

化学工业与工程
Chemical Industry and Engineering
ISSN 1004-9533, CN 12-1102/TQ

《化学工业与工程》网络首发论文

题目：四种智能算法在相平衡数据拟合中的应用
作者：朱炜，刘斌，侯海云，李庆，王新元，樊增禄
DOI：10.13353/j.issn.1004.9533.20181012
收稿日期：2018-03-10
网络首发日期：2018-07-06
引用格式：朱炜，刘斌，侯海云，李庆，王新元，樊增禄. 四种智能算法在相平衡数据拟合中的应用. 化学工业与工程.
<https://doi.org/10.13353/j.issn.1004.9533.20181012>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

四种智能算法在相平衡数据拟合中的应用

朱 炜¹, 刘 斌¹, 侯海云¹, 李 庆¹, 王新元¹, 樊增禄^{2*}

(1. 西安工程大学环境与化学工程学院, 西安 710048; 2. 西安工程大学纺织与材料工程学院, 西安 710048)

摘要：针对 Marquardt-Levenberg 法应用于多元物系相平衡数据拟合时，模型参数剧增，初值难以设定的难题，将四种智能算法，即遗传算法、神经网络、退火算法及粒子群算法，应用于相平衡数据的拟合。以正丙醇（1）+乙腈（2）二元物系汽液相平衡数据的 Wilson 拟合和甲醇（1）+乙腈（2）+ 1-乙基-3-甲基咪唑四氟硼酸盐（3）三元物系汽液相平衡数据的 NRTL 拟合为例，系统讨论了四种算法在应用时的主要影响因素，并将所得结果进行了分析和比较，结果表明遗传算法和粒子群算法可以较好的解决初值难设的问题，并且给出了每种算法的适用范围和使用建议。

关键词：相平衡；参数估值；遗传算法；神经网络；退火算法；粒子群算法

中图分类号：TQ 013.1 **文献标志码：**A

Application of Four Intelligent Algorithms in Phase Equilibrium Data Fitting

Zhu Wei¹, Liu bin¹, Hou Haiyun¹, Li Qing¹, Wang Xinyuan¹, Fan Zenglu^{2*}

(1. School of Environmental and Chemical Engineering, Xi'an Polytechnic University, Xi'an 710048, China; 2. School of Textiles and Materials, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract : The Marquardt-Levenberg (ML) algorithm is the most commonly used algorithm for phase equilibrium data fitting. However, this algorithm belongs to the local optimization algorithm. When ML algorithm was applied to multi-component system phase equilibrium data fitting, it is difficult to find the appropriate initial values for the general thermodynamic researchers owing to model parameters increasing sharply. In this paper, four kinds of intelligent algorithms, namely, genetic algorithm, neural network, annealing algorithm and particle swarm algorithm were applied to fitting the vapor-liquid equilibrium data of n-propanol (1) + acetonitrile (2) binary system by Wilson model and fitting the vapor-liquid equilibrium data of methanol (1) + acetonitrile (2) + 1-ethyl-3-methylimidazolium tetrafluoroborate (3) ([EMIM][BF₄]) ternary system by NRTL model, respectively. The mainly influencing factors of four algorithms on the phase equilibrium data fitting application were discussed. The results were also analyzed and compared. On the basis of above work, the scope of the application and the use of recommendations of each method were proposed.

Keywords : phase equilibria; parameter estimation; genetic algorithm; neural networks; annealing algorithm; particle swarm algorithm

收稿日期：2018-03-10

基金项目：陕西省教育厅科学研究项目(17JK0348)；西安工程大学创新创业训练计划项目（2017086）；国家自然科学基金青年科学基金项目（2160030548）；陕西科技厅国际科技合作与交流计划项目（2017KW -026）。

作者简介：朱炜（1982-），男，博士，讲师，现从事分子热力学，量化计算和分子模拟方面的研究。

通信作者：樊增禄 E-mail: zhuwei@xpu.edu.cn。

相平衡研究中一项很重要的任务是将所获得的有限相平衡实验数据选择合适的热力学模型关联起来,即所谓的相平衡数据拟合。这样在实际工程应用中就可直接应用拟合模型来计算^[1]。通常相平衡数据拟合使用最多的方法是 Marquardt-Levenberg (ML) 法^[2-10]。ML 法属于局部优化算法,当用其进行多元物系相平衡数据拟合时,由于模型参数剧增,初值的选取极其重要。合适的初值不但结果好,而且计算量小。但对一般热力学研究人员而言,初值只能凭经验获取或随机猜测,因而所得结果往往都不理想。

