

文章编号: 2096-1472(2017)-09-01-04

## BH-Quant智能量化策略辅助设计平台的研究与实践

班子寒, 张 阳

(北京航空航天大学计算机学院, 北京 100191)

**摘 要:** 量化交易已在国外高频金融交易领域的广泛应用, 但国内现有量化交易平台存在数据模型单一、策略有限、辅助功能少等不足, 为解决上述问题, 本文设计了一种智能的量化策略辅助设计平台, 该平台引入了深度学习框架, 利用人工智能技术训练模型和设计策略, 并将预测结果通过可视化技术呈现, 为投资者提供按自己的风险偏好选择投资组合提供辅助设计平台。

**关键词:** 量化交易; 机器学习; 深度学习; 投资组合

**中图分类号:** TP391.2 **文献标识码:** A

### Research and Practice of the BH-Quant Intelligent Quantization Strategy Aided Design Platform

BAN Zihan, ZHANG Yang

(School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

**Abstract:** Quantitative trading has been widely applied in foreign high-frequency financial transactions, but there are several deficiencies in the existing domestic quantitative trading platform, such as the single variety of data models, limited strategies and insufficient auxiliary functions. In order to solve the above problems, this paper designs an intelligent quantization strategy aided design platform. The platform introduces a deep learning framework, adopts the technology of the artificial intelligence training model and the design strategy, and presents prediction results through visualization technology, which provides an aided design platform for investors to select portfolios based on their own risk preference.

**Keywords:** quantitative trading; machine learning; deep learning; portfolio

#### 1 引言(Introduction)

近年来国内证券、期货市场震荡频繁, 传统投资策略表现平平, 更多的投资者将目光转向于以绝对收益为目标的量化交易投资策略, 以期获得更好的收益。

量化交易是综合利用金融、数学、物理学、计算机等知识和工具, 通过建立模型进行投资决策。作为一种新生的交易方式, 量化交易随着计算机技术的不断发展而逐渐成熟完善, 并在过去十几年的时间里, 以其稳定优异的业绩表现不断扩大市场规模, 越来越得到投资者们的认可<sup>[1]</sup>。

量化交易在美国等发达的金融市场已经较为成熟, 而在中国正刚刚兴起, 并在快速发展。自2012年5月以来, 中国的资产管理行业出现了一轮监管放松, 业务创新的浪潮, 接着这个机遇, 量化交易由于其特有的优势在证券市场发挥着越来越重要的作用。正是由于国内资本市场的不断创新、政策的积极推动、监管的不断完善, 近年来涌现出了一批以聚宽、优矿等为代表的量化交易平台和量化投资团队, 但交易平台存在数据模型单一、策略有限、辅助功能缺失等不足。为解决上述问题, 本文设计了一种智能的量化策略辅助设计平台, 并进行了技术验证。

#### 2 智能量化策略辅助设计平台的意义(The significance of intelligent quantization strategy aided design platform)

本平台的预期使用者是具有一定编码能力的金融专业从业人员, 设计目的是在于为其打造一套丰富全面的量化交易工具包, 辅助其进行基于机器学习和深度学习的智能量化策略的开发。

首先, 对于较为激进的用户, 在传统回测模拟的基础上, 本文为其封装机器学习和深度学习的方法工具, 以期帮助用户从原始的金融数据中挖掘出更多, 更有价值的市场信息。

其次, 对于相对保守的用户, 辅助提供基于skyline detection的投资组合推荐, 为不同需求的用户提供投资组合。根据不同用户对于风险的不同承受能力, 推荐在可承受风险下, 用户可获得最理想收益的股票或基金组合。

最后, 对于监管部门, 提供异常检测机制, 量化监测市场可能发生的波动和异常状态。

#### 3 智能量化策略辅助设计平台设计(Design of intelligent quantization strategy aided design platform)

##### 3.1 平台总体设计

量化交易是通过使用者编写的“策略”来对历史交易数

据进行特征分析提取，通过计算机自动交易来避免的“人”的感性行为<sup>[2]</sup>。本文是依据此需求建立了一个量化交易策略设计平台，在量化交易策略的设计中引入了SVM、HMM、深度学习等模型，依据历史数据训练模型和回测模拟，根据回测结果优化模型或引入新的模型进而得到好的量化交易策略，发现优质的智能策略，通过引入MPT模型进行投资组合，找到更合理的投资组合方式，如图1所示。

