RESEARCH



Integrated model construction for state of charge estimation in electric vehicle lithium batteries



*Correspondence: lyy9485@163.com

¹ Department of Mechanical and Electrical Engineering and Intelligent Manufacturing, Zhengzhou Vocational College of Information and Technology, Henan Open University, Zhengzhou 450000, China ² Xianghe Kunlun New Energy Materials Co., Ltd, Langfang 065400, China

Abstract

This research addresses the issue of State of Charge (SOC) prediction for electric vehicle batteries by employing a dynamic Kalman neural network model. The model is optimized using a Genetic algorithm to adjust the neural network weights. Additionally, a strategy involving support vector machines for model optimization is proposed. This strategy involves preprocessing the data, selecting appropriate kernel functions for training, and merging prediction results to enhance the stability of the model. Results indicated that the Dynamic Genetic Kalman Neural Network (DGKNN) model achieved the minimum prediction error percentage of only 0.1529% when the correction coefficient was set to 0.7. The DGKNN model consistently exhibited the lowest error percentage, average absolute error, mean square error, and root mean square error when handling small, medium, and large datasets. For instance, in the small dataset, the error percentage was only 0.1518, and the root mean square error was only 0.0604. The research findings demonstrated that the proposed model exhibited high real-time accuracy in predicting battery SOC, enabling real-time monitoring of battery operating parameters. The method proposed in this study can accurately predict the state of battery charge, extend the life of battery packs, and improve the performance of electric vehicles. It has important significance for promoting the development of the electric vehicle industry.

Keywords: Electric vehicles, Lithium batteries, SOC, Integrated model

Introduction

Electric vehicles play a crucial role in the global eco-friendly market, emerging as the preferred mode of sustainable and environmentally friendly transportation. During operation, the remaining electric charge in the lithium batteries of electric vehicles, known as the State of Charge (SOC), significantly influences the performance and user experience (Venkitaraman and Kosuru 2022; Mastoi et al. 2022; Anwar et al. 2022). Accurately and rapidly estimating the SOC of electric vehicle lithium batteries is pivotal for promoting the use and adoption of electric vehicles. However, due to complex nonlinear factors such as discharge properties, processes, and lifespan under different environmental and load conditions, existing efficient models face limitations (Cui et al. 2022a, b; Xu



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativecommons.org/licenses/by/4.0/.

et al. 2022a). Current SOC estimation methods include voltage-based, resistance-based, model-based, and machine learning-based approaches. Yet, these methods have their respective constraints and inadequacies due to the nonlinearity, complexity, and environmental impact of batteries. Therefore, addressing how to comprehensively consider internal battery states, temperature, charge–discharge status, and other influencing factors by integrating various estimation methods is a crucial topic in the field (Dou et al. 2022; Wu et al. 2023a; Mokayed et al. 2023). This study aims to research and develop such an integrated model, exploring a novel approach to constructing a SOC estimation model. The first section introduces the research objectives, the second section constructs an intelligent SOC prediction model, the third section validates the model's effectiveness, and the fourth section presents the research conclusions.

In this paper, a Dynamic Genetic Kalman Neural Network (DGKNN) algorithm based on genetic optimization is proposed by combining the Kalman filter method and neural network method in traditional SOC prediction methods. Compared with the existing methods, the contribution of this study is to combine the advantages of the neural network method and Kalman filter method, and use the neural network method to fit the battery parameters. It simplifies the parameter identification process of the Kalman filter. The model is updated with Kalman filtering to provide noise removal capability and robustness to the model. At the same time, the Genetic Algorithm (GA) optimization also greatly reduces the initial error, helps the battery management system monitor the remaining power of the battery, extends the life of the battery, and improves the safety and stability of the battery.

Research backgrounds

In recent years, research in the field of SOC has gradually deepened. To prevent the battery from overcharging and discharging, the unscented transform and multiple innovations are applied to the particle filter, optimizing the particle distribution and updating the state values according to historical information. The scented particle filter is formed to estimate the charging state of the battery, and the impact of parameter changes on SOC estimation is considered and verified. The algorithm can accurately estimate the real-time SOC changes, and the average error of SOC is less than 0.5%, which has a high precision (Chen et al. 2024). In order to better learn and estimate the state of battery charge, Chen et al. used a multi-core correlation vector machine and whale optimization algorithm to estimate the SOC of lithium-ion battery under different working conditions, and established the SOC estimation model of lithium-ion battery. The weights and kernel parameters of multi-kernel correlation vector machine were automatically adjusted and optimized by whale optimization algorithm, and the estimation accuracy was improved. Compared with other optimization algorithms, this algorithm had a better optimization effect and could estimate SOC more accurately (Chen et al. 2023). Liang and the team presented a rule-based approach that achieved rapid SOC balance while respecting power constraints. The results showed that compared to existing methods, the proposed algorithm exhibited superior performance (Liang et al. 2022). The team led by Xu proposed a novel SOC estimation method, emphasizing the importance of accurate SOC not only for user experience but also for preventing overcharging, overdischarging, and ensuring safe usage. Due to numerous issues with neural networks in

SOC estimation, they introduced an individual migration dynamic step-size Drosophila algorithm, combining it with neural networks for SOC estimation to enhance accuracy. Experimental results demonstrated that, compared to other algorithms, the proposed algorithm exhibited excellent estimation accuracy, with an average absolute error below 0.8% and a root Mean Square Error (RMSE) below 1.4% (Xu et al. 2022b).