人工智能(Artificial Intelligence, AI)^[11]最早提出于 1956 年,它是自然科学与社会科学的交叉。AI 简单来说就是“让计算机能像人一样的去思考和解决问题”。AI 领域中常用到的算法有遗传算法(Genetic Algorithm, GA)^[12],人工神经网络(Artificial Neural Network, ANN)^[13],退火算法(Simulate Anneal Arithmetic, SA)^[14]和粒子群算法(Particle Swarm Optimization, PSO)^[15]等。由于以上所提到的四种人工智能算法理论上均可进行全局优化,因此可将其应用于相平衡数据拟合中。郭宁等用 BP(Back Propagation, 反向传播) ANN 计算了乙醇-环己烷-水体系的汽液平衡^[16]和脂肪酸酯与超临界 CO₂ 体系的相平衡^[17];何学忠等^[18]使用改进的 BP ANN 对大分子体系的相平衡进行预测;林金清等^[19]将 GA 用于含化学反应体系的相平衡计算;安维中等^[20]将遗传算法用于多相多组分化学反应平衡和相平衡的计算;Senol^[21]等采用遗传算法对水+甲酸+酯/醇三元物系进行了 UNIFAC, SERLAS, NRTL, 改进 NRTL, 三参数 Margules 模型关联;Khansary 等^[22]将 GA 和 PSO 用于多元体系液液相平衡作用相互参数的估计;Mirarab 等^[23]将基于 GA 的 ANN 用于 CO₂ 在乙醇+[EMIM][Tf₂N]离子液体溶解度的预测;Li^[24]等将基于自适应 PSO 和混沌理论的 ANN 用于聚合物中气体溶解度的预测。

相平衡研究中的热力学模型包括经验型^[25](如 Margules, Van Laar, Wohl), 半经验型^[25](如 Wilson, NRTL, UNIQUAC) 和预测型(如 UNIFAC^[25], COSMO-RS^[26], COSMO-SAC^[27]) 三大类。其中,经验型只适合于二元物系,不能用于多元物系;基于基团贡献的 UNIFAC 虽然预测效果较好,但必须要有相关物系的基团参数;基于量子力学和统计力学的 COSMO-RS, COSMO-SAC 一般只能给出定性结果,而且对于大分子或复杂分子其准确的量化计算极其耗时^[28, 29]。因此,本文选用半经验模型来进行相平衡数据

拟合。

为了探讨智能算法能否解决多元物系相平衡数据拟合难的问题,本文将正丙醇(1)+乙腈(2)二元物系和甲醇(1)+乙腈(2)+1-乙基-3-甲基四氟硼酸盐([EMIM][BF₄])(3)三元物系汽液相平衡数据,分别利用 ML 法,遗传算法、神经网络、退火算法和粒子群算法进行了数据拟合,对四者的拟合效果进行了全面的分析和比较,并给出了每种算法的使用建议。

1 计算细节

1.1 相平衡数据

相平衡类型及对应的相平衡数据有多种,但不论那种相平衡一旦选中某种半经验型模型,除了相平衡数据预处理过程不同,用该模型拟合的核心过程都是相同的。所以,本文用汽液相平衡数据拟合代表所有相平衡数据的拟合,用二元汽液相平衡数据拟合代表少参数半经验模型的拟合,用三元汽液相平衡数据拟合代表多参数半经验模型的拟合。其中二元和三元汽液相平衡数据分别来自于文献 30 和 31。二元物系共有 15 个实验数据点,使用 Wilson 半经验模型进行拟合,该模型含 2 个参数 Λ_{12} 和 Λ_{21} (Λ_{ij} 为 Wilson 二元交互作用能量参数);三元物系共有 30 个实验数据点,使用 NRTL 半经验模型进行拟合,模型中含 6 个参数,分别是 α_{31}, α_{23} (α_{ij} 为系统混合非随机特征参数), $\Delta g_{13}, \Delta g_{31}, \Delta g_{23}, \Delta g_{32}$ (Δg_{ij} 为 NRTL 二元交互作用能量参数)。

1.2 计算平台及编程环境

Windows 7, 64 位; CPU: Intel i7-4790, 3.60 GHz, RAM: 16 GB, 8 核; Matlab R2010b。

1.3 目标函数

二元物系及三元物系相平衡数据拟合时采用式(1)所定义的目标函数(ARD)。

$$ARD(\%) = \frac{1}{n} \sum_n \left| \frac{\gamma_i^{\text{exp}} - \gamma_i^{\text{cal}}}{\gamma_i^{\text{exp}}} \right| \times 100 \quad (1)$$

上式中 γ_i^{exp} , γ_i^{cal} 分别是组分 i 的实验测定活度系数和计算活度系数, n 是实验数据点数。

1.4 比较标准

以 ML 法计算结果作为不同智能算法间比较的标准,详情见表 1。

表 1 ML 法拟合时的目标函数值及耗时
Table 1 ARDs and time consuming obtained by ML method

物系(模型)	ARD	时间
二元物系 (Wilson)	0.190	0.2 s
三元物系 (NRTL)	3.300	1 hr

1.5 默认参数

1.5.1 遗传算法

种群初始化方法为带约束的随机产生，选择方法为随机均匀分布，再生方法的精英数目和交叉分数分别为 2%和 80%，变异方法为均匀变异，交叉方法为散点交叉，迁移方法为向前，其分数和间隔值分别是 20%和 20，停止准则为种群最大代数(5000)或最小允许误差（1e-6）。

1.5.2 神经网络

使用具有 S 型隐形神经元和线性输出神经元的两层前馈神经网络，采用 LM 反向传播算法对网络进行训练，隐形神经元数目有 10 个，输入参数是液相组成和温度，输出为气相组成，训练集的选择采用随机方式。网络结构如图 1 所示。

