

# 基于层级实时记忆模型的用电行为预测研究

电力用户的用电行为数据中蕴含巨大的商业价值。以层级实时记忆理论为基础,利用稀疏离散表征、空间沉积池技术来表征数据中蕴含的特征,使用KNN分类算法构建分类器,对国网冀北公司的大工业用户用电行为数据进行学习,得到电力容量变更行为预测模型,并与其他主流的电力行为预测方法进行比较。实验证明,HTM模型具有更高的准确度,并可在线预测,减少预测过程中的过拟合问题,提高自适应性。

傅 军<sup>1</sup> 许 鑫<sup>1</sup> 罗 迪<sup>2</sup> 朱天博<sup>1</sup> 魏志丹<sup>3</sup>/1.国网冀北电力有限公司电力科学研究院 2.北京科技大学计算机与通信工程学院 3.北京博望华科科技有限公司

## 0 引言

用电行为包括违约用电、实际电能消费、购电渠道选择以及容量变更等电力用户在消费过程表现出来的行为特征。用电行为从本质上讲是一种商业行为,遵守利益最大化的原则;与一般消费行为不同,影响因素之间存在非线性、强耦合关系,利用已有预测技术难以实现用电行为的准确预测。

用电行为预测可以帮助电力企业掌握电力用户的用电模式、消费习惯和行为偏好,通过向电力用户推荐个性化服务、规避经营的风险等提供增值服务。用电行为预测技术的研究一直是电力数据分析挖掘的热点。文献[1]通过构建LM神经网络预测模型对24个负荷点进行预测,可获得较好的预测准确度;文献[2]将核偏最小二乘回归算法应用于线损率预测,利用历年线损率及其相关数据建

立预测模型,提高预测准确度,并克服非线性因素对预测模型的影响;文献[3]综合考虑了温度及日期类型等因素对日最大负荷的影响,提出了一种采用模糊神经网络的预测模型构建方法,用于短期电力负荷预测,取得了较好的预测结果;文献[4]对影响电费回收的关键因素及其影响程度进行分析,可对用户未来的欠费风险进行预测。

由文献分析可以看出,对于用电行为的预测多使用神经网络算法或者各种回归算法;这些方法对用电行为的动态非线性特征的适应性较差,预测准确度偏低,且需要经常调整模型参数,甚至对模型进行重新训练。层级实时记忆(Hierarchical Temporal Memory, HTM)算法是一种基于脑皮质学习机理的在线机器学习算法,其核心思想是模拟人类智能的来源——新大脑皮质层的工作原理,对信息进行学习、模式识别和预测。2005年,George与Jeff正式提出了HTM的雏形,并给出基本结构和学习过程,通过实验展示了HTM模型的性质,对一些生物学现象进行解释<sup>[5]</sup>。HTM模型现已应用在恶意检测<sup>[6]</sup>、谣言传播<sup>[7]</sup>、手写数字识别<sup>[8]</sup>及金融市场<sup>[9]</sup>中,均取得了优秀的结果。Numenta公司开发了基于HTM理论的Nupic平台,并提供相应的类库和接口来构建预测模型。本文基于HTM理论将稀疏离散表征和空间沉积池技术结合在一起,构建用电行为预测模型,并用于大工业用户电力容量变更行为预测。



傅 军/高级工程师

### 关键词/Keywords

层级实时记忆·  
空间沉积池·  
稀疏离散表征·  
用电行为预测·  
容量变更·

基于大数据的客户增值服务平台关键技术研究及应用(52018K15000G)。

## 1 层级实时记忆原理

### 1.1 层级实时记忆

神经科学发现智能产生于新大脑皮层,新大脑皮层具有层级性:低层级负责学习记忆较为具体的事物,如点、直线和曲线;较高层级负责利用低层级学习到的事物进一步学习并识别高级、抽象的事物,如由点、直线和曲线构成的事物。另外研究发现,人脑的功能区域具有相同的逻辑结构,差别只是在于不同细胞间建立的连接不同;功能区域共用一个算法或者处理方法来处理不同的信息。这个算法在生物学上称为脑皮质算法,HTM算法就是模拟脑皮质算法的工作原理而得到的。一个简单的三层HTM区域结构如图1所示。HTM区域是识别、学习和预测模式的主要单元,每个HTM区域将识别出来的模式和预测结果向上传递给上级HTM区域,并进行进一步的识别、学习和预测。