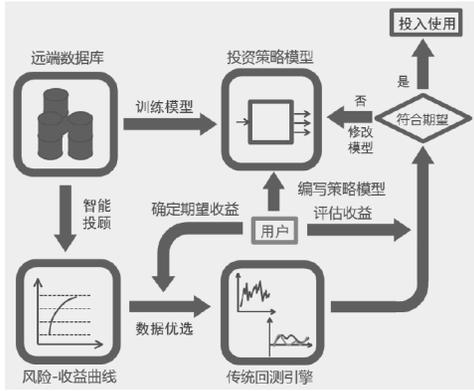


图1 平台总体设计

Fig.1 Platform overall design

### 3.2 平台关键模型

#### 3.2.1 回测模拟模型

##### (1)回测引擎处理逻辑

金融市场中，回测是一种评估方法，通过模拟实盘交易来判断投资策略的优劣，进而为策略是否投入实盘提供可靠依据。

回测引擎需要模拟交易市场和投资者两个客体，并实现两者的交互。其中交易市场由回测代码实现，投资者由用户借助API编写的代码实现。

模拟交易市场由回测框架实现，尽可能真实地再现了实盘。需要注意的是，模拟过程中，市场行为不受用户编写的策略影响，这与实盘有所不同。具体实现时将用户对策略进行诸多限制，以避免用户做出的操作使模拟失真。

回测引擎的另一组成是由用户模拟的投资者。用户主要参与设定初始参数(如股票池、税率、起止时间、时间片细度等)和编写投资策略两个过程。一方面，用户调用设置参数的API来决定交易市场运行的初始状态，回测引擎会按需模拟部分市场而非整个大盘；另一方面，用户调用查询大盘和持仓的API了解市场状态，并编写投资策略做出反馈。投资策略作为测试主体，集合了用户在交易市场中的所有决策，将由回测引擎频繁调用，进而改变用户持仓与收益。

##### (2)回测引擎算法实现

a.回测环境初始化：调用用户编写的initialize函数。

b.模拟交易市场：

每个交易日开始时，将不可动持仓变为可动持仓；

若时间片为天，则执行用户策略，调用用户编写的handle\_data函数；

处理用户策略产生的订单；

若时间片为分钟，则每当交易开盘(9:31—11:30和13:01—15:00)，执行用户策略，调用用户编写的handle\_data函数；

处理用户策略产生的订单；

每个交易日结束时，取消所有未完成订单，更新交易日志。

c.绘各种金融曲线：

通过股票历史数据绘制K线；

通过回测产生的数据绘制收益、基准收益曲线；

计算 $\alpha$ 值、 $\beta$ 值。回测算法流程图，如图2所示。

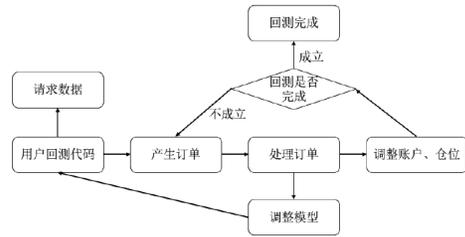


图2 回测算法流程

Fig.2 Regression algorithm flow

#### 3.2.2 智能量化交易策略模型

##### (1)聚类模型-HMM

HMM模型，又称隐马尔科夫模型。HMM与马尔科夫模型相比，不同的地方在于隐藏变量。在股票市场中本文无法准确知道当前时刻的市场状态，而市场状态决定了择时策略<sup>[3]</sup>。但本文可以通过一系列的观察变量来进行猜测，比如通过股票收益率、成交量、主力资金流向、融资余额增长量等观测数据对市场状态(即隐藏状态)进行猜测，并得到第二天的市场状态预测。这样一个基于HMM的量化模型的雏形就出来了。

结合实际使用，本文的模型设置如下：

a.隐藏状态数：6。

b.特征向量：一日对数收益差、五日对数收益差、当日对数高低价差、当日成交量、对数融资余额差，总计五个。

c.假设所有的特征向量的状态服从高斯分布(强假设，便于求解)。平台对用户按需求选取的训练集样本数据提取特征向量，对模型进行训练获得初始模型的各项参数：各个时刻对应的状态序列、混合分布的均值和方差矩阵、混合分布的权重矩阵、状态间转移概率矩阵。