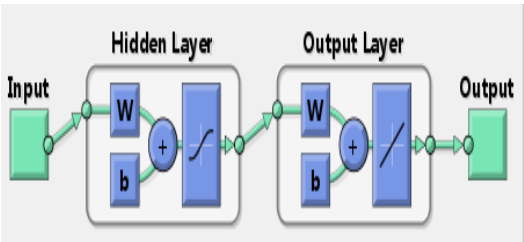


图 1 双层前馈神经网络

Fig. 1 the two-layer feed-forward network

1.5.3 退火算法

最大允许误差设为 1e-6，每次迭代的最大函数个数为（变量数*3000），初始温度设为 100，再退火间隔设为 100，退火函数选择 annealingFast，温度函数选择 temperatureExp，变量类型采用双精度型。

1.5.4 粒子群算法

最大迭代次数为 2000，初始化种子数为 24，局部最优加速度参数和全局最优加速度参数均为 2，初始时刻加权值和收敛时刻加权值分别为 0.9 和 0.4，算法终值阈值设为 1e-25,终止算法的阈值为 250，种子为随机产生。

2 结果与讨论

2.1 遗传算法

将遗传算法分别应用于二元和三元物系的汽液相平衡数据拟合，所得结果如下：

2.1.1 二元物系的拟合结果

表 2 自定义种群数和种群数对于拟合的影响

Table 2 Effects of initial population and population size on the fitting results

No.	自定义种群数	种群数	ARD	时间
1	—	10 ³	1.112	3.1 min
2	441	10 ³	1.113	23.3 s

3	—	10 ⁴	1.112	26.7 min
4	1681	10 ⁴	1.112	2.2 min
5	6561	10 ⁴	1.111	33.9 s
6	—	10 ⁵	1.112	2.1 h
7	41209	10 ⁵	1.112	4.1 min
8	—	10 ⁶	1.112	34.8 min
9	167281	10 ⁶	1.112	38.4 min

对于二元物系的拟合由表 2 可知：自定义种群的存在可极大减少计算量（见编号 2，4，5，7），但当种群数很大时，自定义种群的这种影响就不再明显（见编号 8 和 9）；所得结果均优于 ML 法，但是耗时要比 ML 法多，尤其是种群数很大时。

2.1.2 三元物系的拟合结果

表 3 自定义种群数和种群数对于拟合的影响

Table 3 Effects of initial population and population size on the fitting results

No.	自定义种群数	种群数	ARD	时间
1	—	10 ⁵	7.093	35.0 min
2	—	10 ⁵	6.916	19.7 min
3	—	10 ⁵	4.737	8.2 min
4	—	10 ⁵	4.730	7.7 min
5	—	10 ³	9.505	48.0 s
6	—	10 ⁴	9.510	48.0 s
7	—	10 ⁵	6.990	7.5 min
8	—	10 ⁶	4.410	73.0 min
9	—	10 ⁴	9.447	52.8 s
10	2304	10 ⁴	4.620	50.2 s
11	—	10 ⁵	4.916	8.4 min
12	15625	10 ⁵	4.384	503.3 s
13	—	10 ⁶	4.526	86.2 min
14	117649	10 ⁶	4.143	82.3 min

对于三元物系的拟合由表 3 可知：随种群数增加，ARD 值减小很快，相应计算时间却急剧增加（见编号 5-8）；自定义种群在种群数较小时对 ARD 值的影响很大，随着种群数的增加，自定义种群的这种影响减弱（见编号 9-14）；当种群数为 10⁶时才能取得与 ML 法相近的结果，而且计算时间要比 ML 法长。

此外，也对遗传算法的重现性和相关算法函数的影响进行了考察，详见表 4 和表 5。

表 4 遗传算法的重现性
Table 4 the reproducibility of genetic algorithm

No.	CPU _s	Population size	ARD	Time
1	4	10 ⁵	4.737	8.2 min
2	4	10 ⁵	4.741	7.8 min
3	4	10 ⁵	4.710	7.3 min
4	6	10 ⁵	7.302	12.8 min
5	6	10 ⁵	5.066	9.1 min
6	6	10 ⁵	5.864	7.3 min

由表 4 可知, 由于算法中引入随机数的缘故, 遗传算法的重现性不太稳定。

表 5 交叉函数, 选择函数, 变异函数, 迁移方向
对于拟合的影响

Table 5 the effects of crossover function, selection function, and mutation function and migration direction on the fitting results

序号.	函数	ARD	时间
1	crossoverheuristic	4.173	8.6 min
2	crossoverintermediate	4.371	8.7 min
3	crossoversinglepoint	4.340	8.9 min
4	crossovertwopoint	4.602	11.0 min
5	crossoverarithmetic	4.497	7.2 min
6	selectionremainder	4.177	1.6 h
7	selectionuniform	5.053	7.8 min
8	selectionroulette	4.169	1.1 h
9	selectiontournament	4.607	8.7 min
10	mutationadaptfeasible	4.343	8.4 min
11	both	4.178	12.2 min