Deep learning algorithms have also been widely applied. Feng and colleagues developed a novel bandwidth selection method by extracting features from multiple domains to comprehensively describe target harmonic responses. This method was employed for Vold-Kalman filtering to choose reasonable bandwidth. Experimental results demonstrated the effectiveness and superiority of this adaptive Vold-Kalman filtering in the diagnosis of wind turbine planetary gearbox (Feng et al. 2023). Revach G and team designed a Kalman filter that integrated a structural state-space model and a dedicated recurrent neural network module, capable of learning complex dynamic behaviors from data. The results showed that KalmanNet efficiently addressed nonlinear and model matching problems (Revach et al. 2022). Esenogho and collaborators proposed a neural network ensemble algorithm based on feature extraction to enhance efficiency in credit card fraud detection. The findings revealed that their proposed Long Short-Term Memory (LSTM) network ensemble outperformed all other algorithms in classifier performance, achieving high levels of sensitivity and specificity (Esenogho et al. 2022). As'ad and his team introduced a mechanistic artificial neural network approach, preserving the advantages of form invariance learned from data-driven regression, while incorporating the physical rationality of mechanistic models. By reinforcing certain good mathematical properties in the network architecture, the authors ensured that the learning process adhered to physical constraints, enhancing the success rate of numerical simulations. The advantages of this learning method were prominently demonstrated in multiple finite element analysis instances, and the new approach was validated for ensuring the computational feasibility of multiscale applications (As'ad et al. 2022).

In the development of societal intelligence, embedded SOC technology played a crucial role and underwent extensive research. Within this domain, a series of studies focused on state prediction to enhance efficiency and safety. The outcomes of these studies reflected the in-depth development and diversification trends in the field. Building upon previous research, the current study incorporated new insights and improvements. The innovation of the research primarily manifested in the construction of an integrated model that fully considered the physical characteristics of lithium batteries and the challenges encountered in practical applications. By combining deep learning methods and integrating various outstanding algorithms, the model aimed to enhance prediction accuracy and stability, positively contributing to the development of electric vehicles.

Multimedia quantum representation model design

Research has focused on the design of a dynamic Kalman neural network that combines the advantages of Kalman filtering and neural networks for predicting the SOC of batteries. However, the model faces challenges related to the extensive data requirements and hardware expenses. To address these issues, a GA is introduced for optimizing the weights of the neural network.



Fig. 1 Schematic diagram of parallel connection before series connection



Fig. 2 Artificial neural network structure

DGKNN model design

The predictive SOC estimation process in this study is based on an Reserve Capacity (RC) battery and system. The model treats a set of batteries, combined through parallel and series connections, as equivalent to a single battery to enhance efficiency, as illustrated in Fig. 1.

To tackle SOC prediction challenges, the study introduces the DGKNN. Kalman filtering demands precise parameter identification and high accuracy, while neural networks require a large amount of sample data, leading to increased hardware costs. To overcome these issues, a novel structure is proposed, which is the Dynamic Neural Network. This structure integrates Kalman filtering and neural networks. Kalman filtering excels in handling dynamic changes but requires an accurate model for useful predictions. Traditional Kalman filters often handle problems of linear dynamic systems only. To overcome this limitation, neural networks are introduced, leveraging their powerful nonlinear approximation capabilities in Kalman filtering to enhance prediction accuracy, as depicted in Fig. 2.

A critical concern in the structure of the dynamic neural network is determining the initial weight values, which directly impact stability and prediction effectiveness. Traditional weight selection methods involve random allocation based on training data or initial assignment according to certain rules. However, these methods do not guarantee model accuracy. To address this, the research explores the use of GA to adjust the initial weights of the neural network. GA provides a globally optimal solution, drawing inspiration from natural selection and genetic principles. GA is a random search algorithm based on natural selection and genetic mechanisms. Its core idea is to inherit excellent individuals to the next generation through heredity, variation and other operations. Its advantage is that it does not require any knowledge related to the question to search. The search uses the evaluation function and the process is simple. It is scalable and easy to combine with other algorithms. Applying GA to weight selection enhances the predictive performance of the model. The nonlinear SOC model is represented by Formula (1).

$$\dot{x} = f(x) + \sum_{i=1}^{n} g(u_i).$$
 (1)

In Formula (1), x represents the SOC, and u represents the input. Firstly, it is necessary to construct a dynamic neural network model, which consists of an input layer, hidden layers, and an output layer. In the input parameters, multiple factors influencing SOC are considered, including current, temperature, and battery aging. In the hidden layers, specific transformations (such as nonlinear activation functions) are applied to convert the input parameters into new state parameters. Subsequently, the Kalman filtering theory is employed to train the neural network model. $S(x_i)$ is the activation function whose expression is shown in Formula (2).