以A股市场某只股票从2010年4月1日至2016年3月17日(包含牛熊市)的数据用作测试集为例，得到的状态序列如图3所示。

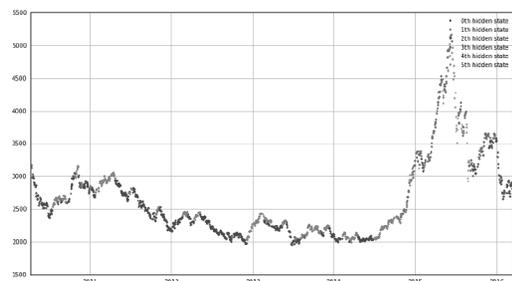


图3 HMM模型状态序列(以2010/4/4至2016/3/17为例)

Fig.3 HMM model state sequence (take 2010/4/4 to 2016/3/17 as an example)

可以看到，不同的市场情况确实通过HMM被学习出来。比如，深蓝色的状态0表示熊市中的下跌阶段，浅蓝色的状态5表示熊市中的反弹阶段等。

按照同样的规则从测试数据集中提取特征向量，作为已训练完成模型的输入，可以得到测试集时间序列对应的隐式状态链。用户可根据某一时刻的状态(涨跌)选择此时执行的策略。同时，已经使用过的时间序列数据作为二次输入对初始

模型进行不断回滚矫正，在使用中提高模型的准确度。

本文随机选取浦发银行(股票代码“SH600000”)为股票池，以2013年全年的数据作为训练集，提取特征向量对模型进行训练，然后应用于2014/5/1至2014/10/1日进行模拟回测交易，得到的策略收益曲线(红线)和基准收益曲线(蓝线)如下。本文可以清晰地看到，通过HMM模型对市场状态的预测，投资者成功避开了2014/6/17日的大跌，清仓保值，而在2014/7/29日的大涨趋势中，投资者持仓增值，获得了相当不错的收益。如图4所示。

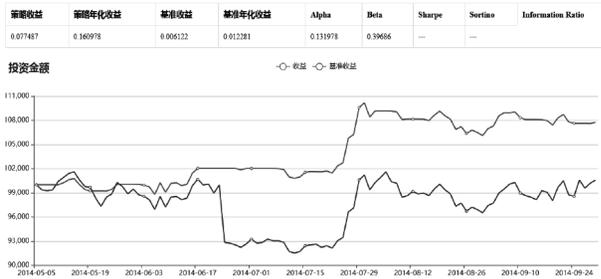


图4 基于HMM回测收益曲线(以2010/5/1至2016/10/1浦发银行为例)

Fig.4 Return curve based on HMM regression (taking 2010/5/1 to 2016/10/1 Pudong Development Bank as an example)

(2)分类模型-SVM

Machine Learning中的支持向量机(SVM)算法常常作为分类器使用。而股票期货等时间序列数据的变化则可以认为是市场状态的表现，例如可以将市场分为震荡状态和趋势状态。因此，本文可以引入SVM算法来对市场状态进行预测。

模型的输入数据为某只股票、期货价格时间序列；通过相应的时间序列来对每一天的状态进行标定状态；由此得到数据集，通过这些数据集作为测试集来训练分类器(由前t个时间点预测t+1时间点的状态)，经过交叉验证、网格优化之后得到市场状态预测模型。

本文采用了沪深300指数在10年内的价格作为时间序列进行了测试，通过标定标签进行训练之后得到分类器之后对市场状态进行了预测，从而得到市场状态图，如图5所示。

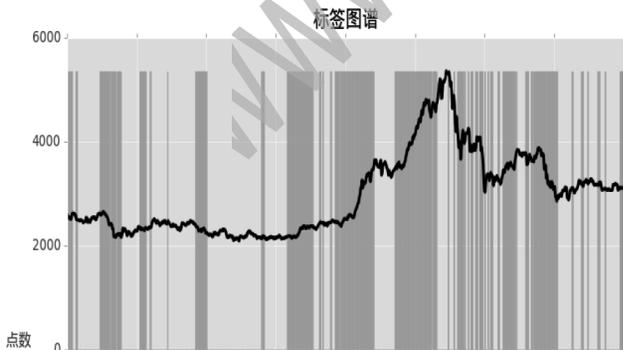


图5 市场状态图(黄色代表趋势状态、白色代表震荡状态)

Fig.5 Market status diagram (yellow represents trend state, white represents shock state)

本文随机选取浦发银行为股票池，以2011/1/1至2012/10/1的数据作为训练集，提取特征向量对模型进行训练，然后应用于2013/4/1至2014/8/1进行模拟回测交易，得到的收益曲线如图6所示。