将表 5 的结果与表 3 中编号 11 相比可见: 不同交叉函数对 ARD 值和耗时影响较小 (见编号 1-5), 其中 crossoverheuristic 结果最好; 选择函数对 ARD 值有一定影响, 虽然 selectionremainder 和 selectionroulette 结果较好, 但耗时都较多 (见编号 6-9); 变异函数对 ARD 和耗时也影响较小 (见编号 10); 选择函数的 both 方向可减小 ARD 值, 且耗时短。

2.1.3 建议

要增加遗传算法的拟合效果, 增加种群数比增加代数更有效, 但计算时间会相应增加; 若粗略知道初值的可能范围, 可使用自定义种群来提高计算效率; 遗传算法使用时最好多算几次再取最优值; 遗传算法的交叉函数, 选择函数等对

ARD 值的影响不大, 计算时不必过于计较。

2.2 神经网络

由于神经网络对二元物系和三元物系的拟合过程和结论相似, 因此只选三元物系的结果进行讨论。该算法的每个结果都是进行了六次训练筛选而得。

2.2.1 三元物系的拟合结果

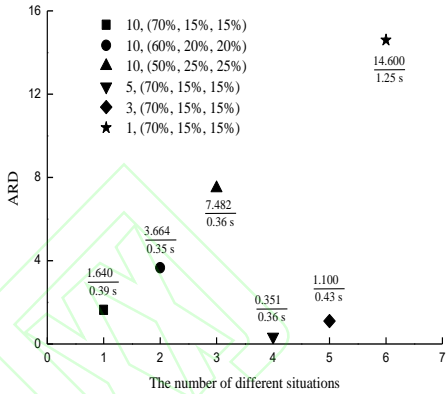


图 2 27 个数据点时神经元数和训练比例 (训练分数, 验证分数, 预测分数) 对于拟合的影响

Fig.2 Effect of number of neurons and training proportion (training, validation, testing) on the fitting results using 27 data points

首先, 采用 27 个数据点对神经网络进行训练, 验证和预测, 结果见图 2。图中短线之上数字表示该次计算的 ARD 值, 短线之下数字表示计算耗时, 下同。由图 2 可知, 当输入数据为 3*27, 神经元数为 10, 预测分数为 20% 时, 所得结果与 ML 法相近; 在训练分数为 70%, 验证分数和预测分数均为 15% 时, 3 个神经元即可得到比 ML 法更好的结果。但此时预测分数只有 15% ~ 20%。

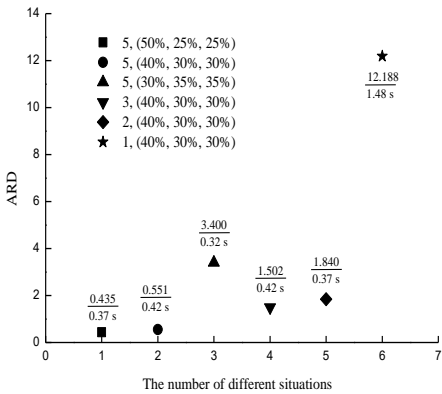


图 3 51 个数据点时神经元数和训练比例 (训练分数, 验证分数, 预测分数) 对于拟合的影响

Fig.3 Effect of number of neurons and training proportion (training, validation, testing) on the fitting results using 51 data points

接着,采用 51 个数据点对神经网络进行训练,验证和预测,结果见图 3。由图 3 可知,当输入数据为 3×51 ,神经元数为 5,预测分数为 35% 时,所得结果与 ML 法相近;在训练分数为 40%,验证分数和预测分数均为 30% 时,2 个神经元即可得到比 ML 法更好的结果。此时预测分数已提高到 30% ~ 35%。为进一步提高网络的预测能力,继续尝试增加数据点。

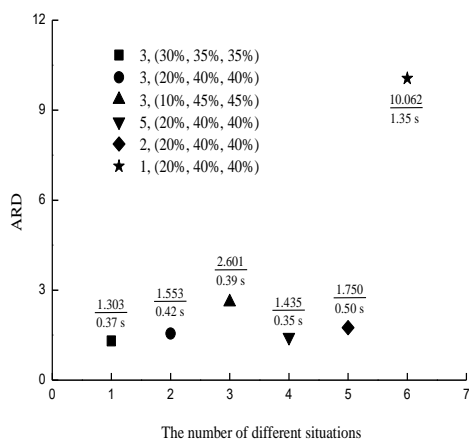


图 4 123 个数据点时神经元数和训练比例(训练分数,验证分数,预测分数)对于拟合的影响

Fig.4 Effect of number of neurons and training proportion (training, validation, testing) on the fitting results using 123 data points

最后,采用 123 个数据点对神经网络进行训练,验证和预测,结果见图 4。由图 4 可知,当输入数据为 3×123 ,神经元数为 3,预测分数为 45% 时,所得结果已优于 ML 法;甚至在训练分数为 20%,验证分数和预测分数均为 40% 时,2 个神经元便可得到比 ML 法更好的结果。此时神经网络只需要非常少的训练集(10% ~ 20%)和神经元(2 或 3)就能拥有很强的预测能力(40% ~ 45%)。

