

文章编号: 1000-5862(2015)01-0073-06

基于材料基因工程的锂离子电池 材料数据生成及数据挖掘平台

肖建茂¹ 汪 浩¹ 欧阳楚英^{2*}

(1. 江西师范大学软件学院, 江西 南昌 330022; 2. 江西师范大学物理与通信电子学院, 江西 南昌 330022)

摘要: 利用材料基因工程方法探索锂离子电池新材料是目前国际上重要的技术手段. 通过高通量计算手段计算生成大量锂离子电池材料的基本物理化学性质数据, 利用数据挖掘技术, 总结材料物理化学性质与材料的组分、组织结构等的构效关系, 进而探索发现新型锂离子电池材料和改性现有材料. 利用材料基因工程基本思想, 设计了利用无机材料晶体结构数据库中的结构数据为源数据, 通过基于 Web 服务器来产生第一性原理计算软件 VASP 的输入文件信息, 并通过 Web 网传输到并行计算中心实现材料性能计算, 生成的材料性质数据再返回到 Web 服务器. 通过对材料结构和性质数据进行挖掘, 总结锂离子电池材料的构效关系, 为锂离子电池材料设计提供技术支持.

关键词: 材料基因工程; 锂离子电池材料; 数据计算; 数据挖掘

中图分类号: TP 311; TB 383 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2015.01.14

0 引言

“材料基因工程”方法与“人类基因组工程”方法类似, 通过高通量的第一性原理计算, 结合已知的可靠实验数据, 用理论模拟去尝试尽可能多的真实或未知材料, 建立其化学组分、晶体结构和各种物性的数据库, 并利用计算机数据处理方法, 通过数据挖掘技术探寻材料结构和性能之间的构效关系模式, 为材料设计提供更多的信息, 拓宽材料筛选范围, 预知材料各项性能, 进而达到缩短材料性质优化和测试周期, 加速材料研究的创新^[1].

以美国为首, 2011年6月24日, 美国总统奥巴马宣布启动一项价值超过5亿美元的先进制造业伙伴关系 (Advanced Manufacturing Partnership, AMP) 计划, 呼吁美国政府、高校及企业之间应加强合作, 以强化美国制造业领先地位, 而材料基因组计划 (Materials Genome Initiative, MGI) 作为 AMP^[2] 计划中的重要组成部分, 投资将超过1亿美元. 而在国内, 2011年12月21日到23日召开了以“材料科学系统工程”为主题的香山科学会议^[3]. 该会的主题报告建议选择“以高温合金为重点的金属材料”和

“以2次电池为重点的能量储存和转换材料”进行示范, 为更大范围的推广积累经验^[3]. 自此, 国内多个课题组开始了利用材料基因工程方法来研究和设计锂离子电池2次电池材料.

锂离子电池是目前综合性能最好的2次电池体系, 具有能量密度高、循环寿命长、对环境污染小等优点, 目前已经在很多消费电子领域获得了广泛的应用. 然而, 随着人们对可移动能源的需求越来越大, 对其要求也越来越高. 比如, 汽车用动力电池需要更高的能量密度和更苛刻的安全性能. 而电池性能的提高, 在很大程度上依赖于电池材料性能的提高^[4]. 近几十年来, 已有大量的研究人员在从事锂离子电池材料的改性和开发, 但收效甚小. 比如, 正极材料到目前为止仅有层状的 LiMO_2 ($M = \text{Co}, \text{Ni}, \text{Mn}$) 体系, 尖晶石结构的 LiMn_2O_4 材料体系以及橄榄石结构的 LiMPO_4 ($M = \text{Fe}, \text{Mn}$) 体系这3类. 究其原因在于目前主流的材料开发是“试错法” (Trial and Error), 而这种模式效率低、周期长、针对性差. 受益于计算技术的快速发展, 利用计算机模拟技术来设计材料, 已经慢慢显现出其强大的功能. 对锂离子电池材料的设计, 近年来也已经取得了一些进展^[5-6]. 而材料基因工程技术正是基于计算机模拟

收稿日期: 2014-12-29

基金项目: 国家自然科学基金 (11234013) 和江西省自然科学基金 (20133ACB21010) 资助项目.

