

文章编号: 2095-2163(2023)03-0143-05

中图分类号: TM933; TM925

文献标志码: A

基于 SSA-SVM 的非侵入式负荷识别

李梓彤, 杨超

(贵州大学 电气工程学院, 贵阳 550025)

摘要: 针对目前非侵入式负荷监测方法对负荷特征相近的电器识别准确率不高的问题, 本文提出了一种基于麻雀搜索算法 (sparrow search algorithm, SSA) 优化支持向量机 (support vector machine, SVM) 的负荷识别方法。该方法除了采用传统的有功和无功作为特征外, 还采用了基波功率因数和频域电流谐波幅值作为新特征, 同时使用麻雀搜索算法对支持向量机的核心参数进行优化, 建立负荷识别模型, 实现对家用电器的有效监测。进而采用 AMPds 数据集对算法进行测试, 通过对比分析, 验证了本文所提方法的有效性。

关键词: 非侵入式负荷监测; 麻雀搜索算法; 支持向量机; 参数优化

Non-intrusive load identification based on Sparrow Search Algorithm Optimized Support Vector Machine

LI Zitong, YANG Chao

(School of Electrical Engineering, Guizhou University, Guiyang 550025, China)

【Abstract】 In order to solve the problem of low accuracy of current non-invasive load monitoring methods in identifying electrical products with similar load characteristics, a load identification method based on sparrow search algorithm (SSA) and optimized support vector machine (SVM) is proposed. This method not only uses traditional active power and reactive power as features, but also uses fundamental power factor and harmonic amplitude of current in frequency domain as new features, and then uses sparrow search algorithm to optimize the core parameters of support vector machine. After that, the load identification model is established to realize the effective monitoring of household appliances. Finally, the algorithm is tested with AMPds data set, and the effectiveness of this method is verified by comparative analysis.

【Key words】 non-intrusive load monitoring; sparrow search algorithm; support vector machine; parameter optimization

0 引言

随着现代社会的快速发展, 智能电网成为一个被广泛讨论的新概念^[1]。居民用户对精准的用电服务需求不断增长, 目前电网公司只向用户提供总电量数据。非侵入式负荷监测 (non-intrusive load monitoring, NILM) 技术最早由 Hart 提出^[2], 具体来说是在电力负荷总进线处获取负荷数据 (电压、电流、功率等), 采用模式识别算法, 通过分析负荷特征量, 实现用户侧负荷类型的辨识及能耗分解。与传统的侵入式负荷监测方法相比, NILM 具有运维简便、基础成本低、信息安全性强、不侵犯用户隐私等优势。

NILM 整个流程分为数据的采集和处理、事件检测、特征提取、负荷辨识等步骤。近年来, 国内外众多学者对 NILM 进行了大量研究, 目前研究重点主要在于负荷特征的提取以及负荷识别算法的优化两个部分。其中, 特征提取主要分为暂态特征和稳态特征, 并且功率是最常用的特征, 文献[3]针对低采样率下识别准确率低的问题提出一种基于极大似然法的负荷识别算法, 采用整数规划和最大似然法进行负荷识别, 平均识别准确率超过 85%。文献[4]将标准化后的负荷电压、电流、V-I 轨迹生成真彩可视化图像, 并将其作为卷积神经网络 (CNN) 的输入进行负荷辨识。文献[5]以稳态电流的时域和频域信息作为特征, 然后采用随机森林进行特征选

基金项目: 贵州省科学技术基金 (黔科合基础[2019]1100)。

作者简介: 李梓彤 (1998-), 女, 硕士研究生, 主要研究方向: 非侵入式负荷监测; 杨超 (1971-), 女, 副教授, 主要研究方向: 配电网规划及电能质量管理。

