One-shot active learning for globally optimal battery electrolyte conductivity

Fuzhan Rahmanian\textsuperscript{1,2}, Monika Vogler\textsuperscript{1,2}, Christian Wölke\textsuperscript{3}, Peng Yan\textsuperscript{3}, Martin Winter\textsuperscript{3,4}, Isidora Cekic-Laskovic\textsuperscript{3}, Helge S. Stein\textsuperscript{1,2,*}

\textsuperscript{1}Helmholtz Institute Ulm, Applied Electrochemistry, Helmholtzstr. 11, 89081 Ulm, Germany
\textsuperscript{2}Karlsruhe Institute of Technology, Institute of Physical Chemistry, Fritz-Haber-Weg 2, 76131 Karlsruhe, Germany
\textsuperscript{3}Helmholtz-Institute Münster (IEK-12), Forschungszentrum Jülich GmbH, Corrensstraße 46, 48149 Münster, Germany
\textsuperscript{4}MEET Battery Research Center, University of Münster, Corrensstraße 46, 48149 Münster, Germany.

Address correspondence to: helge.stein@kit.edu

Abstract

Non-aqueous aprotic battery electrolytes, among other requirements, need to perform well over a wide range of temperatures in practical applications. Herein we present a one-shot active learning study to find all conductivity optima, confidence bounds, and relating trends in the temperature range from 30 °C to 60 °C. This optimization is enabled by a high-throughput formulation and characterization setup guided by one-shot active learning utilizing robust and heavily regularized polynomial regression. Whilst there is an initially good agreement for intermediate and low temperatures, there is a need for the active learning step to globally improve the model. Optimized electrolyte formulations likely correspond to the highest physically possible conductivities within this system when compared to literature data. A thorough error propagation analysis yields a fidelity assessment of conductivity measurements and electrolyte formulation.

Introduction

High-conductivity electrolytes in secondary batteries are of paramount importance for ensuring high performance and reliability of each battery cell chemistry\textsuperscript{1}. In specialty applications such as aerospace or stationary storage\textsuperscript{2} in remote locations, bespoke electrolytes are however necessary. High or low temperatures make the electrolyte a limiting performance factor\textsuperscript{3,4} e.g. in electric vehicles which suffer from relatively narrow optimal temperature windows of 15 °C to 35 °C\textsuperscript{1,2}. Many studies\textsuperscript{2,3,5} have thus been conducted to evaluate lithium-ion battery (LIB) electrolytes at low temperatures in respect to their conductivity, however to date there, is no comprehensive study that identifies optimal formulations and conductivity trends thereof at any practically reasonable temperature.

We therefore present a one-shot active learning study to find the optimally conducting electrolyte formulation at temperatures ranging between -30 °C and 60 °C. Active learning uses machine learned “black-box” models to predict maxima and uncertainties to iteratively discover global maxima and
reduce the model uncertainty. Typically these models are trained in small (down to single sample) batches starting from scratch such that an acquisition function taking the model uncertainty into account is necessary. Through the availability of an already very broadly covering dataset of lithium hexafluorophosphate (LiPF₆) in ethylene carbonate (EC), ethyl methyl carbonate (EMC) and propylene carbonate (PC) totaling 80 electrolyte formulations at 10 temperatures as described by Flores et al., we chose to perform one-shot active learning in a fully exploitative mode. This approach has been shown to increase the so-called “enhancement factor” by Rohr et al. by up to a factor of 10 given the right model and initial training dataset choice. This enhancement factor determines the increase in probability of finding a good material or process. There are however other research modes not explored in this study since the focus is on the pure conductivity optimization of battery electrolytes. However, a recent study by Flores et al. focused on the “understanding driven” research mode. Their symbolic regression approach works well for high temperatures but fails for highly concentrated liquid electrolytes at low temperatures, indicating a change in the physicochemical behavior. This study is enabled through a comprehensive dataset that generated machine learned suggestions that were proven by the experimentalists. Organizationally, this study is the human-in-the-loop version of the fully autonomous active learning study presented by Rahmanian et al.