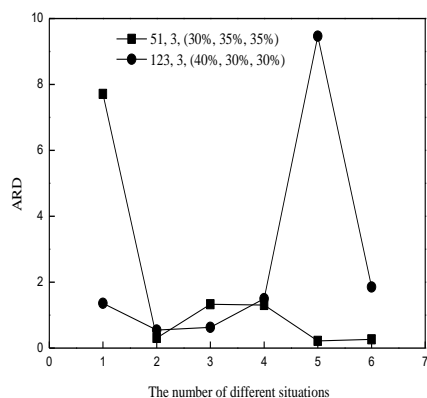


图 5 神经网络的重现性

Fig.5 the reproducibility of neural network

对神经网络重现性的考察结果见图 5。图中图例的数字分别表示输入数据数,神经元数和训练比例(训练分数,验证分数,预测分数)。由图 5 可知,神经网络的重现性很差。这也是本节每个结果都是从六次训练进行筛选的原因。

2.2.2 建议

神经网络输入数据越多,训练所得网络预测能力越强;数据点有限时,若要达到好的预测效果,训练集中必须包含相图中尽可能多的拐点;训练集精度越高,网络预测性能并非越高,其关键取决于训练集中拐点的多少;神经网络若要重现性好,训练集的选择就不可采用随机的方式。

2.3 退火算法

将退火算法分别应用于二元和三元物系的汽液相平衡数据拟合,结果如下:

2.3.1 二元物系的拟合结果

表 6 ML 法和退火算法对于两参数拟合的比较

Table 6 Comparison of ML and annealing algorithms for two-parameter fitting

序号	算法	初值	次数	ARD	时间
1	SA	[1, 1]	5914	0.194	3.2 s
2	ML		42	0.194	0.2 s
3	SA	[10000,	5912	5.819	2.7 s
4	ML	10000]	14	0.194	0.2 s
5	SA	[-10000,	5914	43.522	2.8 s
6	ML	-10000]	16	0.194	0.2 s

对于二元物系拟合由表 6 可知,采用同样的初始值,ML 法要比退火算法效率高很多。

2.3.2 三元物系的拟合结果

表 7 ML 法和退火算法拟合的比较

Table 7 Comparison of ML and annealing algorithms for six-parameter fitting

初值	算法	次数	ARD	时间
[1,1,1,1,0.5,0.5]	SA	5567	9.704	4.2 s
	ML	401	10.039	1.2 s
[-30000,30000,-1000,1000,0.3,0.8]	SA	5067	34.073	3.8 s
	ML	26	3.307	0.4 s
[-10000,10000,-10000,10000,0.5,0.5]	SA	16060	13.179	12.3 s
	ML	401	12.494	1.0 s
[10000, 10000, 10000, 10000, 0.5,0.5]	SA	5544	82.256	3.5 s
	ML	44	15.218	0.3 s

					results				
					序号	初值	最大函数个数	ARD	时间
[-30000,30000, -1000, 1000, 0.3,0.8]	SA	8040	137.746	6.4 s	1		5×10^4	4.756	5.1 s
	ML	23	63.124	0.3 s	2		10^5	4.728	5.2 s
[1000,1000, 1000, 1000, 0.9,0.1]	SA	3055	47.514	2.1 s	3		5×10^5	4.703	5.9 s
	ML	74	3.307	0.3 s	4	[100000,	10^6	4.821	2.3 s
[100000,-1000, 10000, -10000, 0.9,0.1]	SA	7836	5.148	5.2 s	5	-1000,	10^6	4.565	10.0 s
	ML	96	4.052	0.4 s	6	10000,	10^6	4.844	6.4 s
					7	-10000,	10^7	4.770	2.2 s
					8	0.9,	10^7	4.730	6.6 s
					9	0.1]	10^7	4.689	6.6 s
					10		10^8	4.965	2.9 s
					11		10^8	4.835	4.1 s
					12		10^8	4.856	4.0 s

对于三元物系拟合由表 7 可知,采用相同初始值,ML 法的确比退火算法的效率很高,而且初始值的适用范围更宽泛,但结果偶尔会稍差一点。

同时,也对退火算法的重现性和相关算法函数的影响进行了考察,结果见表 8-10。

表 8 初始温度对于拟合的影响及退火算法的重现性
Table 8 Effect of initial temperature on the fitting results and the reproducibility of annealing algorithm

序号	初值	初始温度	ARD	时间
1	[-30000,	10^2	34.073	3.8 s
2	30000,	10^3	27.289	10.6 s
3	-1000,	10^4	15.135	9.4 s
4	1000,	10^5	12.588	2.9 s
5	0.3, 0.8]	10^6	10.363	2.0 s
6		10^8	9.821	2.9 s
7	[100000,	10^2	6.132	2.9 s
8	-1000,	10^3	5.181	2.9 s
9	10000,	10^4	4.782	3.1 s
10	-10000,	10^5	10.470	2.7 s
11	0.9, 0.1]	10^5	46.603	2.1 s
12		10^5	11.384	4.9 s
13		10^5	9.526	2.3 s
14		10^6	12.707	2.1 s
15		10^8	10.666	2.3 s

由表 8 可知,初始温度对 ARD 值影响很大,通常初始温度越高结果越好(见编号 1-9),但也有例外(见编号 7-15),而且不同初始值对应不同的最优初始温度;由编号 10-13 可知,退火算法的重现性很差。