通讯作者: 杨超 Email: 785622539@qq.com

收稿日期: 2022-04-18

择得到最优特征集,最后使用遗传算法优化极限学习进行负荷识别。文献[6]先以有功无功作为特征采用K-Means聚类进行初步分类,然后再使用带有颜色特征的V-I轨迹作为AlexNet神经网络的输入进行精细化分类。文献[7]在低频采样下采用支持向量机进行负荷识别,对大功率电器的识别效果较好,但对小功率电器识别效果差。除了时域特征外还有很多研究人员采用频域特征,文献[8]采用电流谐波幅值作为特征,在此基础上提出增量特征提取的方法,最后使用模糊聚类进行负荷识别,但对功率相近的纯电阻电器识别准确率不高。文献[9]以有功功率、基波功率因数、电压电流三次谐波含量差作为特征,采用优化后的模糊聚类算法进行识别,实现了对低功率用电设备的辨识。针对负荷特征相似的电器,文献[10]以奇次谐波电流为负荷特征,并使用AdaBoost算法进一步筛选特征,采用K邻近与核线性判别方法相结合进行负荷识别。文献[11]中除了考虑传统的电气特征外,还加入了室外温度、大气压强等非电气特征作为输入,采用随机森林算法进行负荷识别,对于电气特征相似的负荷具有良好的识别准确率。

综上所述,针对负荷识别算法的研究有很多,但大部分算法涉及的负荷种类较少,应用场景也比较简单,且大部分采用的是有功、无功等传统特征作为输入。针对上述问题,本文在有功、无功功率的基础上,加入基波功率因数和谐波特征作为输入,提出一种基于麻雀搜索算法优化支持向量机的负荷识别方法,以提高负荷识别的准确率。

1 电气量负荷特征

有功功率 P 和无功功率 Q 是负荷监测的重要特征,当电气运行状态发生变化时,有功、无功产生相应的变化,不同电器稳态运行时有功、无功存在差异,但有功和无功功率不能直接测量,需要通过电压电流数据进一步计算得到,计算公式如下:

$$P = \sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} U_k I_k \cos(\varphi_k) \quad (1)$$

$$Q = \sum_{k=0}^{\infty} Q_k = \sum_{k=0}^{\infty} U_k I_k \sin(\varphi_k) \quad (2)$$

其中, U_k 表示用电设备第 k 次电压谐波的有效值; I_k 表示用电设备第 k 次电流谐波的有效值; φ 为功率因数角。有功、无功虽然可以识别出绝大部分电器,但容易发生特征重叠现象,因此引入基波功率因数,推得的数学公式如下:

$$\lambda_1 = \frac{P_1}{\sqrt{P_1^2 + Q_1^2}} \quad (3)$$

其中, P_1 为基波有功功率, Q_1 为基波无功功率。

此外,不同种类电器的谐波信息也不尽相同,可以作为用电负荷特征,通过快速傅里叶变换(fast Fourier transform)可以将稳态电流的时域信号转换为频域信号,本文采用三次谐波电流幅值作为负荷特征,FFT对稳态电流信号的谐波分解如下:

$$f(t) = c_0 + \sum_{m=1}^{\infty} c_m \sin(m\omega t + \varphi_m) \quad (4)$$

其中, c_0 为直流分量; c_m 为各次谐波的幅值; $m\omega$ 为各次谐波的角频率; φ_m 为各次谐波的相位角。

2 基于SSA-SVM的非侵入式负荷识别

2.1 麻雀搜索算法

麻雀搜索算法(sparrow search algorithm,SSA)^[12]是2020年提出的一种群体优化算法。该算法通过不断更新个体的位置,模拟麻雀的觅食和反捕食行为。相比传统的优化算法,该算法结构简单,控制参数少,且寻优能力强,收敛速度快。在麻雀搜索算法中,将种群分为发现者、跟随者和预警者。种群具有以下特征:发现者通常拥有较高的能源储备,为所有的跟随者提供觅食方向,只要能找到更好的食物来源,每只麻雀都可以成为发现者,但发现者和跟随者在整个种群中的比例是固定的。对此研究内容拟做阐释分述如下。

(1)初始化麻雀种群位置。研究给出的数学表述如下:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & \cdots & X_{1,d} \\ X_{2,1} & X_{2,2} & \cdots & \cdots & X_{2,d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & \cdots & X_{n,d} \end{bmatrix} \quad (5)$$

