The Prediction on Initial Discharge Capacity of AB₅-Based Alloy with Simulated Annealing

Liu Yang^{1,*} and Wu Feng^{1,2}

¹School of Chemical Engineering & Environment, Beijing Institute of Technology, Beijing, 100081, P. R. China

²National Development Center of Hi-Tech Green Materials, Beijing, 100081, P. R. China

The prediction for initial discharge capacity of AB₅-based alloy has tested by employing the simulated annealing method in artificial neural network (ANN). The parameters in cooling schedule were determined and their influences on network prediction performance were discussed. A practical cooling schedule was established and the network with better performances was obtained by associating simulated annealing algorithm with gradient descent algorithm.

(Received September 1, 2003; Accepted December 22, 2003)

Keywords: AB₅-based alloy, artificial neural network, simulated annealing, cooling schedule, gradient-descent

1. Introduction

Among hydrogen storage materials, AB₅-type alloys have attracted much attention. They are used successfully as the negative electrodes of Ni/MH batteries for many advantages in terms of performance and environmental protection. Systematic studies have been done on B-side with Ni substituted partly by Co, Mn, Al.^{1,2)} The method, PLS-BPN (Partial Least Squares-Back Propagation Network), has also been employed in studying the influence of B-side elements on electrochemical properties of alloys.³⁾ However, only gradient descent method was used in the above-mentioned studies.^{4,5)} Owing to the disadvantage of gradient descent algorithm, such as fixed search direction and being easy to trap in local optimum points, the prediction is not so accurate. In this paper, the simulated annealing algorithm is used, which is a method of global optimization and can search the global optimum points of the goal functions effectively.

2. Data Collection

The data are listed in Table 1. The initial discharge capacity (IDC) comes from practical measurement.⁴⁾ In all the samples, the composition of B-side are the same, *i.e.* Ni_{3.5}Co_{0.8}Mn_{0.4}Al_{0.3}, and A-side is misch metal consisting of La, Ce, Nd, Pr with different proportion.

3. Calculations

Artificial neural network⁶⁾ is a new technique for handling and fitting of data and is suitable for building nonlinear mathematical models affected by various factors. According to the Kolmogorov theorem, a three-layer BP network with adequate hidden nodes can approximate any continuous function in a closed interval. However, the optimization rate of the network with traditional learning rule is too slow and much iteration is needed in order to obtain the satisfactory result.

3.1 Simulated annealing algorithm

Simulated annealing algorithm is derived from physical cooling and annealing process where the particles in solids are out-of-order and the inner energy of solids is large at high temperature. The temperature does not reduce gradually until the particles become ordered and the minimum energy is obtained at ground state. According to Metropolis rule,⁷⁾ the probability P for the particles reaching equilibrium state at temperature T is

$$P = \exp\left(\frac{-\Delta E}{K \times T}\right) \quad (1)$$

Where ΔE represents the change of the energy E , and K is Boltzmann constant.

The aim for applying the simulated annealing algorithm to the neural network is to obtain the solution of minimizing $f(X)$, where f is the goal function (*i.e.* error function), and X is the weights and biases value. The structure of simulated annealing algorithm is illustrated schematically in Fig. 1, where i is the number of iterations. ΔX denotes the change of the weights and biases value and is generated generally from Gaussian distribution.

$$\Delta X \sim N\left(0, \sqrt{\frac{T}{\eta}}\right) \quad (2)$$

The condition to accept new solution is $f(Y) < f(X)$ and $P > \alpha$, with α subjected to the uniform distribution in the space interval from 0 to 1 and P given as follows.

$$P = \exp\left(\frac{-(f(Y) - f(X))}{T}\right) \quad (3)$$

3.2 Cooling schedule

The cooling schedule is the core of simulated annealing, which determines the searching accuracy and effectiveness, *etc.* The cooling schedule consists of the initial temperature, the annealing scheme, the final temperature and the length of Markov chain⁸⁾ that controls the iteration number at each temperature.

ΔX is required to be adequately large at the initial process

*Undergraduate Student, Beijing Institute of Technology

Table 1 Chemical composition of the A-side and initial discharge capacity of AB₅-based hydrogen storage alloy.