策略收益	策略年化收益	基准收益	基准年化收益	Alpha	Beta	Sharpe	Sortino	Information Ratio
0.212088	0.159044	-0.038622	-0.026905	-0.017967	0.550819	---	---	---



图6 基于SVM回测收益曲线(以2013/4/1至2014/8/1浦发银行为例)

Fig.6 Return curve based on SVM regression(taking 2013/4/1 to 2014/8/1 Pudong Development Bank as an example)

(3)深度学习模型

深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。应用于股票市场的预测分析，本文的deep learning模型采用三层卷积，两层全连接，经过训练调参后，以一个月(20天)，每天15min为间隔的收盘价作为模型输入，预测未来一周(5天)该股票上涨2%的可能性。输出值为0—1的数字，数值越大代表可能性越高。

该模型适用于大规模的组合投资，应用此模型，本文可以在股票池中放入100只股票，对于每只股票，当模型预测可能性高于既定阈值(如0.9)才出手，这样也许对于某只股票的买卖次数非常少，但是对于整个股票池交易还算频繁，也能保证交易的安全。深度学习神经网络示意图，如图7所示。

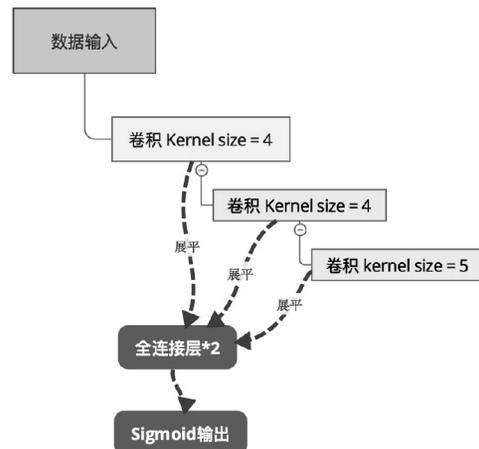


图7 深度学习神经网络示意图

Fig.7 Sketch map of deep learning neural network

3.2.3 投资组合模型

金融产品获取收益的同时，也需要承担一定的损失风险。为分散风险，通常采取同时持有多支产品的方式。投资者在寻求高收益是希望风险尽可能低。投资组合基于MPT模型，根据用户希望的收益给出风险最低的配置建议<sup>[4,5]</sup>。

现代资产配置模型(MPT模型)用收益率的标准差来衡量风险。对于任意一个预期收益值，试图找到一个产品配置权重P，使得组合收益达到预期，且P的方差是最小的。问题的解可由拉格朗日乘法找出。限于计算复杂，故采用随机组合的方法求解。MPT模型算法流程图，如图8所示。



图8 MPT模型算法流程图

Fig.8 Flow chart of MPT model algorithm

图9为随机组合绘制的风险—收益图，其中每一个点代表一种资产配置方式。点阵的左边界曲线呈抛物线形，被称为有效前沿。有效前沿上的点，在固定收益率的情况下能配置出的最小标准差，因此对于相同收益有最小风险。对于相同风险，应选择收益较大的配置，故取抛物线的上半支作为推荐配置提供给用户。风险—收益曲线，如图9所示。