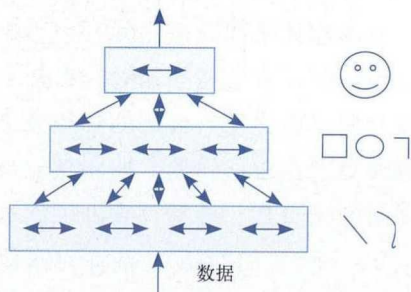


图1 一个简单的HTM模型

HTM模型就是HTM区域对信息进行识别、学习和预测的过程,每个HTM区域都使用相同的方法。HTM区域结构如图2所示。每个区域由很多的细胞柱组成,细胞柱又由多个细胞组成,细胞之间存在着连接,人类的记忆和智能由细胞和细胞

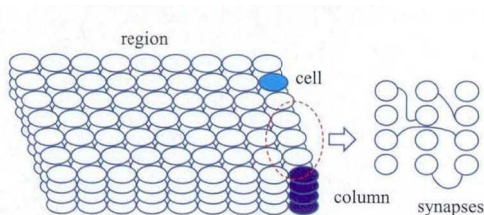


图2 HTM区域结构

间的连接构成。HTM是一种先进的、在线的机器学习算法;本文使用稀疏离散表征和空间沉积池来构建预测模型。

### 1.2 稀疏离散表征

稀疏离散表征是将输入信息稀疏离散化后的表征。解剖学表明,大脑中的信息总是通过数量巨大的神经元细胞中的一小部分活跃细胞来表示。因此,在HTM区域中,用少量、离散的细胞及其连接来表示一个被识别的模式,在编码中表示为一个稀疏矩阵或者稀疏向量。运用稀疏离散表征来表示模式,只需比较模式的关键部分即可识别出整个模式;由于该表示方法是稀疏离散的,因此模式具有优异的鲁棒性和表达能力,改变输入的一小部分不影响整体意义,一旦关键的特征出现即可识别。如不管是遮住人脸的一部分,还是将人脸放大或缩小,甚至将人脸模糊化,都可以识别该人脸属于哪个人<sup>[10]</sup>。

### 1.3 空间沉积池

空间沉积池是HTM模型中的重要过程之一,其主要功能是将输入转化为稀疏离散的模式,即稀疏离散表征。当相同的稀疏离散模式频繁出现时,空间沉积池将该稀疏离散模式识别为一个稳定的模式,并进行学习。本文只讨论单个HTM区域的空间沉积池,学习步骤如下<sup>[11]</sup>:

- 1) 根据输入判定候选细胞柱。
- 2) 从众多候选细胞柱中选择小部分进行激活。
- 3) 加强活跃细胞柱之间连接。

在步骤1)中,空间沉积池接收的输入必须是由固定数量的二进制位组成的输入,HTM区域分配固定数量的细胞柱来接收这些输入,每个细胞柱具有一个相关联的树突区域,每个树突区域有一个潜在突触的集合来表示输入的子集,每个潜在突触都有一个连通值,当连通值超过阈值时,突触变得活跃即连通,突触的连通性由0变成1。初始连通值随机设定,如图3所示。细胞柱活跃突触的个数大于一个阈值时,细胞柱被列入候选细胞柱列表。对于给定的输入,细胞柱具有多少活跃的突触与输入的活跃位(即输入值为1)的数量相关。

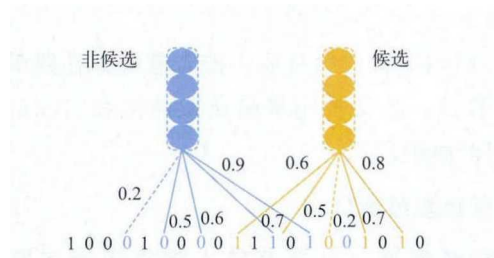


图3 细胞柱接收输入

步骤2)中,为了保证输入的代表足够稀疏离散,只选择很少的一部分候选细胞柱来表示输入。每个候选细胞柱的活跃突触的数量乘以一个促进系数,促进系数由一个细胞柱相对于周围细胞柱的活跃频率动态决定,得到的值从高到低排列,激活值大于阈值。假设阈值为5,则激活值最大的5个细胞柱。设置一个抑制半径,以被激活细胞柱为圆心,抑制半径以内的所有细胞柱都不能被激活。

步骤3)中,如果细胞柱树突区域的神经对应的输入位为1,则增加其连通值(+0.1),否则减少其连通值(-0.1)。通过步骤3),使得频繁出现的模式突触的连通值不断增加,并最终被激活。重复以上步骤,最终HTM区域可以识别并学习到输入中蕴含的模式。

## 2 预测模型构建

### 2.1 算法建模分析

用电行为预测常用建模是时间序列模型、回归模型、神经网络及其改进模型。对于电力容量变更行为的预测来说,电力容量变更行为发生的次数不多,无法充分利用预警作用很大的近期用电变化情况等因素作为预测依据。回归模型、神经网络及其改进模型都需要通过特征工程对模型选取的指标进行降维,存在过拟合的现象,需要间隔一段时间重新进行模型训练。