表 9 最大函数个数对于拟合的影响
Table 9 Effect of max function evaluations on the fitting

由表 9 可知,最大函数个数对结果和重现性几乎无影响,因此无需设置大的最大函数个数。

表 10 退火函数和温度函数对于拟合的影响
Table 10 Effect of annealing and temperature functions on the fitting results

序号	函数	迭代次数	ARD	时间
1	annealingboltz	3432	5.699	2.2 s
		7998	5.104	5.6 s
		6349	6.144	4.5 s
2	temperatureboltz	3773	9.641	2.6 s
		4422	8.928	3.1 s
		7204	5.290	5.4 s
3	temperaturefast	3196	7.782	2.2 s
		6352	5.253	4.4 s
		6045	11.963	4.1 s

由表 10 可知,退火函数对结果影响不大;温度函数对结果影响较大,但是好是坏不能确定,其原因归咎于退火算法重现性较差(见编号 2 和 3)。

2.3.3 建议

退火算法可看作半全局优化,因为初值对结果影响很大。理论上初始温度越高,结果越趋近于全局最优,但实践中该说法不一定正确。要得到好结果,初始值和最优初始温度都必须慢慢尝试,而实现该过程最好的方式是编程。

2.4 粒子群算法

将粒子群算法分别应用于二元和三元物系的汽液相平衡数据拟合,结果如下:

2.4.1 二元物系的拟合结果

表 11 种子数对拟合的影响

Table 11 Effect of population size on the fitting results

序号	种子数	进化次数	ARD	时间
1	20	2000 (1147)	0.194	5.1 s
2	40	2000 (1220)	0.194	6.9 s
3	80	300	0.194	4.5 s
4	160	300	0.194	5.1 s
5	1000	300	0.194	25.4 s
6	10000	300	0.194	3.2 min

二元物系的拟合结果见表 11。表中第三列括号中数字表示实际需要的进化次数，括号外数字表示计算时设定的进化次数，下同。由表 11 可知，通过少种子数多进化次数或多种子数少进化次数，粒子群算法均可取得与 ML 法相同的结果，但耗时要比 ML 法多。

2.4.2 三元物系的拟合结果

表 12 种子数，进化次数对拟合的影响及粒子群算法的重现性

Table 12 Effect of population size and epochs and the reproducibility of particle swarm algorithm

序号	种子数	进化次数	ARD	时间
1	20	10 ⁴ (3542)	9.558	29.5 s
2	20	10 ⁴ (1680)	14.129	13.9 s
3	20	10 ⁵ (2071)	8.959	16.9 s
4	20	10 ⁵ (1652)	9.028	13.5 s
5	20	10 ⁵ (1741)	9.023	14.4 s
6	40	10 ⁴ (2933)	8.959	36.6 s
7	80	10 ²	9.003	2.8 s
8	80	10 ³	8.959	21.6 s
9	80	10 ⁴ (1427)	9.515	30.4 s
10	160	10 ²	9.010	4.2 s
11	160	10 ³	8.974	41.0 s
12	160	10 ⁴ (1524)	8.959	61.3 s
13	320	10 ³	8.959	77.3 s
14	640	10 ³	8.959	2.5 min
15	1280	10 ³	8.959	4.8 min
16	10 ⁵	10 ³	4.124	7.6 h

三元物系的拟合结果见表 12。由表 12 可知：增加种子数和进化次数均可优化结果（见编号 1-16），且采用大幅增加种子数的方式结果更好，但耗时增加严重（见编号 16）；粒子群算法的重

现性较好（见编号 3-5）；当种子数为 10⁵ 时，可达到与 ML 法相近的结果，但耗时却是 ML 法的 7.6 倍。

2.4.3 建议

使用粒子群算法进行多参数拟合若要取得好结果，最好在参数限制条件内设置尽量大的种子数，并以运算时间剧增为代价。

2.5 四种算法的比较

2.5.1 效率

表 13 四种智能算法效率的比较

Table 13 Comparison of efficiencies of four intelligent algorithms

标准	GA	ANN	PSO	SA	ML
二元物系每种算法的最优结果					
ARD	1.111	0.181	0.194	0.194	0.190
时间	33.9 s	0.3 s	5.1 s	3.2 s	0.2 s
三元物系每种算法的最优结果					
ARD	4.169	0.435	4.124	4.501	3.300
时间	1.1 h	0.4 s	7.6 h	2.2 d	1.0 h

四种智能算法用于相平衡数据拟合时的效率比较见表 13。可见，对于二元物系四种算法的效率排序为：ANN > SA > PSO > GA；对于三元物系四种算法的效率排序为：ANN > GA > PSO > SA。

2.5.2 相同点

（1）智能算法中随机数的引入，一方面使算法有可能得到全局最优，另一方面也导致每次计算结果有差异。因而智能算法应用时必须多算几次，再选其中最优者。

（2）智能算法的优势：第一，容易实现，不涉及梯度矩阵的计算，不会出现不收敛等情况；第二，并行计算可极大提高计算速度（除神经网络）；第三，拟合时初值的选择在智能算法中不是难题，因为每种算法都有其专属的初值函数用以产生满足限定条件的初值（除神经网络）。