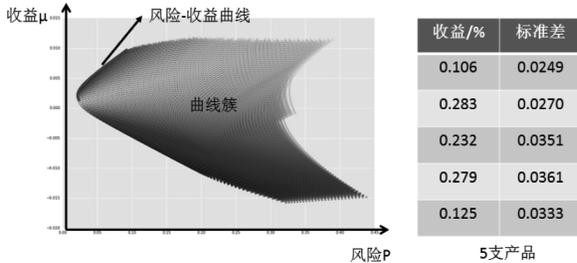


图9 风险—收益曲线

Fig.9 Risk return curve

将产品与推荐配置绘制成图表。可以看出，推荐配置的收益约等于各支产品的平均，但风险比其中任意一支都小。风险—收益率对比图，如图10所示。

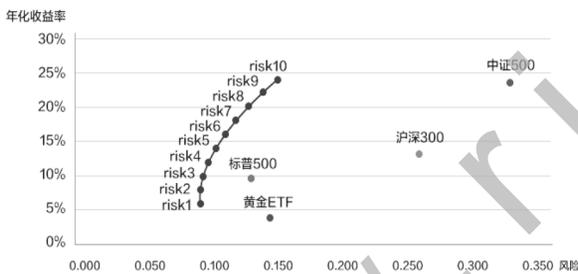


图10 风险—收益率对比图

Fig.10 Risk yield ratio chart

本文随机选取浦发银行、中国国贸、上海机场、黄山旅游、中国医药五只股票作为股票池进行组合投资。从2014/1/1至2014/5/1的回归收益曲线图如图11所示。

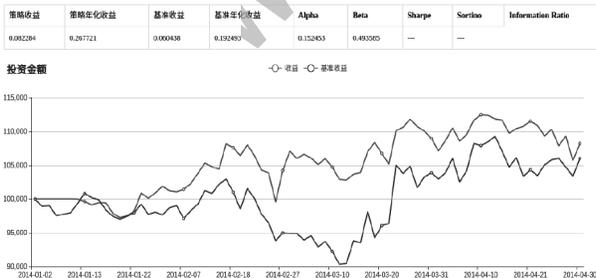


图11 基于投资组合的回归收益曲线(以2014/1/1至2014/5/1为例)

Fig.11 Regression yield curve based on portfolio (from 2014/1/1 to 2014/5/1)

### 3.2.4 异常检测模型

对于股票市场，本文总能找到这样一对或多对股票，两两之间的价格变动趋势具有极强的相关性，基于股票市场的

协整性原则，配对的两只股票总是呈现相同或相反的价格变动趋势，即股票的价格最终会回归至其均衡价值。

那么本文可以这样定义异常：首先本文通过计算股票历史数据计算皮尔森相关系数，挖掘相似度最高的两只股票进行配对。

当配对的两只股票间股价远高于或远低于其均衡价值时，本文就可以根据其协整性进行套利。相关系数较高的两只股票，如图12所示。

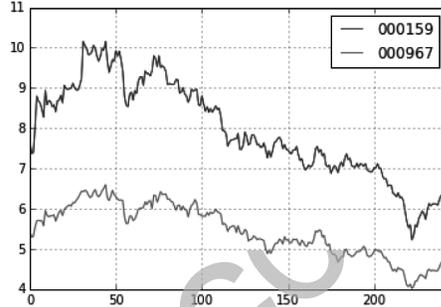


图12 相关系数较高的两只股票

Fig.12 Two stocks with higher correlation coefficients

基于此，本文可以建立基于异常检测机制的套利策略：

(1)初次建仓：当A股票股价高于其基于B股票股价的均衡价值1倍标准差以上时，买入B股票；当A股票股价低于其基于B股票股价的均衡价值一倍标准差以上时，买入A股票。

(2)调仓操作A：当A股票股价高于其基于B股票股价的均衡价值1倍标准差以上时，若A股票的持仓数量大于0，则卖出A股票并买入相同价值的B股票。

调仓操作B：当A股票股价低于其基于B股票股价的均衡价值一倍标准差以上时，若B股票的持仓数量大于0，则卖出B股票并买入相同价值的A股票。

(3)平仓：交易期结束时平仓。

## 4 智能量化策略辅助设计平台应用(Application of intelligent quantization strategy aided design platform)

量化交易是通过使用者编写的“策略”来对历史交易数据进行特征分析提取，通过计算机自动交易来避免的“人”的感性行为。本系统在此需求基础上建立了一个量化交易平台，通过规范统一的接口简化了数据获取、数据分析、数据特征提取等一系列操作，使得使用者能够专注于策略的开发，其主要应用有几个方面。

### 4.1 入门指导

平台为使用者提供了各类金融数据，简易策略示例，统一的金融指标API接口、基础的使用指导等一系列内容，刚接触量化交易的用户可以从这些内容中快速入门，大大降低了学习的门槛，从而在很短的时间内对量化交易有一个粗略但全面的认识，专注于更加重要的“策略”撰写中。

### 4.2 交易评估

对于一般的使用者，平台可以提供趋势分析、行情预测等信息，帮助使用者了解当前形势，同时根据使用者的类型(保守、稳健、激进等)提供对应的投资组合的建议，评估其投资组合的预期收益及风险分析。

### 4.3 投资组合

对于一般性用户平台可以根据用户选择类型提供投资建

议,对于专业金融机构则可以提供相应的策略分析,通过平台中集成的机器学习进行相应的策略优化,从而尽可能达到更高的投资收益比;此外还将对策略风险进行评估,提供更加详细的风险应对措施。

#### 4.4 市场监管

通过一系列机器学习算法,对交易记录进行量化分析,筛选出其中存在的异常行为,从而为监管部门提供咨询建议,及时对异常市场交易做出处理。

### 5 智能量化策略辅助设计平台后期规划(Post planning of intelligent quantization strategy aided design platform)

平台初步完成之后,还需要对已有功能进行完善,后续的一些规划如下:

#### 5.1 引入市场情报

金融市场十分复杂,行情走势不仅仅受到交易信息的影响,还受到政策、各类事件等影响。因此,之后平台将会引入市场情报(包括各类利好利空消息),通过一写方式进行量化后为用户提供API以供引入策略中。

#### 5.2 引入更多机器学习模型

通过机器学习可以更加高效地挖掘出数据中的隐藏信息,找出更多无法通过人工方式发现的市场规律。因此,之后平台中将引入更多的基础性的机器学习模型,用户可以很方便地将其嵌入自己编写的策略中。同时在为相关用户提供咨询建议时,也会提供更加准确且隐藏更深的市场规律。

### 6 结论(Conclusion)

得益于量化交易观念逐渐引起人们的关注,相应的交易

(上接第8页)

由图7可见,比起双曲线渐近线法,基于插值点连线交点的方法重建结果也可以达到预期效果,没有出现由面二义性导致的表面孔洞现象。而二者的差别主要在计算量上,重建上述图形所需时间上表所示,使用插值点连线交点法重建三维表面所需时间少于双曲线渐近线法,而随着数据场规模的增大,这一特点将表现得更为明显。

### 6 结论(Conclusion)

Marching Cubes算法的面二义性问题是因连接方式不确定导致的。当两个相邻体素的公共面是二义性面时,只要保证两个体素在公共面上的等值线连接方式一样,就可以解决这一问题。本文借鉴了双曲线渐近线方法的思路,通过将二义性面进行区域划分来确定一种唯一的连接方式,不但解决了面二义性问题,而且因为计算简单,使得算法更加简洁高效,取得了较好的效果。

#### 参考文献(References)

[1] Minh Chang, et al. Interactive marching cubes algorithm for intraoral scanners[J]. The International Journal of Advanced Manufacturing Technology, 2017, 89(5-8): 2053-2062.

平台也在不断涌现。相比于传统的平台仅提供给用户编写策略的工具和数据,将机器学习等工具引入其中,并且将其封装成策略提供给用户,用户可以将平台已有策略与自身编写策略相结合,从而使得其达到更优的效果;依据机器学习算法编写的市场状态模型,可以更加准确地为用户、监管者提供市场状态信息。

#### 参考文献(References)

[1] 谢堃江. 量化交易策略综述与新策略设计[D]. 浙江: 浙江大学, 2016.

[2] Lawrence R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition[J]. Proceedings of the IEEE, 1989, 77(2): 257-286.

[3] Y. Kawahara, T. Yairi, K. Machida. Change-point detection in time-series data based on subspace identification[J]. Seventh IEEE International Conference on, IEEE, 2007: 559-564.

[4] Hall L O. Exploring Big Data with Scalable Soft Clustering[J]. Proc of the 6th International Conference on Soft Methods in Probability and Statistics. Konstanz, Germany, 2012: 11-15.

[5] Song M, et al. A Comparative Study of Dimensionality Reduction Techniques to Enhance Trace Clustering Performances[J]. Expert Systems with Applications, 2013, 40(9): 3722-3737.

#### 作者简介:

班子寒(1996-), 男, 本科生. 研究领域: 数据挖掘, 深度学习.

张 雷(1996-), 男, 本科生. 研究领域: 数据挖掘, 深度学习.

[2] Seungki Kim, et al. A Novel Interpolation Scheme for Dual Marching Cubes on Octree Volume Fraction Data[J]. Computers & Graphics, 2017, 66(8): 169-178.

[3] Adriano Lopes and Ken Brodlie. Improving the Robustness and Accuracy of the Marching Cubes Algorithm for Isosurface[J]. IEEE Transactions on Visualization and Computer Graphics, 2003, 9(1): 16-29.

[4] Durst, M. J. Additional reference to Marching cubes[J]. Comput. Graph, 1988, 22(2): 72-73.

[5] P. Cignoni, et al. Reconstruction of topologically correct and adaptive trilinear surfaces. Computers and Graphics, 2000, 24(3): 399-418.

[6] Nielson G, Hamann B. The asymptotic decider: resolving the ambiguity in marching cubes[C]. Proceedings of Visualization'91, Los Alamitos CA, 1991.

#### 作者简介:

王 铮(1985-), 男, 硕士, 助教. 研究领域: 计算机图形学.

李瑞明(1989-), 男, 硕士, 助教. 研究领域: 算法研究.