通信作者: 欧阳楚英 (1976-), 男, 江西吉安人, 教授, 博士, 主要从事锂离子电池材料物理方面的研究.

技术对材料设计的快速发展和大量经验积累的基础上,让材料设计更加系统化、科学化、规范化。

材料基因工程计划的实现包括 3 个方面: (i) 材料高通量计算系统; (ii) 材料制备和性能测试系统; (iii) 材料性能数据库与信息平台。就第 1 个方面,材料的高通量计算系统而言,近十几年来,锂离子电池材料的计算机模拟和设计已经有了足够的经验积累,在计算方法上已经能够涵盖锂离子电池的动力学和倍率性能^[7]、结构稳定性和循环性能^[8]、表面稳定性^[9]等各个方面。从材料物性计算软件角度来说,基于密度泛函理论(DFT, Density Functional Theory)的第一性原理计算软件目前已经比较成熟,已有大量实践表明这类软件对材料基本物理和化学性质的计算精度完全可以满足要求。其中最典型的代表就是 VASP (Vienna ab initio simulation package) 软件包^[10],该软件基于密度泛函理论及缀加平面波(PAW, projector augmented-wave) 赝势法^[11],可以正确求解并描述各类固体材料中的电子运动方程,进而获得材料的基态能量、电子能态、电荷密度和分布等信息。根据这些信息和一定的物理和化学关系,从理论上可以获得材料的机械、热力学、电磁学、光学等基本物理化学性质。基于上述原因,目前高通量的材料计算仅仅是一个系统集成的问题,在材料基因工程计划中不难解决。第 2 个材料性能测试系统是实验测试系统的建立,是一个难题,难以在短期内解决。第 3 个问题是材料性能数据库和平台的建立。锂离子电池虽然已经问世 30 多年,但对各类电池材料的基本物理化学性能数据库还没有建立。实验上虽然已经有了大量的测量,但由于成本高,且大量科研工作者的数据无法统一到一个数据库平台。因此,利用高通量计算技术结合第一性原理精确计算方法,可以比较方便和系统地获得各类已知的锂离子电池材料的性能数据,同时也可以获得具有潜力成为锂离子电池材料的物质的性质。

数据挖掘技术是一个新兴的、潜力巨大的研究领域,涉及的学科领域和方法较多,例如应用于数据库管理系统(DBMS)中用于语义查询优化、完整性约束和不一致检测等^[12]。随着海量数据和大数据时代的来临,数据挖掘技术也在日益普及,该技术也日益渗透到其他应用领域,包括材料科学、生物医学、金融分析、通信电子等领域^[13]。随着锂离子电池材料基因工程方法研究的开展,数据挖掘技术在构建锂离子电池材料构效关系上将起到关键作用。该领域的应用是关于材料性质数据库的数据挖掘,先从数据库中生成规则,然后由锂离子电池材料领域专

家进行验证,验证后规则就可加入知识库中,从而达到知识发现的效果。

本文基于材料基因工程的思路,选择锂离子电池材料为示范,利用无机材料晶体结构数据库(ICSD, the inorganic crystalline structure database)中的晶体结构数据为源数据,并连接到高性能并行计算中心计算材料的各种性能数据,建立了一个高通量计算平台,通过该平台来计算并生成比较完备的锂离子电池材料结构和性能数据库。利用该数据库,通过数据挖掘技术来探索锂离子电池材料的构效关系,为锂离子电池材料设计提供技术支撑。