$$S(x_i) = \frac{1}{1 + e^{-x_i}}.$$
(2)

When constructing the error model, the weight set is first established as shown in Formula (3).

$$V = [V_1, V_2, \dots, V_3].$$
(3)

And the input value set is defined by Formula (4).

$$u = [u_1.u_2, \dots, u_n]^T.$$
(4)

The error model can then be described by Formula (5).

$$\dot{x} = Ax + W \cdot S(x_i) + V \cdot u + ME.$$
(5)

In Formula (5), ME represents the error model, and A represents the correction constant. After discretization, it becomes Formula (6).

$$x(k+1) = Ax(k) + W \cdot S(x_i(k)) + V \cdot u + ME(k).$$
(6)

Finally, the error model is obtained as shown in Formula (7).

$$ME = f(x) + g(u) - (Ax + W \cdot S(x) + V \cdot u).$$
⁽⁷⁾

In Formula (7), g(u) represents the mapping relationship. The basic assumption of the model is that the system's state is determined by a known state transition model and an observation model. In the research scenario, the state transition model is the studied neural network model, and the observation model is the actual measured value of SOC. During the model training process, the study needs to update the weights in the neural network model based on the observed SOC values to obtain the best prediction results. The observation model is given by Formula (8).

$$y(k) = x(k+1) - Ax(k) = H(k)w(k) + ME(k).$$
(8)

In Formula (8), w(k) represents the weight vector, and H(k) represents the mapping of w(k) to y(k). The system update Formula is given by Formula (9).

$$\hat{w}(k+1) = w(k) + \zeta(k).$$
 (9)

In Formula (9), $\zeta(k)$ represents the process noise. The weight error covariance prediction matrix is given by Formula (10).

$$P^{-}(k) = \Phi(k)P^{-}(k-1)\Phi^{T}(k) + Q_{w}.$$
(10)

In Formula (10), $\Phi(k)$ represents the Jacobian matrix, Q_w represents the variance of the error. The Kalman filtering gain of the model is given by Formula (11).

$$K(k) = P^{-}(k)H^{T}(k)\left(H(k)P^{-}(k)H^{T}(k) + R_{k}\right)^{-1}.$$
(11)

In Formula (11), $P^-(k)$ represents the error covariance interface matrix, H(k) represents the transpose of the Jacobian matrix, R_k represents the noise covariance matrix. However, since Kalman filtering requires an accurate model, it is often challenging to find the optimal solution quickly during training. Therefore, the study introduces GA for improvement. The GA is responsible for finding initial weights in the training of the neural network model, effectively enhancing the model's predictive performance. By designing appropriate adaptive functions and selection strategies, the efficiency and convergence of GAs are ensured. This leads to a dynamic Kalman neural network model based on genetic optimization for SOC prediction. The genetic optimization process is illustrated in Fig. 3.

For the optimization of initial weights in this paper, it is assumed that *N* DGKNN weight vectors $w(0) = \{w_1(0), w_2(0), ..., w_N(0)\}$ are the initial population. The algorithm set the number of terminating iteration steps *s*, and compare the initial value x(0) of *x* with the initial population $w(0) = \{w_1(0), w_2(0), ..., w_N(0)\}$ plugged into formula (5) to get the value of the next state $x(1) = \{x_1(1), x_2(1), ..., x_N(1)\}$. Let the true SOC value at this time be r(1), Formula (12) represent the fitness function $f_i(k)$.



Fig. 3 Genetic optimization process

$$f_i(k) = 1/|x_i(k) - r(k)|, i = 1, 2, \dots, N.$$
(12)

In Formula (12), r(k) represents the true value, and $x_i(k)$ represents the state value. The selection operation is performed on the population, and the new generation population will be selected from the previous generation population according to probability. The selection probability of each individual is calculated according to the fitness function. The probability of being selected is expressed as Formula (13).

$$p_i(k) = f_i(k) / \sum_{j=1}^N f_j(k), i = 1, 2, \dots, N.$$
 (13)

In Formula (13), *i* and *j* represent different individuals. Then randomly select a number *R* from [0,1], if $R < p_i(k)$, $i = 1, 2, \dots, N$, then the *i*th individual is selected. If $R < p_{i+1}(k)$, then the i + 1 individual is selected, and so on, until all individuals have been traversed. Cross-operation is conducted to the population. It is assumed that the crossover probability is ω_c , then a real number *R* is randomly generated between the interval [0,1]. If $R < \omega_c$. Then two individuals $\{w_m(k), w_n(k)\}$ are randomly selected from the population and then crossed according to Formula (14).

$$\begin{cases} \tilde{w}_m(k) = \alpha w_m(k) + (1 - \alpha) w_n(k) \\ \tilde{w}_n(k) = \alpha w_n(k) + (1 - \alpha) w_m(k) \end{cases}$$
(14)

In Formula (14), α represents a random number in the range [0,1], and $w_m(k)$ and $w_n(k)$ represent different individuals. To mutate the population, it is assumed that the mutation probability is ω_m . A real number *R* is randomly generated between the interval [0,1], if $R < \omega_m$, then randomly select an individual $w_i(k)$ from the population, and then mutate according to the following Formula (15).