其中, n 为麻雀数量, d 表示要优化的变量维度。所有麻雀的适应值可由如下公式来描述:

$$F_X = \begin{bmatrix} f([X_{1,1} & X_{1,2} & \cdots & \cdots & X_{1,d}]) \\ f([X_{2,1} & X_{2,2} & \cdots & \cdots & X_{2,d}]) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f([X_{n,1} & X_{n,2} & \cdots & \cdots & X_{n,d}]) \end{bmatrix} \quad (6)$$

其中, F_X 中每行表示个体麻雀的适应值。在SSA中,适应值较好的发现者在搜索过程中优先获得食物。

(2) 种群中发现者负责寻找食物的方向和位置, 引导追随者向食物移动。发现者位置更新如下:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{max}}\right) & R_2 < ST \\ X_{i,j}^t + Q \cdot L & R_2 \geq ST \end{cases} \quad (7)$$

其中, t 表示当前迭代次数; $X_{i,j}^t$ 表示迭代 t 次时第 i 个麻雀的第 j 维值; $iter_{max}$ 表示最大迭代次数; $\alpha \in (0, 1]$ 是一个随机数; $R_2 \in [0, 1]$ 表示报警值; $ST \in [0.5, 1]$ 表示安全阈值; Q 是服从正态分布的一个随机数; L 是 $1 \times d$ 维的矩阵, 矩阵中的每个元素都是 1。

发现者位置更新规则如下: 若 $R_2 < ST$, 意味着周围没有捕食者, 发现者可以进入广泛搜索模式。若 $R_2 \geq ST$, 则意味着有麻雀发现了捕食者, 所有麻雀都需迅速飞往安全区域。

(3) 跟随者。会频繁地监视发现者, 进而争夺食物。跟随者的位置更新方式如下:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^t - X_{i,j}^t}{i^2}\right) & i > \frac{n}{2} \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \cdot A^+ \cdot L & i \leq \frac{n}{2} \end{cases} \quad (8)$$

其中, X_{worst} 表示当前全局最差位置; X_p 为发现者占据的最佳位置; A 是 1 和 -1 的 $1 \times d$ 维矩阵, $A^+ = A^T (AA^T)^{-1}$ 。当 $i > \frac{n}{2}$ 时, 第 i 个跟随者处于较差位置, 需飞往其他区域进行觅食。当 $i \leq \frac{n}{2}$ 时, 跟随者 i 将在最佳位置 X_p 附近觅食。

(4) 预警者。一般占种群的 10%~20%, 当个体麻雀感知到危险时, 会迅速向安全区域移动, 其位置更新如下:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot |X_{i,j}^t - X_{best}^t| & f_i > f_g \\ X_{i,j}^t + K \cdot \left(\frac{|X_{i,j}^t - X_{worst}^t|}{(f_i - f_w) + \varepsilon}\right) & f_i = f_g \end{cases} \quad (9)$$

其中, X_{best} 表示当前全局最优位置; β 为步长控制参数, 服从均值为 0、方差为 1 的正态分布; $K \in [-1, 1]$ 是一个随机数; ε 为常数, 用于避免分母为 0; f_i 表示当前个体的适应度值, f_g 、 f_w 分别表示当前全局最优和最差的适应度值。当 $f_i \neq f_g$ 时, 麻雀位于种群的边缘位置, 容易遇到危险, 需要向最优位置移动来获得更高的适应度值。当 $f_i = f_g$ 时, 麻雀处于种群中心位置, 当意识到危险时, 该麻雀向其他同伴靠近, 以此远离危险区域。

2.2 支持向量机

支持向量机 (Support Vector Machine, SVM) 是一种基于统计学理论的监督型学习算法^[13], 常用于小样本的分类。主要思想是通过核函数将低维样本映射到高维空间, 从而在高维空间中求出最优分类超平面使得样本线性可分。

当样本集为 $T = \{(x_i, y_i) \mid i = 1, 2, \dots, N\}$, $x_i \in R^d$, $y_i \in Y = \{1, 2, \dots, M\}$, 最优分类超平面方程如下:

$$w^T x + b = 0 \quad (10)$$

其中, x 为输入样本; w 为权重向量; b 为偏置量。

目标函数为:

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } y_i (w^T x_i + b) \geq 1 - \xi_i \end{cases} \quad (11)$$

其中, C 为惩罚因子; ξ_i 为松弛常量, $\xi_i \geq 0, i = 1, 2, \dots, n$ 。

SVM 的分类模型为:

$$f(x) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b\right) \quad (12)$$

其中, α_i 为样本训练中得到的拉式乘子; $K(x_i, x)$ 为核函数; g 为核函数半径; b 为对应的偏差。 C 和 g 的选择决定了 SVM 分类效果的好坏, 需要对这 2 个变量进行寻优来得到更好的模型。

2.3 SSA-SVM 算法在 NILM 的实现

算法流程如图 1 所示。SSA-SVM 负荷识别具体步骤如下。

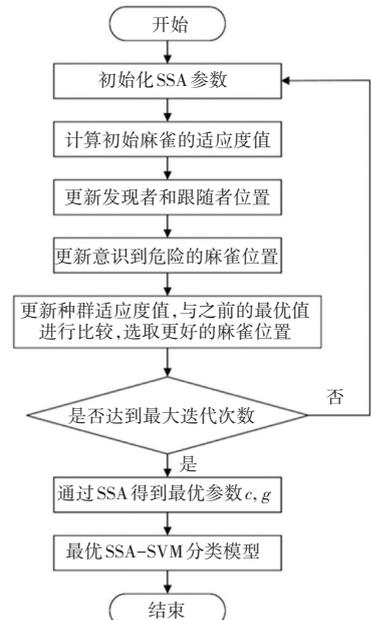


图 1 NILM 方法流程图

Fig. 1 Non-intrusive load identification process

(1)首先提取负荷特征作为模型的输入。建立训练集和测试集样本。

(2)初始化麻雀搜索算法具体参数。包括麻雀数量、最大迭代次数、发现者和预警者麻雀所占比例、SVM 参数 (C, g) 的上边界和下边界。

(3)计算初始所有麻雀的适应度值。根据适应度高低按照比例将麻雀分为发现者和跟随者。确定当前最优适应度值 f_g 和该麻雀对应的最优位置 X_{best} 。

(4)根据预警值的大小,按照式(7)更新发现者的位置,根据式(8)更新跟随者的位置。

(5)根据式(9)更新意识到危险的的麻雀位置,使其迅速向安全区域移动。

(6)计算当前所有麻雀的适应度值,并与之前的最优适应度值进行比较,取全局最优的适应度值以及对应的麻雀位置。

(7)判断是否达到最大迭代次数。若满足条件则输出最优参数,得到最优模型;否则转向步骤(3)。

由于 SVM 的惩罚因子 C 和核函数参数 g 的选取对分类效果影响很大,本文采用麻雀搜索算法对 C 和 g 进行寻优,以得到最优的 SSA-SVM 分类模型。

3 算例分析

本文采用 AMPds 数据集^[14]进行识别效果验证,该数据集记录了 2012 年 4 月 1 日到 2014 年 3 月 3 日期间加拿大一户居民住宅的能耗数据,其中包含了 21 个电表数据、2 个水表数据和 2 个天然气表数据。用电数据包括有功、无功、视在功率、电压、电流、频率、基波功率因数等,采样频率为 1 min。本文从中选取 5 种常用的用电设备作为验证,电器类别标签记为 1,2,3,4,5。负荷信息见表 1。

表 1 负荷信息表

Tab. 1 Load information table

标签	名称	P/W
1	卧室灯	21
2	洗衣机	100~500
3	干衣机	250~5 000
4	冰箱	100~150
5	热泵	1 200~2 000

SSA-SVM 算法的训练集和测试集包含 4 个负荷特征,分别是有功功率、无功功率、基波功率因数和三次谐波电流幅值,形成特征向量:

$$F = [E_1, E_2, E_3, E_4]$$

其中,随机选取 80% 的样本数据作为训练集,20% 的样本数据作为测试集。SSA-SVM 算法参数

设定:麻雀数量为 10,发现者比例为 70%,跟随者比例为 30%,预警麻雀比例为 20%,预警值为 0.6。最大迭代次数为 20,SVM 参数 (C, g) 的下边界设置为 $[0.1, 2^{-5}]$,上边界设置为 $[10, 2^4]$ 。以分类预测错误率作为优化的目标函数值,SSA 优化 SVM 后的分类效果及适应度曲线如图 2 所示。