1 数据生成和数据库结构

1.1 无机材料晶体结构数据库

利用材料基因工程思路来设计锂离子电池材料,首先需要建立锂离子电池材料的结构和性能数据库。这里包括已知的锂离子电池材料和未知的可能用于锂离子电池的材料。首先需要大量相关材料的结构源数据,而源数据最方便的来源是晶体材料数据库。目前国际上有几个著名的晶体学数据中心,这些数据中心的主要作用之一是收集、储存和提供已知化合物的晶体结构数据。其中 2 个十分重要的数据库为:剑桥结构数据库(CSD, Cambridge structural database)和无机晶体结构数据库(ICSD)。本文的研究选择了无机晶体结构数据库,把该库中的所有含 Li 元素的材料和结构作为 1 级源数据,计算和分析这些结构的材料作为锂离子电池材料的可行性和优缺点。选定一些目标后,再进行原子替换、结构组合、原子掺杂等,进而生成更多的未知(ICSD 库中不存在)的结构和材料,形成 2 级源数据。同时,还可以选择和 Li 元素具有类似性质的含 Na、K 等材料的结构,将 Na 和 K 替换成 Li,生成 2 级源数据。2 级目标晶体结构数据的产生思想可参阅文献[1]。

晶体按其内部结构可分为 7 大晶系^[14]。已知晶体形态超过 4 万种,它们都是按 7 种结晶模式发育生长,即 7 大晶系:立方晶系、四方晶系、三方晶系、六方晶系、斜方晶系、单斜晶系、三斜晶系。无机晶体结构数据库(ICSD)自 1913 年开始出版,至今已包含近 10 万条化合物目录,是国际最权威的无机晶体结构数据库。该数据库包括信息有:晶体中的元素或者化学名和化学式、矿物名和相名称、晶胞参数、空间群、原子坐标、热参数、位置占位度、R 因子及有关文献等各种信息^[15]。

1.2 材料性能计算软件的输入和输出数据

目前国际上已经有了许多第一性原理计算程序代码,本文设计的平台选择了 VASP 程序包来实现。该程序包的输入文件有 4 个文本文件,且固定文件名为: INCAR, POSCAR, POTCAR 以及 KPOINTS。INCAR 文件包括了第一性原理计算过程中的各种计算参数的设置、精度控制、算法选择等。VASP 软件对所有参数都提供一个默认值,因此,设置这些参数的唯一目的是达到人为需要的精度或人为选择计算方法。KPOINTS 文件是提供第一性原理计算中需要的倒格子空间中的取样,与计算的精度直接相关。POSCAR 提供的是计算的晶胞的详细信息,包括晶格参数和每个原子的坐标。POTCAR 文件提供了第一性原理计算选择的赝势。VASP 程序包已经提供了各种元素的赝势原始文件, POTCAR 生成只需要根据 VASP 提供的原始文件对计算的材料中涉及的元素进行组合。生成上述 4 个文件,可以在明确计算需求后,直接由计算机完成,也可以由专业人员根据需求进行特殊设置。

VASP 软件的输出信息包括基态能量本征值(总能)、优化后的晶格和原子坐标、电子态密度、电荷密度分布、电子自旋及原子磁矩、波函数等基态性质。这些信息除波函数 WAVECAR 以二进制文件存储外,其他都储存在 OUTCAR、DOSCAR、OSCICAR、CONTCAR、CHGCAR 等固定格式的文本文件中。通过提取这些文件中的有效信息,并做简单处理,可以获得计算所需要的材料的基本物理和化学性质。由于文件都是以固定格式存储,这些工作可以由计算机来完成。

1.3 材料结构和性能数据库生成的平台设计

整个锂离子电池材料结构和性能数据库生成的硬件平台主要包括一个 Web 服务器兼数据库存储器以及一个并行计算机中心。Web 服务器满足客户从 Internet 网络终端查询数据库中的数据 and 输送计算任务,同时,整个锂离子电池材料结构和性能数据库也存储在服务器上。而并行计算中心可以是自建的并行计算 Cluster,也可以是超算中心的一个账户。平台设计的框架如图 1 所示。