2.5.3 不同点

（1）粒子群算法没有交叉，选择和变异过程，比遗传算法更容易实现，但效率不如遗传算法。

（2）神经网络效率最高（耗时少，精度高），但所得模型参数较多且物理意义不明确，可将其用于各种热力学模型都难以描述或拟合效果很差的场合。

3 结 论

本文将四种智能算法引入相平衡数据拟合，通过对各算法具体应用时的分析和讨论，可得出以下几点结论：

（1）少参数拟合使用 ML 法经济；多参数拟合

推荐使用 GA 和 PSO; 若 ML, GA, PSO 和 SA 拟合效果均不理想, 可尝试采用 ANN 进行拟合。

(2) 理论上智能算法均可实现全局优化, 但实践中只能趋近于全局最优。若采用智能算法+局部优化的方式可获得更优结果。

(3) 拟合时将 ARD 的允许误差设置为 10^{-4} 为宜, 否则会因精度太高而浪费大量时间。

(4) 智能算法为多元相平衡数据拟合难问题提供了一种可行的解决方案。

参考文献:

- [1] Walas S M. Phase Equilibria in Chemical Engineering [M]. Oxford: Butterworth Publishers, 1985
- [2] Soujanya J, Satyavathi B, Sankarshana T. Isobaric ternary vapour-liquid equilibrium of methanol (1)+diisopropyl ether(2)+isopropyl alcohol(3) along with methanol+isopropyl alcohol binary data at atmospheric and sub-atmospheric pressures [J]. Fluid Phase Equilib, 2015, 405: 31-36.
- [3] Kamiński W, Górak A, Kubiczek A. Modeling of liquid-liquid equilibrium in the quinary system of water, acetone, n-butanol, ethanol, and ionic liquid [J]. Fluid Phase Equilib, 2014, 384: 114-121.
- [4] Lei Z, Xi X, Dai C, *et al.* Extractive Distillation with the Mixture of Ionic Liquid and Solid Inorganic Salt as Entrainers [J]. AIChE J, 2014, 60(8): 2994-3004.
- [5] Lei Z, Dai C, Liu X, *et al.* Extension of the UNIFAC Model for Ionic Liquids [J]. Ind Eng Chem Res, 2012, 51(37): 2135-2144.
- [6] Shao X, Ge H, Li Z, *et al.* Solubility of methylphosphonic acid in selected organic solvents [J]. Fluid Phase Equilib, 2015, 390: 7-13.
- [7] Pirouzi A, Nosrati M, Haghtalab A, *et al.* Experiment and correlation of osmotic coefficient for aqueous solution of carboxylic acids using NRTL nonrandom factor model [J]. Fluid Phase Equilib, 2012, 327: 38-44.
- [8] Ricardo F P, Moilton R F. Solubility of salicylic acid in water + salt (NaCl, KCl, NaBr, Na₂SO₄ and K₂SO₄) at 293.5-313.3K [J]. Fluid Phase Equilib., 2012, 330: 48-51.
- [9] Padaszyński K, Okuniewski M, Domańska U. Solid-liquid phase equilibria in binary mixtures of functionalized ionic liquids with sugar alcohols: New experimental data and modeling [J]. Fluid Phase Equilib, 2015, 403: 167-175.
- [10] Rogośić M, Sander A, Kojić V, *et al.* Liquid-liquid equilibria in the ternary and multicomponent systems involving hydrocarbons, thiophene or pyridine and ionic liquid (1-benzyl-3-methylimidazolium bis(trifluoromethyl sulfonyl)imide) [J]. Fluid Phase Equilib, 2016, 412: 39-50.
- [11] 温正. 精通 MATLAB 智能算法[M]. 北京: 清华大学出版社, 2015: 1-10
Wen Zheng. Proficient in MATLAB intelligent algorithm [M]. Beijing: Tsinghua University Press, 2015(in Chinese)
- [12] 雷英杰, 张善文. MATLAB 遗传算法工具箱及应用 (第二版) [M]. 西安: 西安电子科技大学出版社, 2014: 45-105.
Lei Yingjie, Zhang Shanwen. MATLAB Genetic Algorithm Toolbox and its Application (2nd) [M]. Xi'an: Xi'an University of Electronic Science and Technology Press, 2014(in Chinese).
- [13] 飞思科技产品研发中心. 神经网络理论与 MATLAB7 实现[M]. 北京: 电子工业出版社, 2005.
FEISI Technology Product R&D Center. Neural Network Theory and MATLAB7 Implementation [M]. Beijing: Electronic Industry Press, 2005(in Chinese).
- [14] 康立山, 谢云, 尤矢勇, 等. 非数值并行算法(第一册): 模拟退火算法[M]. 北京: 科学出版社, 1994.
Kang Lishan, Xie Yun, You Shiyong, *et al.* Non-numerical Parallel Algorithm (Volume 1): Simulated Annealing Algorithm [M]. Beijing: Science Press, 1994: (in Chinese).
- [15] 钱锋. 粒子群算法及其工业应用[M]. 北京: 科学出版社, 2013.
QIAN F. Particle Swarm Algorithm and its Industrial Application [M]. Beijing: Science Press, 2013(in Chinese).
- [16] 郭宁, 崔志芹, 云志, 等. BP 神经网络计算乙醇-环己烷-水体系汽-液平衡 [J]. 南京工业大学学报, 2002, 24(4): 91-93.
Guo Ning, Cui Zhingqing, Yun Zhi, *et al.* Vapor-liquid equilibrium data of ethanol-cyclohexane-water using BP artificial neural network[J]. Journal of Nanjing university of technology, 2002, 24(4): 91-93(in Chinese).
- [17] 郭宁, 云志, 邵荣, 等. 神经网络计算脂肪酸酯与超临界 CO₂ 体系相平衡[J]. 南京工业大学学报, 2003, 25(3): 56-61.
Guo Ning, Yun Zhi, Shao Rong, *et al.* Calculation of phase equilibrium of systems containing supercritical carbon dioxide and fatty acid esters using artificial neural networks[J]. Journal of Nanjing University of technology, 2003, 25(3): 56-61(in Chinese).
- [18] 何雪忠, 张香平, 张锁江, 等. 改进的 L-M 算法用于大分子体系相平衡的神经网络预测[J]. 化工学报, 2005, 56(3): 392-399.
He Xuezhong, Zhang Xiangping, Zhang Suojian, *et al.* Improved L-M algorithm for ANNs prediction of phase equilibrium in macromolecule system[J]. CIESC Journal, 2005, 56(3): 392-399(in Chinese).
- [19] 林金清, 李浩然, 韩世钧. 应用遗传算法求解含化学反应体系的相平衡[J]. 化工学报, 2002, 53(6): 616-620.
Lin Jinqing, Li Haoran, Han Shijun. Computation algorithm for simultaneous chemical and phase equilibrium by using genetic algorithm[J]. CIESC Journal, 2002, 53(6): 616-620(in Chinese).
- [20] 安维中, 胡仰栋, 袁希钢. 多相多组分化学反应平衡和相平衡计算的遗传算法[J]. 化工学报, 2003, 54(5): 691-694.
An Weizhong, Hu Yangdun, Yuan Xigang. Calculation of multiphase and multicomponent chemical equilibrium using genetic algorithm[J]. CIESC Journal, 2003, 54(5): 691-694(in Chinese).
- [21] Senol A, Bilgin M, Baslioglu B, *et al.* Modeling phase equilibria of ternary systems (water + formic acid + es-