$$\tilde{w}_i(k) = w_i(k) + \eta, \eta \sim N\left(0, \sum\right).$$
(15)

In Formula (15), η follows a normal distribution. $w_i k$ represents a random individual. The specific configuration of the genetic neural network layer is: the total number of iterations is 100, the crossover probability is 0.3, the mutation probability is 0.1, and the population size is 150. In this paper, the structure of the neural network is set as 14-10-1, with a total of 150 weights and 11 thresholds, so the individual coding length of the GA is 161, which is the sum of the number of weights and thresholds. By repeating these steps, the process is like the population evolving continuously, and each iteration's chromosome is more adapted to the environment than the previous one. When the preset evolutionary steps are reached, the chromosome with the highest fitness among the remaining chromosomes is the sought-after optimal solution, representing the best network weight configuration.

Improvement design for SVM stability in DGKNN model

After optimizing the initial weights of the DGKNN model through GA, significant progress has indeed been achieved. However, this method still has certain limitations, particularly in terms of prediction stability. Therefore, incorporating the principles of

Support Vector Machine (SVM), the research plan involves a series of enhancements to the studied model to improve prediction stability. Firstly, the data needs preprocessing before proceeding with SVM training. During this stage, an appropriate kernel function must be selected. If the data can be completely separated by a straight line in a two-dimensional plane, a linear kernel is the best choice. On the other hand, if the data exhibits a circular distribution in a two-dimensional plane, the radial basis function is a better choice. The research employs cross-validation and grid search to select the optimal model and parameters, as depicted in the model fusion architecture shown in Fig. 4.

After obtaining preprocessed data and training the SVM model, the next step involves merging the prediction results. Specifically, the SVM prediction results and DGKNN model prediction results are integrated to form the final prediction. Different weights can be assigned to the predictions of the two models, and the two results are then combined to obtain the final result. The prediction result fusion is illustrated in Formula (16).

$$SOC_{final} = \alpha * SVM_{predict} + (1 - \alpha) * DGKNN_{predict}.$$
 (16)

In Formula (16), α represents the weight in the range [0,1], *SVM_predict* represents the SVM prediction result, and *DGKNN_predict* represents the DGKNN model prediction result. This approach maximizes the predictive capabilities of both SVM and DGKNN models while improving prediction stability through rational weight allocation. The SVM optimization strategy diagram is shown in Fig. 5.

By incorporating SVM into the research model, the study aims to achieve more robust prediction results. Throughout this process, the research first preprocesses the data, then trains the SVM, generates prediction results, and finally merges these results with the predictions of the DGKNN model. Subsequently, the study continuously optimizes and adjusts the research model in practical use, making it better suited for real-world environments and providing more accurate and robust SOC predictions.

The life of the battery is limited, and the use of the battery for a long time will reduce the battery capacity and cause the attenuation of the battery life. In real electric vehicle operation, the discharge rate of the battery will change with the increase of the driving speed. Therefore, the connection between the discharge rate and the capacity of the battery must be studied in order to better meet the needs of the car. The relationship between capacity and different discharge rates was studied by conducting discharge



Fig. 4 Model fusion architecture



Fig. 5 SVM optimization strategy

experiments under high performance battery testing equipment. It is found that the terminal voltage of the battery drops rapidly at the beginning, and the voltage changes caused by different discharge rates are different. The rate of voltage drop is positively correlated with the rate of charge and discharge. The smaller the rate of discharge, the slower the voltage drop. After the voltage drops rapidly, it enters the voltage plateau, the terminal voltage changes slowly in this stage. The smaller the discharge rate, the longer the duration of this stage, and the battery is suitable for working in this stage. Finally, when the battery is about to run out of power, the discharge rate will become lower, and the performance of the power battery will be improved. When the voltage at both ends of the battery drops sharply to its set minimum current, the battery is not suitable for work at this stage.

Performance evaluation of SOC prediction models

The study conducted an in-depth examination of the performance of SOC prediction models. The study evaluated the DGKNN model based on genetic optimization and four other predictive models under The U.S. Environmental Protection Agency (EPA) scenario. EPA working conditions are relatively complex, including urban working conditions, high-speed working conditions, high temperature working conditions, low temperature working conditions and intense driving conditions. The final EPA results are obtained by weighting after the completion of the five working conditions. A comparison of the performance of these models in handling small, medium, and large datasets was carried out.