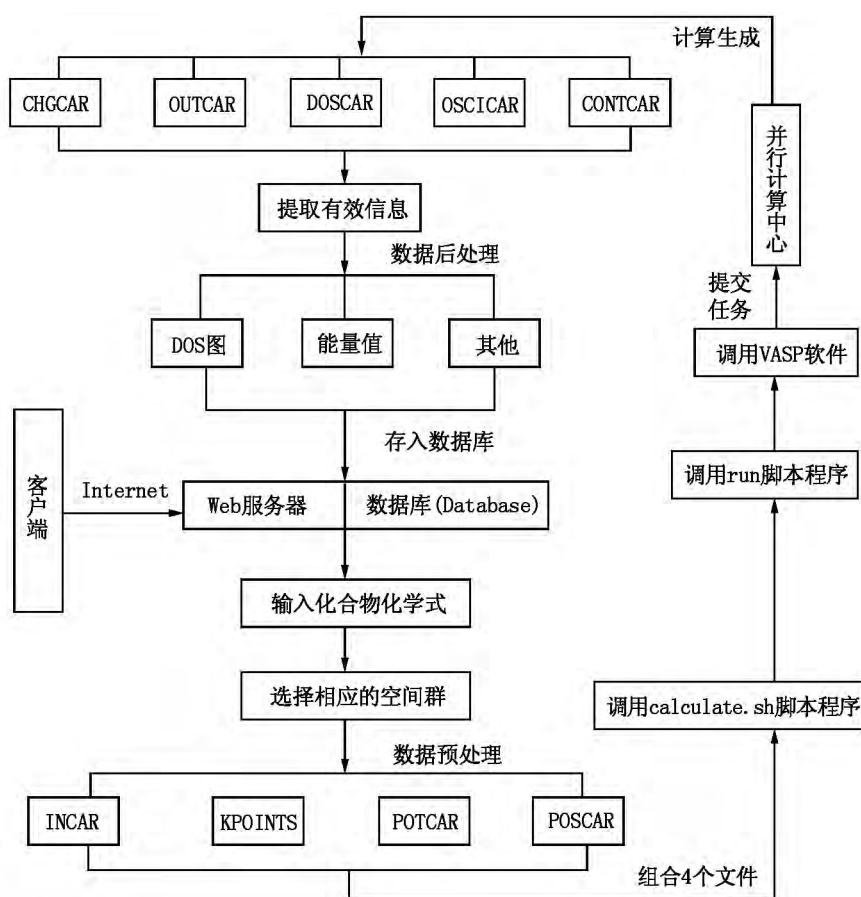


图1 材料结构和性能数据库生成的平台设计框架图

从图1可以看出,该平台工作基本原理是基于用户从 Internet 终端上访问 Web 服务器,并通过输

入化学式查询相应某种材料的性质。如果数据库中已经有了该材料的数据,则返回数据到用户终端。如

果数据库中不包含该材料数据,则从终端输送计算该材料性质的指令,服务器可自动完成材料计算的输入文件的产生,也可根据用户需求对参数做相应修改,然后输入文件数据传递给并行计算中心,并行计算中心自动启动计算任务,计算结束后,收集并整理相应数据返回到 Web 服务器并按相应格式储存在数据库中,并同时反馈给用户。

整个平台对材料性能数据库生成主要有 2 个部分:输入文件数据预处理和输出文件数据后处理。数据预处理主要从 ICSD 库中导入晶体结构对应的 cif 文件,并从 cif 文件中提取材料的化学计量比、晶胞参数、原子坐标等信息,进而根据晶胞参数和原子坐标生成 POSCAR 文件,根据化学计量比生成 POTCAR 文件,根据所需精度和特殊需求生成 INCAR 和 KPOINTS 文件。数据后处理包括计算出提取出能量值,进而计算平均电压,提取电子态密度信息,进而做出态密度图,提取结构优化后的原子坐标和晶胞参数等信息,和 ICSD 库中 CIF 文件进行比较等。

