生2,乔延江1

(1北京中医药大学,北京 100102:2天津天士力现代中药资源有限公司,天津 300402)

摘要:目的:基于组合相似度进行丹参样本质量的稳定性评价。方法:采集丹参水提液的色谱指纹图谱数据,建立组合相似度,基于样本间组合相似度、共有峰夹角余弦相似度和所有峰夹角余弦相似度对9个丹参样本进行系统聚类分析,对丹参样本质量的稳定性进行评价。结果:9个丹参样本的组合相似度与所有峰夹角余弦相似度的计算值接近,2种相似度方法的系统聚类结果与丹参色谱指纹图谱的主成分分析结果一致,结果表明,9个丹参样本较好地聚为2类。结论:组合相似度可用于9个丹参样本峰比例稳定性的初步评价,为原料药质量的稳定性评价提供一种相似度方法。

关键词:丹参;组合相似度;色谱指纹图谱

基金资助:国家科技重大专项(No.2010ZX09502-002),北京中医药大学自主课题(No.2009JYB22-JS030)

Quality stability evaluation of *Radix Salviae Miltiorrhizae* by applying combination similarity

ZHAN Xue-yan¹, SHI Xin-yuan¹, SUN Qi-sheng², QIAO Yan-jiang¹
(¹Beijing University of Chinese Medicine, Beijing 100102, China; ²Tianjin Tasly Modern TCM resources Co, Ltd, Tianjin 300402, China)

Abstract: Objective: To investigate the application of combination similarity to quality stability evaluation. Methods: The chromatographic fingerprints of *Radix Salviae Miltiorrhizae* were collected and combination similarity was established by the chromatographic peaks area of 9 samples of *Radix Salviae Miltiorrhizae*. The quality stability of 9 samples of *Radix Salviae Miltiorrhizae* were evaluated by the HCA (Hierarchical Cluster Analysis) based on combination similarity, cosine of common peaks and cosine of all peaks between every two samples. Results: The calculated results of combination similarity was close to cosine of the angle calculating all peaks area for all 9 samples of *Radix Salviae Miltiorrhizae*. The PCA (Principle Component Analysis) result of 9 samples confirmed the HCA result of 9 samples based on combination similarity and based on cosine of the angle calculating all peaks area, which 9 samples of *Radix Salviae Miltiorrhizae* were classified into 2 groups. Conclusion: Combination similarity could preliminary evaluate the peaks proportion stability of 9 samples of *Radix Salviae Miltiorrhizae* and becomes an alternative similarity measure to evaluate the quality stability of Chinese crude drug.

Key words: Radix Salviae Miltiorrhizae; Combination similarity; Chromatographic fingerprints

Fund assistance: National Science and Technology Major Project(No.2010ZX09502-002), Subjective Selected Subjects Program of Beijing University of Chinese Medicine(No.2009JYB22-JS030)

色谱指纹图谱相似度是评价中药质量稳定性的有效手段之一,其中,相似度计算方法是色谱指纹图谱相似度应用的关键技术。目前色谱指纹图谱相似度的计算方法主要有3类:数量相似度、夹角余弦和相关系数等定性相似度、距离相似度和程度相似度^[1]等定量相似度,后续的相似度方法研究中分别对3类相似度方法进行了改进,如改进Nei系数法^[2]、组合相似度^[3]、加权马氏距离^[4]、改进程度相似度^[5-7],以及提出总量统计矩相似度^[8]、相对熵奇异值法^[9]和指纹图谱差别矢量统计检验^[10]等新相似度方法。目前,向量夹角余弦相似度在药材质量稳定性评价中应用广泛,共有峰夹角余弦衡量物种稳定的成分间色谱峰峰比例的一致性,适合药材的品种鉴别,所有峰夹角余弦直

接计算共有峰和非共有峰的峰面积,衡量所有色谱峰对应的化学成分峰比例的稳定性,用于制剂中原料药成分间配比稳定性的初步评价。组合相似度将非共有峰峰面积以一定的权重纳入计算,并基于色谱指纹图谱峰面积数据确定非共有峰的权重,对夹角余弦进行了修正。本研究将组合相似度用于丹参色谱指纹图谱的相似性分析,基于丹参样本峰面积数据建立的组合相似度能更灵敏地反映样本内在成分的峰比例变化,适合衡量丹参原料药质量的稳定性。