No.	Composition of A-side (at %)				C ₀ (mAh/g)
	La	Ce	Nd	Pr	
1	0.317	0.135	0.223	0.325	239.4
2	0.176	0.206	0.124	0.494	240.5
3	0.528	0.029	0.372	0.071	244.7
4	0.433	0.04	0.433	0.095	246.4
5	0.27	0.135	0.27	0.325	233.7
6	0.15	0.206	0.15	0.494	234.3
7	0.45	0.029	0.45	0.071	226.9
8	0.649	0.04	0.216	0.095	244.2
9	0.405	0.135	0.135	0.325	243.4
10	0.225	0.206	0.075	0.494	247.2
11	0.675	0.029	0.225	0.071	248.1
12	0.288	0.04	0.577	0.095	228
13	0.18	0.135	0.36	0.325	239.8
14	0.1	0.206	0.2	0.494	228.8
15	0.3	0.029	0.6	0.071	220.8
16	0.794	0.066	0.109	0.031	255.6
17	0.831	0.069	0.078	0.022	270.7
18	0.277	0.023	0.545	0.155	218.3
19	0.272	0.272	0.355	0.101	248.2
20	0.43	0.43	0.109	0.031	258
21	0.45	0.45	0.078	0.022	256.3
22	0.15	0.15	0.545	0.155	234.1
23	0.181	0.36	0.355	0.101	233.2
24	0.287	0.573	0.109	0.031	257.7
25	0.3	0.6	0.078	0.022	259.2
26	0.1	0.2	0.545	0.155	255.6
27	0.518	0.026	0.355	0.101	229.5
28	0.819	0.041	0.109	0.031	247.8
29	0.857	0.043	0.078	0.022	259
30	0.286	0.014	0.545	0.155	223.5

Fig. 2 Influence of initial temperature T_0 on simulated annealing.

probability is required to the value-added solution of goal function, and the initial acceptable probability P must be greater than 0.8. The precision with different initial temperature is illustrated in Fig. 2. It can be concluded that the higher the initial temperature is, the larger the probability to accept the value-added solution is, thus time required to reduce the temperature becomes longer, such as $T_0 = 50$ and 100. However, if the initial temperature is too low, it is difficult to find the global optimum point, such as $T_0 = 5$. The initial temperature $T_0 = 20$ is selected according to the rate of convergence and the accuracy of training.

The disturbance coefficient η determines the searching range. The results with different η at initial temperature $T_0 = 20$ are illustrated in Fig. 3. It is obvious that the range of searching becomes smaller and the iteration expedites as η becomes large in the initial process, thus the solution reducing error function is easy to be found. Since the global search in the initial process is insufficient at low temperature, it is possible to either trap in local minimum or require a long search to trap out the local minimum. In this case, the accuracy is difficult to reach or reaches after a long time if η is too large, while the iteration is slow due to the global search if η is small. If ΔX is suited to the temperature, a large number of iteration to reduce the temperature is required. Only if the temperature reduces low enough and the ΔX is appropriate, is the iteration speed fast. All factors considered the proper η is 100.

Fig. 1 Algorithm structure of simulated annealing.

of simulated annealing in order to search larger space. It is known that the range of ΔX becomes large as the initial temperature increases, according to the property of Gaussian distribution in eq. (2). Furthermore, a large acceptable

Fig. 3 Influence of disturbance factor η on Simulated Annealing.

Table 2 Influence of α in annealing scheme on simulated annealing.

α	Time (s)	R	ΔE	ΔX
0.998	503	0.928	6.27E-08	3.04E-06
0.995	430	0.928	-3.88E-05	9.21E-04
0.990	127	0.928	2.20E-08	-5.64E-06
0.980	1249	0.928	8.59E-04	-7.36E-03
0.950	2000	0.898	-5.61E-11	-6.07E-09
0.800	2000	0.927	-8.30E-08	1.00E-05
0.650	2000	0.904	1.33E-08	7.60E-06

The annealing scheme is related to the method producing ΔX . The approximate annealing scheme for Gaussian distribution follows as $T_{k+1} = T_k \cdot \alpha$ ($k = 0, 1, 2 \dots$), and generally α is taken from 0.6 to 0.99. When T_0 is 20 and η is 100, the influence of different α on prediction performance is listed in Table 2.

The velocity of temperature reduction becomes slow as α increases, and the searching space becomes large simultaneously, so it is better to find a stable state at different temperature and more accurate to search the global minimum, such as $\alpha = 0.998$ and 0.995. Whereas, either the global minimum can be accurately found if it is located in the range or the iteration is damped into local minimum if the global minimum is not located in the range and in this case it is more possible to reduce the temperature without reaching the stable state or to converge with a slower rate due to a long time to trip out the local minimum, such as $\alpha = 0.8$ and 0.65. Considering all the factors, α is selected as 0.99.

In the preceding section, most of the final temperature T_E is 10^{-8} . If the goals do not reach below this temperature, it is difficult to find the global minimum with further reduction in temperature. The length of Markov chain can be determined according to the variance dimension and is constant generally.