数据库采用固定的格式,通过上述基于第一性原理计算的平台,可以系统、有序、高效地把材料化合物对应的物性参数结合计算机语言完整地罗列出来,构建数据完整、质量高、可靠性强、可持续发展的储能材料与高性能结构材料数据库,构建基础数据子库、过程数据子库、以及性能数据子库。下表以锂离子电池正极材料 LiMn_2O_4 为例,给出了计算得到的基本数据信息,其中结构数据和 ICSD 库中的格式一致。

表 1 LiMn_2O_4 计算参数设置和计算数据结果,晶格常数 a b c 对应小括号内为低温下 LiMn_2O_4 实验数据

计算参数	赝势	PAW
	切断能	520 eV
	K-网格	$3 \times 3 \times 3$
	交换关联	GGA + U
晶格 ^[16]	a	0.815 9(0.819 9) nm
	b	0.825 8(0.824 8) nm
	c	0.833 1(0.828 0) nm
	$\alpha = \beta = \gamma$	90°
	晶系	正交晶系
计算结果	原子个数	56
	平均嵌 Li 电压	3.92 V
	嵌 Li 容量	$148 \text{ mAh} \cdot \text{g}^{-1}$
	禁带宽度	0.5 eV
	Mn 原子磁矩	$3.0/3.8 \mu_B$

2 数据挖掘与构效关系的建立

探索和挖掘锂离子电池材料的构效关系,首先

要明确锂离子电池材料必须或希望具备的性质,这些性能决定了锂离子电池最终表现出来的优良性能。下一代锂离子电池的目标是具有更高的能量密度、更好的倍率性能、更优的循环性能、更好的安全性能等。因此,挖掘锂离子电池材料的构效关系可以围绕这几个方面来进行。比如,为了提高电池的倍率性能,希望锂离子在电池材料中的扩散快,那么,锂离子在怎样的电池材料中扩散快这个问题,可以通过大量计算锂离子在不同结构的材料中锂离子的迁移数据,通过数据挖掘技术,找出锂离子迁移材料相关的结构、组分等的关系。

研究构效关系在材料科学中是一个重要的基础理论研究课题。晶体材料的性能受结构的支配,与此同时,晶体材料的组分又会影响相关材料的结构。晶体结构和组分因素主要包括:晶体对称性、结构基元间键合类型、物质类别、晶体结构缺陷、原子之间的计量比等。在晶体组成、结构与性能三者关系中,结构起到呈上启下的作用,因此,晶体材料的结构是关键因素。

本文通过建立一个锂离子电池材料性能和结构数据库,对该数据库进行动态管理,通过数据挖掘技术,来研究锂离子电池材料的构效关系。数据挖掘就是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中抽取和精化新的模式或知识,挖掘数据中潜在有用的信息的过程^[17]。近年来,数据挖掘领域的研究相当活跃,无论在理论上,还是在实用技术上都取得了可观的成果。数据挖掘在物理化学领域中的应用主要集中在检索和处理原始数据,而本文所做的关于二代锂离子电池新材料方面的探索研究中涉及到的数据处理就是一个典型的数据挖掘问题,无论是从大量各种锂离子化合物对应的晶体结构中获取相关参数信息还是从计算生成的大量数据中,包含的海量数据提取有价值的物性参数都是一个极其复杂繁琐的数据抽取和挖掘过程。通过收集的海量数据库辅以数据挖掘技术,实现高性能材料的预测和设计。本文设计的一个简单的数据挖掘模式如图 2 所示,通过图 2 可以认识到数据挖掘是一个知识获取的过程,用户从提交需求开始,整个系列数据挖掘的过程就是对数据不断进行预处理后处理以及优化的过程,其中涉及到的如数据定义、挖掘算法、模式筛选等。

从图 2 可以看出,整个数据挖掘过程的目标是锂离子电池材料的性能和结构之间的关系,而基础是大量正确可靠的数据来源。通过文中前面介绍的

通过第一性原理计算可以获得大量的锂离子电池材料的相关数据,再根据探索和设计下一代锂离子电池材料的目标导向,设定目标,通过图 2 给出的模

式,可以实现整个数据挖掘过程.具体的工作,目前还不能进行深入的讨论,需等待数据生成工作完成之后进行.