材料

1. 仪器 Agilent-1100高效液相色谱仪(HP真空脱气泵、HP四元泵、HP自动进样器、HP柱温箱、HP二极管阵列检测器

通讯作者:乔延江,北京市望京中环南路6号北京中医药大学,邮编:100102,电话:010-84738619,E-mail:yjqiao@263.net

(DAD)。98-1-B型电子调温电热套(天津市泰斯特仪器有限公司)。

2. 试药 乙腈(色谱纯,Fisher公司),三氟乙酸(分析纯,Fisher公司),乙醇(分析纯)。对照品丹参素钠(批号:110855-200507)、原儿茶醛(批号:110810-200506)和丹酚酸B(批号:111562-200504)均购自中国药品生物制品鉴定所,纯度均大于98%。9批不同产地的丹参Radix Salviae Miltiorrhizae药材(样本号:s1-s9)由天津天士力现代中药资源有限公司孙启生工程师鉴定并提供。

方法

- 1. 样品溶液的制备 丹参药材50 下干燥4h,粉碎过筛 (60目)备用。精密称定丹参样品粉末2.5g,置于圆底烧瓶中,准确加入25mL去离子水,精密称定瓶子及药液总重并记录。待水完全浸没药粉,加热至沸腾,回流提取2h,冷却放置20min,再精密称重,补足损失的水分。0.45µm微孔滤膜滤过,滤液作为供试品溶液。
- 2. HPLC测试条件 色谱柱: Agilent TC- C_{18} (4.6mm×250mm,5 μ m)。流动相: 乙腈(A)-0.1%三氟乙酸水(B)体系。梯度洗脱程序见表1。流速:1.0mL/min。柱温:30。检测波长:270nm,进样量:10 μ L。

表1 丹参水提液的HPLC梯度洗脱条件

时间(min)	0.1%三氯乙酸(%,V/V)	乙腈(%,\//\)
0	95	5
21	70	30
45	78	22
50	0	100
55	95	5
60	95	5

- 3. 色谱指纹图谱的方法学考察 对丹参水提液色谱指纹图谱的测定方法进行稳定性、精密度和重复性试验。
- 4. 数据处理方法 将丹参水提液的色谱指纹图谱导入国家药典委员会《中药色谱指纹图谱相似度评价系统》2004A版,时间窗宽度设为0.1,以s9四川中江丹参样本色谱指纹图谱为参照谱,以丹参素钠、原儿茶醛、丹酚酸B为参照进行保留时间的多点校正,得到9批不同来源的丹参药材色谱指纹图谱的峰面积数据。

组合相似度Scos - P_{\sharp} ,以组合相似度的最大峰比例同态性值Hmax为目标,基于9个丹参样本色谱指纹图谱峰面积数据确定相似度公式中的参数值,建立组合相似度模型,组合相似度的建立参照文献[3]。用组合相似度衡量丹参样本色谱指纹图谱的相似程度,基于9个丹参样本的组合相似度进行系统聚类分析(HCA),对丹参样本水提液内在化学成分的稳定性进行初步评价。数据处理过程中相关计算程序自行编写,采用Matlab软件(Mathwork Inc.)计算,Originpro 7.5作图。

结果

- 1. 丹参水提液色谱指纹图谱数据的采集
- 1.1 色谱指纹图谱的方法学考察
- 1.1.1 稳定性试验:取s9四川中江的丹参药材样品,按方法 "1."项下的方法制备样品溶液,分别放置0、3、6、9、12、24h后,按方法"2."项中上述色谱条件进行分析,记录色谱图。结果表明丹参素、原儿茶醛和丹酚酸B的保留时间*RSD*均小于0.1%,峰面积的*RSD*为0.4%、0.7%、0.8%,供试品溶液在24h内稳定。
- 1.1.2 精密度试验:取s9四川中江的丹参药材样品,按方法"1."项下的方法制备样品溶液,在方法"2."项中的液相色谱条件下重复进样6次,记录色谱图。结果表明,丹参素、原儿茶醛和丹酚酸B的保留时间的RSD均小于0.3%,峰面积的RSD为0.3%、0.8%、0.8%,符合色谱指纹图谱的方法学要求,仪器精密度良好。
- 1.1.3 重复性试验:取s9四川中江的丹参药材样品6份,按方法"1."项下的方法制备样品溶液,按"2."项下液相色谱条件进行分析,记录色谱图。丹参素、原儿茶醛和丹酚酸B的保留时间RSD均小于0.3%,峰面积的RSD为1.4%、2.9%、3.2%,符合色谱指纹图谱的要求,方法重现性良好。
- 1.2 丹参色谱指纹图谱的峰面积数据 按照与样品相同的色谱条件采集丹参素钠、原儿茶醛和丹酚酸B 3个混合标准品溶液的色谱图。采用保留时间对照法和色谱峰紫外光谱比较法,确定了指纹图谱中的2号、3号和23号峰的对应成分分别是丹参素、原儿茶醛和丹酚酸B。