DGKNN model evaluation

With the increase of electric vehicles, users' driving habits have changed, and the acceleration section of actual user conditions was slightly stronger than EPA. The reason was that electric vehicles had stronger performance and greater acceleration. EPA was a product of many years ago, and the main collection of vehicle data was also different from the latest. It may have some differences from the actual test. However, EPA was still used by automobile companies, indicating that the error was within the acceptable range. In designing the experimental validation for the model, the

Project	Content
Experimental conditions	Multiple cycles of NEDC
Maximum operating speed of the vehicle	120 km/h
Average speed	36.01 km/h
The entire process takes time	3074 s
Sampling interval	0.1 s
SOC initial value	0.44
Prediction model	DGKNN based on genetic optimization
Model input variables	Battery voltage (Vn — 1), battery cur- rent, battery temperature
Model output variables	×(it)
Number of population in genetic optimization algorithm	N=50
Terminate iteration steps	F50
Correction constant	MI < 1

Table 1 Experimental settings



Fig. 6 Computational complexity analysis

experiments utilized multiple cycles of the NEDC structure. The vehicle's maximum operating speed was set at 120 km/h, with an average speed of 36.01 km per hour. The entire experimental process took 3074 s, with sampling occurring every 0.1 s, as detailed in Table 1.

As shown in Table 1, the initial SOC value of the battery was set to 0.44. The experiment used a genetic optimization based DGKNN for prediction. The input variables of the model included battery voltage (Vn - 1), battery current, and battery temperature. The output variable of the model was x(it). In terms of genetic optimization algorithms, the population size was set to 50 and the number of termination iteration steps was 50. And a correction constant MI was set, with a value less than 1.

The training environment of this experiment: CPU was Intel(R) Core(TM)i9-10920X, memory was 16G. The GPU was GeForce RTX 2080 Ti and the video memory was 11G. The operating system was Windows 64-bit, and the CUDA11.2 library file was installed. The development language was Python, Pytorch1.7.1 framework. Figure 6 showed the CPU usage of the proposed model. During the operation of DGKNN model, the CPU usage did not exceed 50%, and the average was 38.21%, which had low requirements for system hardware configuration. The optimization effect analysis was shown in Fig. 7.



Fig. 7 Genetic optimization effect analysis

Figure 7 illustrated that without the application of GA optimization, DGKNN's initial prediction error was significant. Conversely, when the GA was applied, this error significantly diminished. This observation strongly suggested that GA optimization was crucial in the application of DGKNN. Furthermore, with the optimization of weights through the GA, a substantial improvement was observed. The algorithm significantly reduced the error in the initial predictions, alleviating the stiffness observed in the DGKNN model without optimization. This enhancement boosted model performance, further enhancing the precision and practicality of prediction results. DGKNN prediction effectiveness was detailed in Fig. 8.

Figure 8 showed a significant reduction in prediction error when DGKNN was used based on genetic optimization compared to Deep Neural Network (DNN) and Genetic models. This finding provided crucial reference information for research and design. Detailed analysis of prediction errors revealed the superiority of DGKNN after genetic optimization. Genetic optimization not only improved prediction accuracy within stages but also optimized accuracy in multi-stage predictions, reducing error fluctuations. In contrast, DNN and Genetic models exhibited lower fitting to the true value curve, with larger prediction error fluctuations. Genetic optimization-based DGKNN outperformed DNN and Genetic models in prediction errors, emphasizing the importance of genetic optimization in controlling prediction system errors. Prediction errors under different correction coefficients were outlined in Table 2.

In Table 2, considering the percentage error, the percentage error ranged from 0.1529 to 0.3307. At a correction coefficient of 0.9, the percentage error reached its maximum value of 0.3307, indicating the maximum deviation of predicted results from actual values. Conversely, when the correction coefficient was 0.7, the percentage error was minimal, only 0.1529, indicating the highest accuracy of the model's predictions at



Fig.8 Analysis of DGKNN prediction effect

Correction factor	Error percentage	Mean absolute error	MSE	RMSE
0.1	0.2078	0.0435	0.0236	0.1536
0.2	0.2289	0.0462	0.0314	0.1773
0.3	0.2614	0.0527	0.0381	0.1952
0.4	0.1733	0.0363	0.0172	0.1311
0.5	0.1789	0.0350	0.0156	0.125
0.6	0.1987	0.0405	0.0205	0.1432
0.7	0.1529	0.0324	0.0154	0.1243
0.8	0.2875	0.0572	0.0432	0.2079
0.9	0.3307	0.068	0.0607	0.2464

Table 2 Prediction error under different correction coefficients

this correction coefficient. Moving on to the mean absolute error (MAE), this metric fluctuated between 0.0324 and 0.068. At a correction coefficient of 0.9, the MAE was maximum at 0.068, signifying the largest average absolute error between predicted and actual values. However, at a correction coefficient of 0.7, the MAE was minimal at 0.0324, indicating the smallest average gap between predicted and true values. Regarding the Mean Square Error (MSE), a similar trend was observed, with the range fluctuating from 0.0154 to 0.0607. The maximum value occurred at a correction coefficient of 0.9, while the minimum value was attained at a correction coefficient of 0.7, consistent with the patterns observed in the previous two metrics. Finally, considering the RMSE, a standard measure of prediction error, its values varied from 0.1243 to 0.2464. The maximum RMSE was observed at a correction coefficient of 0.9, indicating larger fluctuations in predicted values. On the other hand, the minimum RMSE occurred at a

correction coefficient of 0.7, suggesting relatively accurate prediction results. It is evident that, under a correction coefficient of 0.7, all four error metrics were minimized, indicating the highest accuracy of the model. The reason of the high prediction accuracy of DGKNN was that it took into account the influence of measurement noise and time. In addition, because the internal resistance, capacitance and other parameters of the battery were not easy to measure, it was very difficult to identify the parameters of other models. However, DGKNN adopted a neural network method to fit the parameter relationship inside the battery, which greatly simplified the process of parameter identification and thus reduced the difficulty of modeling. Table 3 showed the difference between the MAE, MSE and RMSE of the DGKNN algorithm, AGA-UKF-EKF algorithm (Wu et al. 2023b) and ACHF algorithm (Serat et al. 2023).