- ter or alcohol) through UNIFAC-original, SERLAS, NRTL, NRTL-modified, and three-suffix Margules: Parameter estimation using genetic algorithm[J]. *Fluid Phase Equilib*, 2016, 429: 254–265.
- [22] Khansary M A, Sani A H. Using genetic algorithm (GA) and particle swarm optimization (PSO) methods for determination of interaction parameters in multicomponent systems of liquid–liquid equilibria[J]. *Fluid Phase Equilib*, 2014, 365: 141–145.
- [23] Mirarab M, Sharifi M, Ghayyem M, *et al.* Prediction of solubility of CO₂ in ethanol–[EMIM][Tf₂N] ionic liquid mixtures using artificial neural networks based on genetic algorithm[J]. *Fluid Phase Equilib*, 2014, 371: 6–14.
- [24] Li M, Huang X, Liu H, *et al.* Prediction of gas solubility in polymers by back propagation artificial neural network based on self-adaptive particle swarm optimization algorithm and chaos theory[J]. *Fluid Phase Equilib*, 2013, 356: 11–17.
- [25] Smith J M, Van Ness H C, Abbott M M. Introduction to Chemical Engineering Thermodynamics (7th)[M]. New York: McGraw Hill Higher Education, 2005.
- [26] Klamt A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena [J]. *J Phys Chem*, 1995, 99(7): 2224–2235.
- [27] Lin S, Sandler S I. COSMO-SAC: A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model [J]. *Ind Eng Chem Res*, 2002, 41(5): 899–913.
- [28] Lei Z, Chen B, Li C, *et al.* Predictive Molecular Thermodynamic Models for Liquid Solvents, Solid Salts, Polymers, and Ionic Liquids [J]. *Chem Rev*, 2008, 108(4): 1419–1455.
- [29] Xue Z, Mu T, Gmehling J. Comparison of the a Priori COSMO-RS models and group contribution methods: Original UNIFAC, Modified UNIFAC (Do), and Modified UNIFAC (Do) Consortium [J]. *Ind Eng Chem Res*, 2012, 51(36): 11809–11817.
- [30] Zhu W, Qing L, Liu B, *et al.* Effect of imidazolium-based ionic liquid on vapor-liquid equilibria of 2-propanol + acetonitrile binary system at 101.3 kPa [J]. *Fluid Phase Equilib*, 2016, 409: 383–387.
- [31] Li Q, Zhu W, Liu B, *et al.* Measurement and correlation of the vapor–liquid equilibrium for methanol + acetonitrile + imidazolium-based ionic liquids at 101.3 kPa [J]. *J Chem Thermodyn*, 2016, 101: 25–30.