色谱指纹图谱的结果表明,9批丹参药材色谱指纹图谱均含有丹参素钠、原儿茶醛、丹酚酸B的色谱峰,所以用丹参素钠、原儿茶醛、丹酚酸B的色谱峰为参照对9批丹参药材色谱指纹图谱的保留时间进行多点校正,得到峰面积数据,见表2。9批丹参药材有16个共有峰和18个非共有峰。

2. 组合相似度模型的建立 按照方法"4."项下建立组合相似度的方法确定9个不同参照样本下组合相似度中的参数值。9个不同参照样本下组合相似度的峰比例同态性值H随 值增大迅速升高,缓慢下降。表3显示了9个丹参参照样本下组合相似度峰比例同态性值H得到优化的 区间,值的优化区间和分别是组合相似度比所有峰夹角余弦和共有峰夹角余弦和似度具有更高峰比例同态性值H所对应的 区间。9个参照样本下值的优化区间共同区间为[0.036,0.045],表明当0.036

0.045时,任何一个参照样本下建立的组合相似度比共有峰夹 角余弦相似度和所有峰夹角余弦相似度具有更高的峰比例同 态性值H,组合相似度比夹角余弦相似度对丹参样本的峰比例 变化响应更灵敏。

- 3. 丹参样本的质量稳定性分析
- 3.1 九个丹参样本两两间的相似度 计算9个丹参样本

不同来源的丹参水提液色谱指纹图谱的峰面积 表2 s1 s2 s3 s5 s7 s8 s9 0.0 521.2 0.0 0.0 0.0 598.5 0.0 0.0 0.0 1 1 479.7 1 293.3 1 330.0 1 679.6 1 780.0 1 257.0 1 395.2 1 656.9 1 663.1 2 1 847.3 1 130.9 2 967.4 3 2 035.4 2 482.5 2 516.8 6 91.6 2 300.1 1 641.1 57.7 0.0 80.9 141.0 390.8 0.0 127.0 95.4 0.0 4 101.3 87.8 135.9 437.3 304.9 147.2 176.2 166.6 75.3 5 949.1 1536.5 802.8 1484.4 856.9 406.1 1 772.8 1 556.3 864.3 6 388.0 371.2 397.6 336.0 544.1 7 0.0 531.5 343.3 400.4 269.6 372.7 227.9 393.3 1 012.9 431.5 376.7 310.8 317.7 8 460.9 482.5 317.1 624.2 931.8 485.1 596.5 402.0 602.9 9 283.3 358.6 538.2 266.1 6717 460.4 307.3 594.8 593.2 10 413.7 387.3 406.1 509.1 0.0 538.6 500.8 302.1 557.3 11 0.0 0.0 0.0 0.0 603.9 298.4 0.0 0.0 0.0 12 91.7 99.4 0.0 0.0 0.0 0.0 13 102.0 0.0 0.0 0.0 93.4 68.2 213.8 1 286.7 307.5 265.5 165.7 0.0 14 235.0 180.5 162.5 202.0 354.1 919.7 301.7 428.9 365.5 15 1 987.9 1 721.1 1 669.4 1 923.8 2 090.9 1 137.6 1 667.1 1 404.5 2 119.8 16 349.2 396.8 394.7 447.3 1 260.7 0.0 409.1 337.3 459.0 17 754.2 567.1 781.9 589.6 330.7 643.8 600.3 926.0 834.5 18 2 354.0 2 267.7 2 492.2 2 943.8 1 395.0 3 186.5 2 533.3 2 024.7 2 484.7 19 282.1 464.8 432.0 792.2 347.0 663.5 430.7 465.5 541.8 20 1 044.2 1 125.6 1 091.7 1 296.3 1 520.2 1 279.6 1 193.2 1 390.1 1 944.7 21 2 485.8 1 988.7 1 708.6 2 096.1 2 062.9 2 669.4 3 216.0 2 561.7 2 789.1 22 16 570.4 23 507.5 21 409.2 26 601.5 20 177.0 28 500.7 23 700.3 28 778.3 16 989.1 23 24 1 198.2 1 425.0 1 455.0 1 825.1 743.4 1 847.7 1 340.7 1 136.0 1 206.4 362.5 143.5 378.0 0.0 0.0 0.0 0.0 152.9 282.8 25 1659.8 1 347.9 1 857.0 2 142.6 2 369.0 2 012.6 0.0 0.0 1 824.0 26 0.0 0.0 0.0 2 641.4 1 859.2 3 936.4 1 822.0 2 147.3 0.0 27 1 691.9 2 207.0 0.0 307.7 0.0 28 0.0 0.0 0.0 0.0 1 326.7 0.0 0.0 251.5 408.4 603.9 435.5 166.1 1 225.8 29 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 581.0 30 2 667.6 389.4 0.0 0.0 0.0 0.0 0.0 0.0 2724.3 31 3 851.4 32 0.0 791.3 0.0 0.0 0.0 0.0 1 850.2 0.0 1 930.9 201.7 219.3 597.6 100.4 1 117.2 0.0 33 53.3 59.7