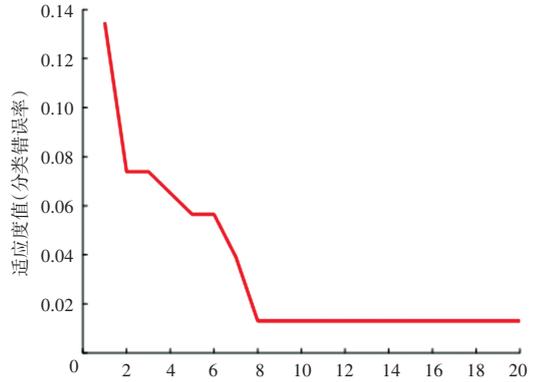


图 2 SSA-SVM 适应度曲线

Fig. 2 SSA-SVM fitness curve

使用 Matlab 软件对本文提出 SSA-SVM 负荷识别模型进行测试,经 SSA 优化后得到的最优参数为: $C = 6.2611, g = 12.6048$ 。得到最优的分类模型后,用测试集样本进行验证,结果如图 3 所示。为了测试模型的性能,模型的识别准确率计算公式为:

$$Accuracy = \left(1 - \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \right) \times 100\% \quad (13)$$

其中, N 为测试集中样本总数目; y_i 为测试集中样本的真实标签值; \hat{y}_i 为模型预测的测试集输出值。

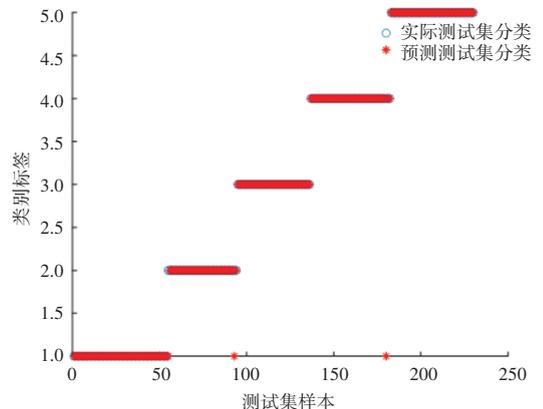


图 3 SSA-SVM 分类结果

Fig. 3 The classification results of SSA-SVM algorithm

为了体现 SSA-SVM 算法的负荷识别效果,使用相同的数据集,分别采用传统的 SVM 算法和 GA-SVM 算法进行负荷识别,其分类结果如图 4、图 5 所示。3 种方法的负荷识别准确率见表 2。