From the RMSE data in the table, the DGKNN algorithm could effectively filter noise and better fit the actual working conditions of the battery. The smaller the ratio of the three pairs, the better the outcomes. The smaller the ratio indicated that the experimental model had better accuracy and could better predict the data. The average absolute error of the DGKNN algorithm was 0.32%, and the maximum error was 0.84%. Compared with the errors of the other two algorithms, the error of the DGKNN algorithm in estimating the SOC of lithium battery was smaller, indicating that the improved algorithm can estimate the battery capacity in real time and reduce the influence of capacity on the estimated SOC of the battery. Compared with other algorithms, the improved algorithm had higher precision and its results were more reliable.

Comprehensive model validation

The study compared the predictive effectiveness of four models for estimating the remaining capacity of lithium-ion batteries. These models included the research design model, Support Vector Regression Model (SVR), Genetic Algorithm-Backpropagation Model (GA-BP), and Particle Swarm Optimization-Adaboost Model (PSO-adaboost). The predictive results were illustrated in Fig. 9.

As shown in Fig. 9, four different models were compared in SOC prediction: the Research Design model, GA-BP model, PSO-adaboast model, and SVR model. Firstly, the Research Design model exhibited outstanding performance in SOC prediction. It achieved high accuracy, a well-fitting prediction curve, and minimal errors. This model accurately estimated battery states with high stability. Following that, the GA-BP model ranked second in prediction performance, closely following the Research Design model. Despite its slightly higher complexity, the GA-BP model demonstrated robustness and accuracy. Thirdly, the PSO-adaboast model ranked third in prediction performance, while the SVR model performed the least effectively. The main reason for this phenomenon was that the initial weights in DGKNN were random numbers. Although the GA

Algorithm	DGKNN	AGA-UKF-EKF	ACHF
MAE	0.0032	0.0042	0.0067
MSE	1.6105×10^{-5}	2.00773×10^{-5}	4.8521×10^{-5}
RMSE	0.00040	0.0046	0.0070

Table 5 Companson of estimated	d errors
--------------------------------	----------



Fig. 9 Comparison of prediction effect of different models

Modeling methods	Dataset level	Error percentage	MAE	MSE	RMSE
DNN	Minor level	0.2654	0.1011	0.1085	0.3279
	Middle level	0.2689	0.1025	0.1099	0.3315
	Grand level	0.271	0.1032	0.111	0.3334
SVR	Minor level	0.2035	0.0668	0.0068	0.0802
	Middle level	0.2058	0.0702	0.0072	0.0846
	Grand level	0.2093	0.0724	0.0075	0.0866
GA-BP	Minor level	0.1901	0.0605	0.0053	0.0719
	Middle level	0.1932	0.0631	0.0056	0.0746
	Grand level	0.195	0.0648	0.0059	0.0768
PSO-adaboast	Minor level	0.1913	0.0632	0.0058	0.0761
	Middle level	0.1939	0.0654	0.0061	0.0782
	Grand level	0.196	0.0665	0.0063	0.0793
Research model	Minor level	0.1518	0.0503	0.0039	0.0604
	Middle level	0.1542	0.0517	0.0041	0.064
	Grand level	0.1561	0.0525	0.0042	0.0648

Table 4 Comparison of prediction errors of different models

has been optimized, the results of the GA itself had randomness, while the parameters of SVR had no random numbers. At the same time, compared with neural networks, FCDE-SVR had fewer requirements on the amount of data. Therefore, although the FCDE-SVR algorithm required a period of training, the algorithm was more complex, but stable with high precision. Error analysis was presented in Table 4.

As shown in Table 4, across all three data set levels, DNN exhibited the highest error percentages, MSE, and RMSE values. For example, in the large data set, its error percentage was 0.271, MSE was 0.111, and RMSE was 0.3334. This may indicate that DNN's

predictive performance for these data sets was inferior to other models. In contrast, the Research model consistently showed the lowest error percentages, MAE, MSE, and RMSE across all three data set levels. For instance, the error percentage for the small data set was only 0.1518, and the RMSE was only 0.0604, highlighting the Research model's superior predictive accuracy. As for the other three models (SVR, GA-BP, and PSO-adaboast), their error metrics exhibited similar trends across small, medium, and large data sets. Generally, as the data level increased, the error metrics also increased, possibly due to a decrease in predictive accuracy with increased data complexity. For instance, in the small data set, GA-BP and PSO-adaboast had similar RMSEs, both around 0.07. However, in the large data set, GA-BP's MSE increased to 0.0768, while PSO-adaboast's increased to 0.0793. Considering these factors, the Research model outperformed the other models, indicating that its structure or parameter settings were more suitable for predicting this type of data. Table 5 showed the comparison of errors of different models at 6 different temperatures.