73、个四分令参照件本下组合作队反中,但时况化区1	表3	不同丹参参照样本下组合相似度中	值的优化区间
---------------------------	----	-----------------	--------

0.0

0.0

0.0

725.6

0.0

0.0

34

参照样本	优化区间	优化区间
s1	0.014 - 0.140	0.022-0.076
s2	0.020 - 0.185	0.023-0.161
s3	0.021-0.123	0.022 - 0.106
s4	0.003-0.116	0.017-0.045
s5	0.036 - 0.700	0.000 - 1.000
s6	0.005 - 0.263	0.008-0.067
s7	0.006 - 0.116	0.009-0.063
s8	0.004-0.200	0.000 - 1.000
s9	0.030 - 0.110	0.033-0.090

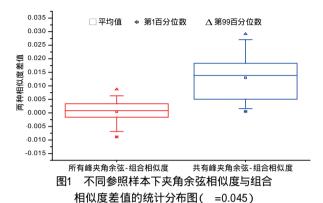
两两样本间共有峰夹角余弦相似度值、所有峰夹角余弦相似度值和 =0.045时的组合相似度值。对所有峰夹角余弦相似度与 =0.045组合相似度的差值和共有峰夹角余弦相似度与 =0.045组合相似度的差值进行统计分析,两组差值的统计分布图,见图1。图1显示所有峰夹角余弦与组合相似度的差值中位数0.001,分布范围为-0.008-0.008,差值的第25百分位数-第75百分位数集中分布区间为-0.002-0.003;36个共有峰夹角余弦与组合相似度差值的中位数为0.015,分布范围0.001-0.030,差值的第25百分位数-第75百分位数集中分布区间为0.005-0.020。以上结果表明,不同参照样本下共有峰夹角余弦相似度值与 =0.045组合相似度值相差较大,两者计

0.0

0.0

0.0

算值有显著性差异(P<0.001)。所有峰夹角余弦相似度值与 =0.045组合相似度值接近,组合相似度以0.045的权重计算非 共有峰的影响,相对于直接将非共有峰峰面积纳入计算的所有 峰夹角余弦相似度,计算值没有显著性差异。 =0.036组合相 似度计算值得到相似的结果, =0.036组合相似度与所有峰夹 角余弦计算值接近。



3.2 丹参样本的分类 当 =0.045时,基于样本间组合相似 度、共有峰夹角余弦和所有峰夹角余弦采用"最短距离法"的 类间距离进行系统聚类分析,9个丹参样本的系统聚类结果,见