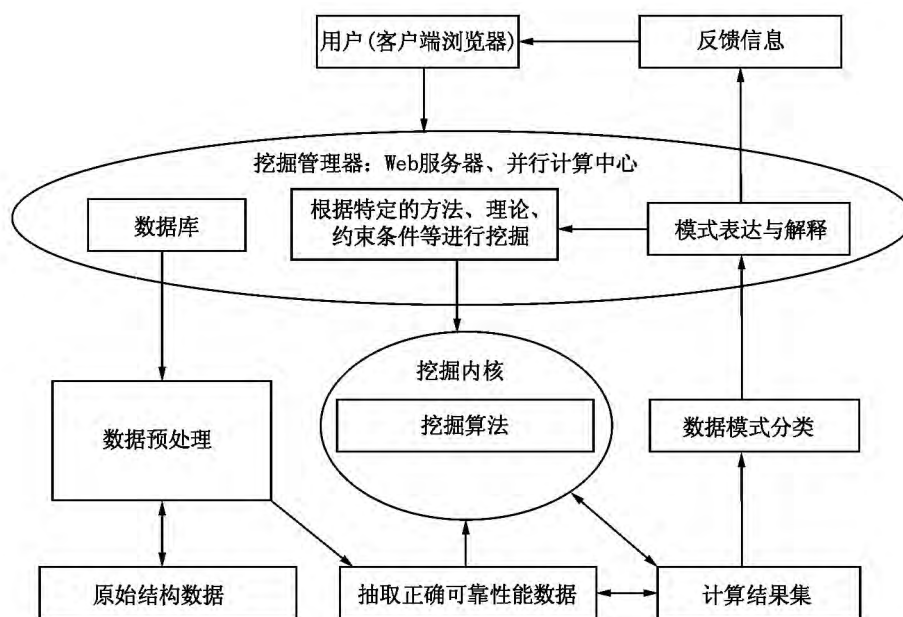


图 2 数据挖掘模式图

通过数据挖掘技术,从大量锂离子电池材料的性质中总结出来了什么样的结构和组分材料将具备哪些良好的性能.然后,将依据这些结构,探索设计出新型的锂离子电池材料.通过高通量计算,可以合理大胆去尝试和创新大量材料化合物结构并预测其性质,筛选和设计新材料,实现通过新的思维模式来挖掘新材料,缩短材料探索周期,从而加速材料研究的创新.

3 结论

锂离子电池的发展目前面临找瓶颈阶段,下一代锂离子电池的开发成败依赖于锂离子电池材料设计和创新.应用材料基因工程方法的基本思路,可以加速新材料开发与创新的步伐.本文基于该基本思路,设计了一个锂离子电池材料结构和性能数据库的生成方法和平台.该平台以无机材料晶体结构数据库为源数据,通过高通量第一性原理计算技术,可建立一个基于 Web 的可动态管理的锂离子电池材料结构和性能数据库.作为一个材料基因工程设计新材料的示范,本文设计的平台,通过构建一定的数据挖掘模式和方法,设定一定的数据生成和提取规则,可用于发现锂离子电池材料的构效关系,实现从

单一目标的计算材料学向材料基因组工程的质的转变.