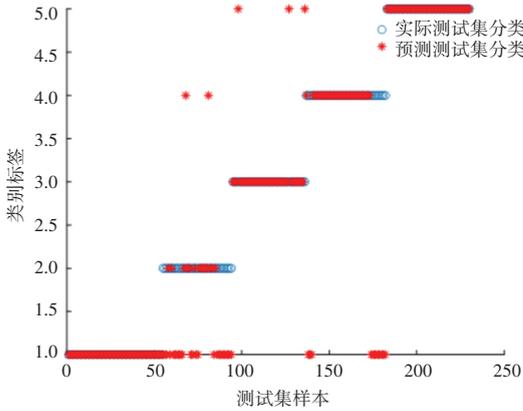


图4 传统 SVM 分类结果

Fig. 4 The classification results of SVM algorithm

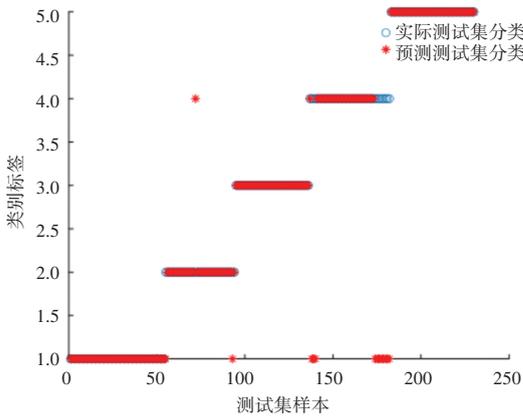


图5 GA-SVM 分类结果

Fig. 5 The classification results of GA-SVM algorithm

表 2 3 种算法识别准确率对比

Tab. 2 Comparison of identification accuracy of three algorithms

标签	传统 SVM	GA-SVM	SSA-SVM
1	100	100	100
2	37.50	92.50	95.00
3	92.86	100	100
4	73.91	73.91	97.83
5	100	100	100
总准确率	82.61	93.48	98.69

由表 2 可以看出,SSA-SVM 算法比其他 2 种算法整体上有更高的识别准确率,测试准确率可达到 98.69%,且对于多状态电器的识别效果要比传统 SVM 和 GA-SVM 好。

4 结束语

针对负荷特征相近的电器识别准确率不高的问题,本文提出了一种基于 SSA-SVM 的负荷识别模型,首先在传统有功、无功功率及基波功率因数的基础上引入电流谐波作为特征,然后采用 SSA 算法对 SVM 的核心参数 C 和 g 进行寻优,通过 SSA-SVM 模型实现负荷识别,实验结果表明,本文方法在多场景情况下具有较高的识别准确率及稳定性。

参考文献

- [1] 余贻鑫. 智能电网的技术组成和实现顺序[J]. 南方电网技术, 2009,3(02): 1-5.
- [2] HART G W. Nonintrusive appliance load monitoring [J]. Proceedings of the IEEE,1992,80(12): 1870-1891.
- [3] KONG Liang, YANG Dongsheng, LUO Yanhong. Non-intrusive load monitoring and identification based on maximum likelihood method [C]// 2017 IEEE International Conference on Energy Internet (ICEI). Beijing, China:IEEE, 2017:268-272.
- [4] 丁昊,杨乐,石鸿凌,等. 利用数据可视化实现智能非侵入式负荷辨识[J]. 华中科技大学学报(自然科学版), 2021,49(10): 85-90.
- [5] 安琪,王占彬,安国庆,等. 基于随机森林-遗传算法-极限学习机的非侵入式负荷识别方法[J]. 科学技术与工程, 2022,22(05): 1929-1935.
- [6] 解洋,梅飞,郑建勇,等. 基于 V-I 轨迹颜色编码的非侵入式识别方法[J]. 电力系统自动化,2022,46(04): 93-102.
- [7] DUFOUR L,GENOUD D,JARA A, et al. A non-intrusive model to predict the exible energy in a residential building [C]//2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). New Orleans, LA, USA: IEEE, 2015: 69-74.
- [8] 孙毅,崔灿,陆俊,等. 基于差量特征提取与模糊聚类的非侵入式负荷监测方法[J]. 电力系统自动化,2017,41(04): 86-91.
- [9] 杜刃刃,杨超,蒲阳. 基于稳态特征和 IGWO-FCM 模糊聚类的非侵入式负荷监测方法[J]. 电测与仪表,2021,58(01): 152-157.
- [10] 宋旭帆,周明,涂京,等. 基于 k-NN 结合核 Fisher 判别的非侵入式负荷监测方法[J]. 电力系统自动化,2018,42(06): 73-80.
- [11] 李如意,张鹏,刘永光,等. 基于随机森林的分侵入式家庭负荷辨识方法[J]. 电测与仪表, 2021,58(04):9-16.
- [12] XUE Jiankai, SHEN Bo. A novel swarm intelligence optimization approach: Sparrow search algorithm[J]. Sys-tems Science and Control Engineering, 2020, 8(1):22-34.
- [13] 丁世飞,齐丙娟,谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报,2011,40(01): 2-10.
- [14] MAKONIN S, POPOWICH F, BARTRAM L, et al. AMPds: A public dataset for load disaggregation and eco-feedback research [C]//2013 IEEE Electrical Power Energy Conference (Epec). Halifax, NS, Canada: IEEE, 2014:1-6.