As can be seen from the table, the errors of all algorithms increased significantly at 0 °C, because the battery activity decreased at a lower temperature, and the battery internal resistance increased, resulting in a cumulative increase in errors during the estimation process. Compared with other methods, the proposed model had obvious advantages. Especially at low temperatures, the accuracy of the model fluctuated the least compared with that at normal temperatures, which indicated that the algorithm proposed in this paper had better robustness at different temperatures. Battery state analysis is depicted in Fig. 10.

In Fig. 10, the Research Design model demonstrated real-time capabilities along with precise analysis and processing of battery operating parameters. These parameters include current, terminal voltage, temperature, vehicle speed, and battery power. Real-time monitoring of these parameters not only aids in anticipating and preventing potential faults but also allows users to understand the battery's usage conditions and take measures to improve battery performance and extend its lifespan. This helps users gain better insights into the status of electric vehicles and their batteries, ensuring the safe and efficient use of electric vehicles.

Results and discussion

In this paper, a DGKNN based on genetic optimization is proposed, which combines the advantages of both the neural network method and Kalman filter method. It uses the neural network method to fit battery parameters to simplify the parameter identification process of Kalman filter. The model is updated with Kalman filter to provide noise removal capability and robustness. At the same time, the initial error is greatly

1						
RMSE	−10 °C	0 °C	10 °C	25 °C	35 °C	45 °C
EKF (Liu et al. 2023)	1.27	1.21	1.13	1.07	1.10	1.33
CKF (Cao et al. 2023)	1.35	1.19	1.16	1.04	1.11	1.38
UKF (Ly and El-Sayegh 2023)	1.36	1.24	1.09	1.08	1.15	1.29
Research model	1.22	1.15	1.05	1.02	1.09	1.27

 Table 5
 Comparison of errors of different models at 6 different temperatures



Fig. 10 Real-time change curve of each battery parameter

reduced by GA optimization. The results of this paper show that DGKNN's prediction effect is significantly better than that of the initial neural network, and DGKNN does not need to undergo long-term and large data training like the initial neural network, which means that if the battery is replaced, the initial neural network will have to undergo long-term training, which will have a great blow to the practicality of its algorithm. This also means that DGKNN has an advantage over initial neural networks in real-time SOC prediction. Therefore, DGKNN based on genetic optimization is a very suitable algorithm for SOC prediction.

In the selection of the correction coefficient, the research is obtained by comparing and analyzing the known standard values and the measured results. When the deviation of the measured result is relatively large, the correction coefficient can be used to correct the measured data to make it closer to the real value. When the error of the measurement result is relatively large, the correction coefficient can be used to adjust the measurement data, reduce the error, and improve the accuracy and reliability of the measurement. However, since the initial weights of DGKNN are obtained by the method of random generation before genetic optimization, both steps will lead to randomness of the weights, resulting in the generated weights being not fixed, and the results of each prediction will be different, which will make the algorithm unstable and the prediction effect will be good or bad. At the same time, the prediction effect of both charge and discharge data is not ideal.

In order to solve the problem of the DGKNN model, the FCDE-SVR algorithm is proposed. The algorithm forms clusters with different features through feature clustering and data screening, which effectively reduces the amount of data and the extracted data features are more obvious, which is suitable for modeling. Using SVR as the sublearning machine of ensemble learning, each cluster is trained by SVR, and then each sublearning machine is integrated to get the final result. The method obtains the feature relationship between data by combining density and distance, and updates the weights in the process of ensemble learning based on this, and realizes the selection of sub-data sets and the integration of the final prediction model.

Conclusion

In addressing the real-time prediction of the remaining battery capacity in electric vehicles, this study aimed to design a SOC prediction model with high accuracy and stability. Combining the advantages of Dynamic Kalman Neural Network with GA and SVM, the research employed a fusion strategy to tackle the substantial demand for sample data and device hardware overhead. The results revealed that, with a correction coefficient of 0.7, the predictive error percentage of the designed model in the study was minimized to only 0.1529%. The average absolute error reduced to 0.0324, while the RMSE decreased to 0.0604. When handling datasets of different scales, the designed model in the study consistently maintained an advantage in predictive accuracy. In the case of large-scale datasets, the error percentage was 0.1518, and the RMSE was 0.0604. Meanwhile, other models such as DNN, SVR, GA-BP, and PSO-adaboast exhibited relatively poorer performance across various dataset levels, especially in the scenario of large-scale datasets. In summary, the SOC prediction model proposed in this study, based on DGKNN combined with GA and SVM, demonstrates a significant advantage in accuracy and real-time performance. SOC estimation is not only affected by temperature, but also its accuracy is affected by the current remaining capacity. In SOC estimation, the default capacity is unchanged, while in the actual situation, the battery capacity is constantly changing with the temperature and the number of cycles, and the relative capacity of the battery can be calibrated in the future.