4 参考文献

- [1] Wu M S, Xiao Jianmao, Xu B, et al. Use of materials genome initiative in lithium secondary batteries materials design [J]. J Chin Ceram Soc 2014(1): 42.
- [2] Obama. By the numbers: \$ 774 billion [EB/OL]. [2014-11-14]. <http://www.whitehouse.gov/blog/2011/11/14/numbers-774-billion>.
- [3] 陈立泉. 材料科学系统工程 [EB/OL]. [2011-12-30]. <http://www.xssc.ac.cn/ReadBrief.aspx?ItemID=968>.
- [4] Tarascon J M, Armand M. Issues and challenges facing rechargeable lithium batteries [J]. Nature 2001, 414: 359.
- [5] Ouyang Chuying, Chen L Q. Physics towards next generation Li secondary batteries materials: a short review from computational materials design perspective [J]. Science China, Physics, Mechanics & Astronomy, 2013, 56(12): 2278.
- [6] Zhou F, Cococcioni M, Marianetti C A, et al. First-principles prediction of redox potentials in transition-metal compounds with LDA + U [J]. Phys Rev B, 2004, 70: 235121.
- [7] 胡秀霞, 欧阳楚英. 离子迁移动力学的计算材料学方法

- 概述及其在锂离子电池材料中的应用 [J]. 江西师范大学学报: 自然科学版, 2013, 37(6): 551-556.
- [8] van de Walle A. A complete representation of structure-property relationships in crystals [J]. *Nat Mater*, 2008 (7): 455-458.
- [9] Ouyang Chuying, Zeng X M, Sljivancanin Z, et al. Oxidation states of Mn atoms at clean and Al_2O_3 -covered $\text{LiMn}_2\text{O}_4(001)$ surfaces [J]. *J Phys Chem C*, 2010, 114: 4756.
- [10] Kresse G, Furthmüller J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set [J]. *Phys Rev B*, 1996, 54: 11169.
- [11] Blöchl P E. Projector augmented-wave method [J]. *Phys Rev B*, 1994, 50: 17953.
- [12] 张维群. 数据挖掘研究和应用的现状与前景 [J]. 统计与信息论坛, 2004, 19(1): 95-96.
- [13] 林建勤, 林筑英. 数据挖掘与智能化信息处理研究 [J]. 贵州大学学报: 自然科学版, 2003, 20(3): 296-298.
- [14] 黄昆. 固体物理学 [M]. 北京: 高等教育出版社, 2004.
- [15] ICSD. Stage-Inorganic-Crystal-Structure-Database-ICSD [EB/OL]. [2014-09-21]. <http://sciencestage.com/inorganic-crystal-structure-database-icsd>.
- [16] Rodriguez-Carvajal J, Rousse G, Masquelier C, et al. Electronic crystallization in a lithium battery material: columnar ordering of electrons and holes in the spinel LiMn_2O_4 [J]. *Phys Rev Lett*, 1998, 81: 4660.
- [17] 龙腾芳. 数据挖掘技术在农业领域中的应用研究 [J]. 微计算机信息, 2005, 21(8): 42.

The Platform for Lithium Ion Battery Materials Data Calculation and Data Mining Based on Materials Genome Initiative

XIAO Jianmao¹, WANG Hao¹, OUYANG Chuying^{2*}

(1. School of Software Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China;

2. College of Physics and Communication Electronic, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: One important technique mean used to discover and design novel lithium ion battery materials is the materials genome initiative (MGI) method. A large amount of lithium ion battery materials basic physical and chemical properties can be generated through high-throughput calculations. The correlations between materials composition and structure and their physical and chemical properties can be analyzed with data mining techniques, and then can be used to guide the discovery of novel materials and designing of conventional materials for lithium ion battery. With the basic idea of the MGI, this paper takes the Inorganic Crystal Structure Database as source data, the input files of the first principles calculations software VASP are created through a Web server, and transferred to a parallel computing center. The materials properties data are then calculated and returned to the Web server. Through data mining technique, the correlations between lithium ion battery materials composition and structure and their physical and chemical properties can be summarized, which provides strong support to the design of the next generation lithium ion battery materials.

Key words: materials genome initiative; lithium ion batteries; data generation; data mining

(责任编辑: 冉小晓)