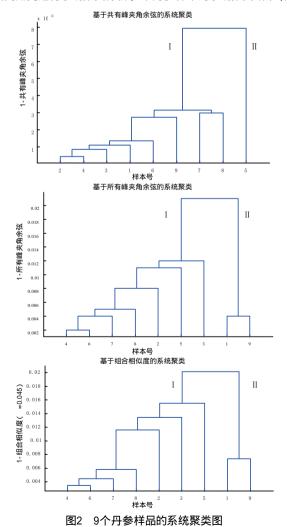
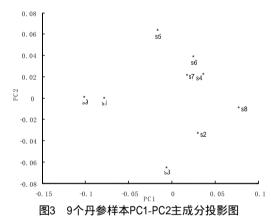


图2。共有峰夹角余弦相似度的系统聚类结果表明s5自成一类, 其它样本聚为一类,而且由纵坐标的距离知s5样本离其它样本 很远。所有峰夹角余弦相似度和组合相似度的系统聚类结果 一致,9个丹参样本很好地分为两类,s4、s6、s7、s8、s2、s3、s5 7个样本依次聚为一类,s1和s9聚为另一类。以上系统聚类结果 表明9个丹参样本的非共有峰对聚类有影响,将非共有峰的影 响纳入计算的组合相似度和所有峰夹角余弦相似度的聚类结 果与共有峰夹角余弦的聚类有很大的差别。

将9个丹参样本的峰面积数据除以各个样本总峰面积后 进行主成分分析,前两个主成分的累积方差贡献率为83.55%, 前3个主成分的累积方差贡献率为88.94%,对前两个主成分和 前3个主成分进行投影作图,结果见图3-图5,其主成分分析 (PCA)投影结果显示s4、s6和s7的距离很近,图2中所有峰夹 角余弦和组合相似度的系统聚类结果与其相符,共有峰夹角余 弦相似度的聚类结果显示以上3个样本距离较远,与9个丹参样 本PCA投影结果不相符。



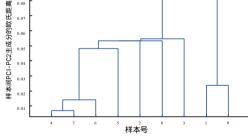
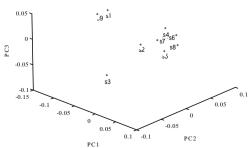


图4 9个丹参样本PC1-PC2主成分的系统聚类图



9个丹参样本PC1-PC2-PC3主成分投影图 图3中9个丹参样本PC1-PC2主成分的投影图表明s1和s9样

本投影在一起,其它的7个样本投影空间很近,可以看成一类。基于9个样本PC1-PC2主成分间欧氏距离采用"最短距离法"的类间距离进行系统聚类分析,聚类结果见图4。9个样本PC1-PC2主成分聚为两类,1和9为一类,其它7个聚为一大类,与图2中9个丹参样本的所有峰夹角余弦相似度和组合相似度的系统聚类结果一致。图5中9个丹参样品PC1-PC2-PC3主成分投影图和图6中9个样本PC1-PC2-PC3主成分间欧氏距离的系统聚类结果得到相同的结论,9个丹参样本的PC1-PC2-PC3主成分投影结果和聚类结果与图2中所有峰夹角余弦相似度、组合相似度的系统聚类结果一致,一定程度上证实了以上两种相似度聚类结果的合理性。综上分析,组合相似度和所有峰夹角余弦相似度的聚类结果以及PC1-PC2、PC1-PC2-PC3主成分投影结果均表明该9个丹参样本分两类,s1和s9样本一类,其它样本为第2大类,s4、s6和s7样本为第2类的中心,s2、s5、s8与s4、s6、s7 3个样本的距离较近,s3与s4、s6、s7 3个样本的距离较远。

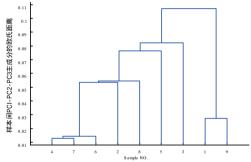


图6 9个丹参样本PC1-PC2-PC3主成分的系统聚类图 结论

将组合相似度用于9个丹参样本的相似性分析,基于样本色谱指纹图谱数据建立组合相似度模型,非共有峰权重的共同优化区间为[0.036,0.045]。当 =0.045时,两两样本间组合相似度计算值与共有峰夹角余弦相似度存在显著性差异(P<0.001),与所有峰夹角余弦相似度值接近,无显著性差异。组合相似度模型S_{cos} - P_#中,非共有峰的权重 值小,此时对于该组样本,可以用组合相似度和所有峰夹角余弦两种相似度的计算值对原料药质量的稳定性进行初步评价。因此,组合相似度可用于该9个丹参样本峰比例稳定性的初步评价,为原料药质量的稳定性评价提供一种相似度方法。