HTM模型是模拟人类大脑认知机理的在线学习算法;模型具有良好的自适应性,可以根据输入数据不断地调整模型来实现自适应。HTM区域的输入必须是二进制模式,空间沉积池能够极大减少特征工程的工作量。近年来HTM被广泛应用于模式识别领域,在图片分类上取得了令人瞩目的

成效。本文将此模型用于大工业用户电力容量变更规律获取的机器学习。

### 2.2 特征指标选取

抽取国家电网冀北公司2014年1月~2016年5月所有大工业用户的营销数据,包括大工业用户的容量变更记录、用户基本信息以及每月缴费等信息。在已有容量变更信息基础上进行了信息扩展,选取近5个月的用电量、用电增长率以及距上一次容量变更的时间距离,以用户基本信息作为特征,部分关键特征有:前一个月用电量、前两个月用电量、距上一次增容月份数、用电量平均值、用电量方差、近3个月用电平均值、近6个月用电平均值、近3个月容量变更次数、历史容量变更次数、当前月份、用电趋势以及所在地区等。

### 2.3 训练数据处理

在计算得到上述特征指标后,月用电量、原有运行容量等指标为连续型数值。HTM的编码能够在一定程度上实现降噪,且给定最大值和最小值后,不论是属性规范化到一个小的固定范围还是不予处理,数值间的相对分布是一样的。为了处理方便,也方便对数据的预测效果进行对比,将数据集进行数据转换,即将数值都规范化为规定区间的浮点数。最后得到的训练数据全部为区间(0,255)的数值,并将训练数据转化为灰度图像。本文将训练数据分为训练集和测试集,将所有数据分成“预测下个月会进行容量变更”和“预测下个月不会进行容量变更”两类,每一条训练数据为一个灰度图片,得到文件结构如图4所示。



图4 训练集和测试集结构

2.4 构建HTM预测模型

整个HTM模型是一个网状结构，完整的HTM网络包含多个区域，每个区域都实现了HTM模型。一个HTM预测模型包括参数调整、编码、空间沉积池和分类器，其组成部分如图5所示。

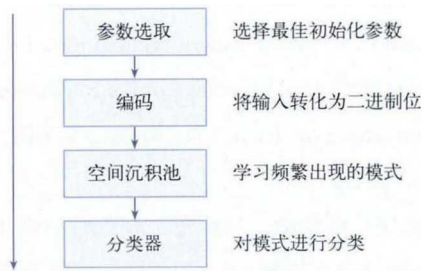


图5 HTM模型组成

HTM模型的参数包括传感器、空间沉积池和分类器参数。空间沉积池的参数包括确定细胞柱的数量、每个细胞柱中细胞的数量、激活阈值和抑制半径等。

适当的编码有助于对目标特征进行预测，但是好的预测模型还和其他参数有关。为了训练一个更加合适的模型，得到更好的预测效果，还需要对空间沉淀池、时间沉淀池的规模和神经的敏感程度进行相关实验。根据经验对一些参数进行简化，如让活跃细胞每次的增量与不活跃细胞每次的减量相等，让时间沉淀池的细胞柱数量与空间沉淀池的细胞柱数量相等。并且根据经验对一些参数采用不等间隔的取值，这样可以大大减少循环次数的量级。

在Nupic框架中首先调用nupic.engine.Network模块得到一个对象，然后使用addRegion ()方法，按照先后为Network添加编码器、空间沉积池和分类器三个HTM区域对象。同时，下级HTM区域的输出作为上级HTM区域的输入，调用Link方法将三层HTM区域连接起来。ImageSensor区域将数据转化为二进制的输入流，其输出作为输入传送到第二个区域空间沉积池，从输入流中识别出频繁出现的模式，将这些模式作为第三个区域分类器的输入，并将这些模式进行分类。对于HTM网络的最后一个区域，本文选择KNN分类器作为空间沉积池学到模式

的分类器。

3 预测结果分析

本文得到样本集共88 372条记录，其中下个月会容量变更的数据40 346条，下个月不会容量变更的数据48 036条，所有样本集按照种类进行排序。

3.1 整体准确度

采用第一种结构的二分类样本集，并将所有样本集平均分成10份，训练集占1份，构建9个测试集，分别使用这些测试集来检测模型分类的准确度。设置学习/推理模式参数为“1”，“1”代表学习/推理模式开启。“0”代表学习/推理模式关闭。这意味着用训练集得到HTM分类器后，进行测试时同时进行学习。以这种方式，HTM模型不仅可以推断实例属于什么类别，而且根据真实的类别修正模型。测试中设置分为9个部分，计算识别每一部分的准确度。该实验重复10次，最后得到的平均准确度如图6所示。