The final cooling schedule is fixed as $\eta = 100$, $T_0 = 20$, $\alpha = 0.99$, $T_E = 10^{-8}$, and the length of Markov chain is 26. The predicted results with this network and the interpolating and extrapolating performance are illustrated in Figs. 4, 5 and 6 respectively. The comparison of simulated annealing algorithm and traditional gradient descent is listed in Table 3.

Fig. 4 Predicted results and linear regression analysis of simulated annealing algorithm.

Fig. 5 Interpolating performance and linear regression analysis of simulated annealing algorithm.

Fig. 6 Extrapolating performance and linear regression analysis of simulated annealing algorithm.

Table 3 Comparison of predicted performance and cost between different algorithms.

	Correlation coefficient (R)			Time (s)
	All data	Interpolating	Extrapolating	
Gradient descent	0.962	0.795	0.799	436
Simulated annealing	0.919	0.846	0.913	3000
Mixing method	0.993	0.843	0.863	960

3.3 Improved simulated annealing

The searching direction of simulated annealing is stochastic so simulated annealing is suited for large range search but easy to cost long time in accurate search, while the searching direction of gradient descent is fixed so iteration time is short but easy to trap in local minimum. The global minimum can be found in shorter time with combination of these two methods than traditional methods. The searching process is to find the range where the optimum solution lies with simulated annealing, and then to search accurately with gradient descent. If the assigned goal is not obtained with gradient-descent, it indicates that the searching process possibly trap

Fig. 7 Iteration process of mixing algorithm.

in the local minimum and should continue the iteration with simulated annealing, and repeat till the assigned goal reaches. The process is illustrated in Fig. 7.

The optimized results with the network that associates simulated annealing and gradient descent is illustrated in Fig. 8. The comparison with purely gradient descent and simulated annealing respectively is listed in Table 3. It can be concluded that the accuracy with mixing method is greater than with purely either gradient descent or simulated annealing while the time is greatly reduced.

4. Conclusions

The simulated annealing algorithm is adopted in the network and the cooling schedule is determined with the initial temperature T_0 being 20, the disturbance coefficient η being 100, the cooling rule being $T_{k+1} = T_k \times 0.99$, the length of Markov chain being 26 and the final temperature T_E being 10^{-8} . This schedule is used to predict with a correlation coefficient 0.9664 and a better interpolating and extrapolating performance. It is practically adapted to the prediction of initial discharge capacity of AB₅-type alloys with A-side consisting of La, Ce, Pr, Nd. The prediction accuracy of the network with simulated annealing is good, only the time cost is large. By incorporating it with the gradient descent method, both predictions accuracy and velocity can be greatly improved, with a correlation coefficient of predicted results 0.993 and time 960 s. When combining simulated annealing with gradient-descent, the training velocity is

Fig. 8 Predicted results and linear regression analysis of improved simulated annealing algorithm.

increased and good accuracy is obtained. The network adopting the cooling scheme brought forward in this paper and the method mixing simulated annealing and gradient-descent can be employed for predicting the electrochemical properties of AB₅-type alloys.

Acknowledgements

This work is supported by National Research Fund for Fundamental Key Project No. 973 (2002CB211800). We thank Prof. C. M. Hong of Beijing University due to his helpful discussion for this paper.

REFERENCES

- 1) T. Sakai, K. Qguro, H. Miyamura N. Kuriyama, A. Kato, H. Ishikawa and C. Iwakura. *J. Less-Common Met.* **161** (1990) 193–202.
- 2) Y. X. Gai and W. Z. Tang: *Proceeding of '94 China Material Discussion in autumn*, ed. by China Material Society, (The Chemical Industry Press, Beijing, 1995), 425.
- 3) J. Guo, C. H. Li, H. L. Liu, T. S. Huang and N. Y. Chen: *Mater. Sci. and Eng.* **B47** (1997) 96–100.
- 4) J. Guo, C. H. Li, H. L. Liu and N. Y. Chen: *J. Electrochem. Soc.* **144**(7) (1997) 2276–2278.
- 5) L. H. Kok, M. J. Hong, L. Yi and P. Wu: *J. Mater. Chem.* **9** (1999) 837–843.
- 6) J. J. Hopfield: *IEEE Circuits Devices* **11** (1988) 3–10.
- 7) H. H. Szu and R. L. Hattley: *Phys. Lett. A* **122** (1987) 157–162.
- 8) M. Z. Guo, J. F. Jiang and J. M. Li: *Computer Engineering* **26** (2000) 63–66.