非共有峰的权重 值为0.045的组合相似度对丹参样本进行系统聚类,聚类结果与所有峰夹角余弦相似度的聚类结果相同,9个丹参样本很好地聚为两类。9个丹参样本二维和三维的PCA投影得到与系统聚类一致的分类,证明了组合相似度和所有峰夹角余弦相似度聚类结果的合理性。因此,该9个丹参样本分为两类,s1和s9样本一类,其它样本为第2类,其中,s4、s6和s7样本为第2类的中心,s2、s5、s8与s4、s6、s73个样本的距离较近,s3与s4、s6、s73个样本的距离较远。

参考文献

[1] 刘永锁,孟庆华,蒋淑敏,等.相似系统理论用于中药色谱指纹

图谱的相似度评价.色谱,2005,23(2):158-163

LIU Yong-suo,MENG Qing-hua,JIANG Shu-min,et al.Similarity system theory to evaluate similarity of chromatographic fingerprints of traditional Chinese medicine. Chinese Jounal of Chromatography, 2005,23(2):158-163

- [2] 孟庆华,刘永锁,王健松,等.色谱指纹图谱相似度的新算法及 其应用.中成药,2003,25(1):4-8
 - MENG Qing-hua,LIU Yong-suo,WANG Jian-song,et al.A new algorithm and application of similarity between chromatographic fingerprints.Chinese Traditional Patent Medicine,2003,25(1):4-8
- [3] 詹雪艳,史新元,段天璇,等.色谱指纹图谱组合相似度的算法, 色谱,2010,28(11):1071 - 1077
 - ZHAN Xue-yan,SHI Xin-yuan,DUAN Tian-xuan,et al.Combination similarity algorithm on chromatographic fingerprints. Chinese Jounal of Chromatography,2010,28(11):1071-1077
- [4] Xiang S M,Nie F P,Zhang C S.Learning a Mahalanobis distancemetric for data clustering and classification.Pattern Recognition,2008,41:3600
- [5] 刘永锁,曹敏,王义明,等.相似系统理论定量评价中药材色谱 指纹图谱的相似度.分析化学,2006,34(3):333-337 LIU Yong-suo,CAO Min,WANG Yi-ming,et al.Similarity system theory to evaluate the similarity of high performance liquid chromatography fingerprint of traditional Chinese medicine quantitatively.Chinese Jouannal of Analytical Chemistry,2006, 34(3):333-337
- [6] 詹雪艳,史新元,展晓日,等.基于相似系统理论的相似度计算方法的改进.分析化学,2010,38(2):253-257
 ZHAN Xue-yan,SHI Xin-yuan,ZHAN Xiao-ri,et al.Improvements in similarity algorithms based on similarity system theory.Chinese Jouarnal of Analytical Chemistry,2010,38(2):253-257
- [7] 谷瑞敏,郭治昕,刘巍巍,等.中药色谱指纹图谱相似度评价新模型及其论证.中成药,2009,31(1):1-4
 GU Rui-min,GUO Zhi-Xin,LIU Wei-wei,et al.Study of a new method on similarity evaluation for chromatographic fingerprint as applied in traditional Chinese medicine.Chinese Traditional Patent Medicine,2009,31(1):1-4
- [8] 贺福元,周宏灏,邓凯文,等.指纹图谱的一种定性定量研究新方法:总量统计矩分析法.药学学报,2008,43(2):195-201 HE Fu-yuan,ZHOU Hong-hao,DENG Kai-wen,et al.A new qualitative and quantitative analytical method of chromatographic fingerprints:total quantum statistical moment.Acta Pharmaceutica Sinica,2008,43(2):195-201
- [9] Wang Kang, Du Kai, Li Hua. The similarity algorithm study on fingerprint of traditional Chinese medicine based on relative entropy. Computers and Applied Chemistry, 2007, 24(1):49-52
- [10] Feng Gan,Runyi Ye.New approach on similarity analysis of chromatographic fingerprint of herbal medicine. Journal of Chromatography A,2006,1104:100-105

(收稿日期:2011年7月11日)