Author contributions

Yuanyuan Liu wrote the main manuscript text; Wenxin Dun reviewed and revised the manuscript. Both authors approved this submission.

Funding Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval and consent to participate No human or animal subjects or materials were used.

Consent for publication Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 6 February 2024 Accepted: 6 March 2024 Published online: 14 March 2024

References

Anwar MB, Muratori M, Jadun P, Hale E, Bush B, Denholm P, Ma O, Podkaminer K (2022) Assessing the value of electric vehicle managed charging: a review of methodologies and results. Energy Environ Sci 15(2):466–498

As'ad F, Avery P, Farhat C (2022) A mechanics-informed artificial neural network approach in data-driven constitutive modeling. Int J Numer Methods Eng 123(12):2738–2759

Cao H, Wu Y, Bao Y, Feng X, Wan S, Qian C (2023) UTrans-Net: a model for short-term precipitation prediction. Artif Intell Appl 1(2):106–113

Chen K, Zhou S, Liu K, Gao G, Wu G (2023) State of charge estimation for lithium-ion battery based on whale optimization algorithm and multi-kernel relevance vector machine. J Chem Phys 158(10):104110–104120

Chen P, Jin X, Han XF (2024) Joint estimation of state of charge and state of health of lithium ion battery. J Electrochem Energy Convers Storage 21(01):11008–11020

Cui Z, Wang L, Li Q, Wang K (2022a) A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network. Int J Energy Res 46(5):5423–5440

Cui Z, Kang L, Li L, Wang L, Wang K (2022b) A hybrid neural network model with improved input for state of charge estimation of lithium-ion battery at low temperatures. Renew Energy 198:1328–1340

Dou Y, Kan D, Su Y, Zhang Y, Wei Y, Zhang Z, Zhou Z (2022) Critical factors affecting the catalytic activity of redox mediators on Li–O₂ battery discharge. J Phys Chem Lett 13(30):7081–7086

Esenogho E, Mienye ID, Swart TG, Aruleba K, Obaido G (2022) A neural network ensemble with feature engineering for improved credit card fraud detection. IEEE Access 10:16400–16407

Feng K, Ji JC, Ni Q (2023) A novel adaptive bandwidth selection method for Vold-Kalman filtering and its application in wind turbine planetary gearbox diagnostics. Struct Health Monit 22(2):1027–1048

Liang G, Rodriguez E, Farivar GG, Ceballos S, Townsend CD, Gorla NB, Pou J (2022) A constrained intersubmodule state-ofcharge balancing method for battery energy storage systems based on the cascaded H-bridge converter. IEEE Trans Power Electron 37(10):12669–12678

Liu K, Sun Y, Yang D (2023) The administrative center or economic center: which dominates the regional green development pattern. A case study of Shandong Peninsula urban agglomeration, China. Green Low-Carbon Econ 1(3):110–120

Ly A, El-Sayegh Z (2023) Tire wear and pollutants: an overview of research. Arch Adv Eng Sci 1(1):2–10

Mastoi MS, Zhuang S, Munir HM, Haris M, Hassan M, Usman M, Bukhari SS, Ro JS (2022) An in-depth analysis of electric vehicle charging station infrastructure, policy implications, and future trends. Energy Rep 8:11504–11529

Mokayed H, Quan TZ, Alkhaled L, Sivakumar V (2023) Real-time human detection and counting system using deep learning computer vision techniques. Artif Intell Appl 1(4):221–229

Revach G, Shlezinger N, Ni X, Escoriza AL, Van Sloun RJ, Eldar YC (2022) KalmanNet: neural network aided Kalman filtering for partially known dynamics. IEEE Trans Signal Process 70:1532–1547

Serat Z, Fatemi SAZ, Shirzad S (2023) Design and economic analysis of on-grid solar rooftop PV system using PVsyst software. Arch Adv Eng Sci 1(1):63–76

Venkitaraman AK, Kosuru VSR (2022) A review on autonomous electric vehicle communication networks-progress, methods and challenges. World J Adv Res Rev 16(3):13–24

Wu Y, Zhao Z, Hao X, Xu R, Li L, Lv D, Huang X, Zhao Q, Xu Y, Wu Y (2023a) Cathode materials for calcium-ion batteries: current status and prospects. Carbon Neutraliz 2(5):551–573

Wu Z, Zhao Y, Zhang N (2023b) A literature survey of green and low-carbon economics using natural experiment approaches in top field journal. Green Low-Carbon Econ 1(1):2–14

- Xu Y, Chen X, Zhang H, Yang F, Tong L, Yang Y, Yan D, Yang A, Yu M, Liu Z, Wang Y (2022a) Online identification of battery model parameters and joint state of charge and state of health estimation using dual particle filter algorithms. Int J Energy Res 46(14):19615–19652
- Xu H, Wang S, Fan Y, Qiao J, Xu W (2022b) A novel Drosophila-back propagation method for the lithium-ion battery state of charge estimation adaptive to complex working conditions. Int J Energy Res 46(11):15864–15880

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.