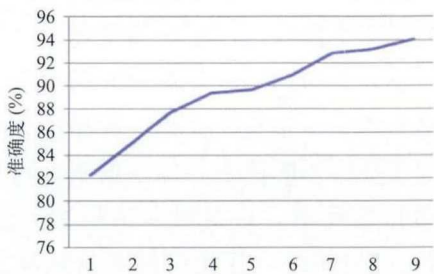


图6 各部分的识别准确度

图6显示出了第一部分的准确度约为82%。当测试集不断增大时，模型边学习边进行预测，准确性不断提高，最后达到94%。结果表明，该HTM模型具有学习新知识和修改自身的能力。

3.2 准确度比较

实验采用相同的样本集，用现有主流的机器学习算法构建预测模型，并对容量变更行为进行预测。结果如下表所示。

表 各种算法的准确度比较

| 模型  | 十折交叉验证 | 10%训练，90%测试 |
|-----|--------|-------------|
| HTM | 95.02% | 94.11%      |
| KNN | 92.73% | 81.31%      |

(续)

| 模型     | 十折交叉验证 | 10%训练, 90%测试 |
|--------|--------|--------------|
| LM神经网络 | 93.16% | 79.28%       |
| SVM    | 95.45% | 83.90%       |

为了比较HTM和其他集成预测方法的性能, 本文使用十折交叉验证用于计算平均值的准确度<sup>[12]</sup>。

从实验结果中可以看出, 在十折交叉验证中, HTM表现优秀, 只有SVM模型能够超越其准确度, 但是在使用10%的样本数据作为训练集、90%的样本数据作为测试集进行检验时, HTM由于其在线学习能力很强, 能够根据测试集不断调整自身模型; 可有效减少过拟合现象, 达到一个较高的准确度。

## 4 结束语

本文基于HTM模型构建了用电行为的预测模型, 并以容量变更行为的预测为例来验证模型的准确性。在实验部分, 本文分析了HTM模型和其他模型的预测结果对比, HTM模型在学习模式有更好的表现, 能够达到一个更大的识别率, 而且在小训练集的情况下具有优势。在测试集不断增多的情况下能够不断修正存储的模型, 同时不断提高识别的准确率。目前对于其研究只是使用了单层的HTM网络, 且使用的不是时间序列数据而是静态数据, 仍然需要将其延伸到多层, 并且对于预测的某些细节部分的研究还有很长的路要走。

### 参考文献

- [1] 罗静. 电力负荷数据预测方法模型设计与分析[D]. 上海: 复旦大学, 2008.
- [2] 王海燕. 基于核偏最小二乘回归分析的线损率预测[J]. 计算机仿真, 2012, 29(11): 323-326.

- [3] 于海燕, 张凤玲. 基于模糊神经网络的电力负荷短期预测[J]. 电网技术, 2007, 31(3): 68-72.
- [4] 赵永良, 秦萱, 吴尚远, 等. 基于数据挖掘的高压用户电费回收风险预测[J]. 电力信息与通信技术, 2015, 13(9): 57-61.
- [5] George D, Hawkins J. A hierarchical bayesian model of invariant pattern recognition in the visual cortex [C]. In Proceedings of the International Joint Conference on Neural Networks, IEEE, 2005: 1812-1817.
- [6] 仇慎健, 张仕斌, 刘莘光. 基于HTM模型的恶意Android应用检测[J]. 四川理工学院学报: 自然科学版, 2015, 28(4): 50-56.
- [7] 王薇. 基于HTM的谣言传播个体记忆一预测模型研究[D]. 武汉: 华中科技大学, 2013.
- [8] 陈雪, 朱敏, 钟煜, 等. 基于HTM的离线手写签名识别及改进[J]. 四川大学学报: 工程科学版, 2011, 43(1): 146-150.
- [9] Gabrielsson P, Konig R, Johansson U. Hierarchical temporal memory based algorithmic trading of financial markets [C]. Computational Intelligence for Financial Engineering & Economics, IEEE, 2012: 1-8.
- [10] Ahmad S, Hawkins J. Properties of sparse distributed representations and their application to hierarchical temporal memory [J]. Computer Science, 2015.
- [11] Numenta platform for intelligent computing [EB/OL]. [2015-05-12]. <http://numenta.org/htm-white-paper.html>
- [12] 牛晓太. 基于KNN算法和10折交叉验证法的支持向量选取算法[J]. 华中师范大学学报: 自然科学版, 2014, 48(3): 335-338.

(收稿日期: 2017-03-21) **EA**

欢迎订阅 欢迎投稿 欢迎刊登广告