Methods

An initial dataset of 80 electrolyte formulations was considered using an automated formulation and characterization setup. Each electrolyte formulation contains EC, propylene carbonate PC, e EMC in a solvent/co-solvent mixture and LiPF₆ as the conducting salt. Formulations were characterized between -30 °C and 60 °C, at increments of 10 °C as described by Krishnamoorthy et al. Conductivity measurements were repeated 5 to 7 times. Uncertainties for the experimental values are expressed by the min/max spread of the measured values. For each datapoint, the electrolyte formulation, conductivity and measurement temperature were recorded. Across all formulations the ratio of (EC + PC) : EMC was fixed either at 3:7 or 1:1 by weight and the concentration of LiPF₆ was selected between 0.2 and 2.1 mol/kg. For model training and one-shot active learning, the mass ratios of PC and LiPF₆ were normalized and referenced as $r_{PC} = \frac{PC}{(PC + EC)}$ and $r_{LiPF_6} = \frac{LiPF_6}{(PC + EC)}$. The (EC + PC) : EMC ratio was not considered during model training.

Model training and one-shot active learning

The dataset size poses the challenge of finding well performing models that are simple and interpretable. We therefore settle on polynomial regression for our purely exploitative active learning study. Contrary to Flores et al. we do not consider inter temperature trends in model training. The basic setup is a strongly regularized polynomial regressor aiming to avoid multicollinearity, which is measured by variance inflation factors. The polynomial regressor, ridge regularization, and hyperparameter tuning are performed using the scikit-learn library. From the fitted polynomial model a fine subsampling is performed across $10^4$ formulation ratios at a fixed grid spacing of 1 ratio-%. From this fine subsampling, the top one percentile for each temperature was reported to the experimentalists.
resulting in a total of 100 formulation suggestions. A subset of 24 formulations was chosen by the experimentalists covering all suggestions for -30°C, 20°C and 60°C. Subsequent to the formulation and conductivity measurements of the new formulations, the model was retrained on the entire dataset. For hyperparameter tuning (details see S5) we performed a Bayesian search\textsuperscript{20,23} with a threefold cross validation. The best parameters are then fed to the surrogate model. This search uses ridge regularized polynomial models to favor low polynomial degrees. To assess the model uncertainty for both aleatoric and epistemic uncertainty after the learning shot (and the possible necessity for a second learning shot) we build a pipeline applying the model agonistic prediction interval estimator (MAPIE)\textsuperscript{24}. This estimator uses jackknife plus\textsuperscript{25} to estimate the uncertainty\textsuperscript{26} of the model for a 95\% prediction interval. A summary of this study’s workflow can be seen in Figure 1. Summarizing there are three stages in this pipeline: 1) model training, 2) formulation suggestion and measurement, 3) retraining and refinement of the model and uncertainty quantification.

**Figure 1:** Schematic diagram of this study’s pipeline consisting of initial model training, suggesting formulation to experimentalists for measurement of requested formulations and retraining the model by one-shot active learning with uncertainty quantification

**Results and Discussion**

**Pre-shot model training**
This study uses the same dataset underlying the study presented by Flores et al.\textsuperscript{8} using the formulation and characterization setup reported by Krishnamoorthy et al.\textsuperscript{10}. The herein presented one shot active learning approach is model free, meaning that we do not utilize any physics or chemistry knowledge except correct pose of the input (formulation) and output (conductivity) and a compartmentalization of the problem by temperature.

The global trends of electrolyte conductivity, captured by our model, are shown in Figure 2, which illustrates the conductivity ($\sigma$) over $r_{LiPF_6}$ and $r_{PC}$ at -30°C, -10°C, 30°C, and 50°C (additional temperatures see S2). For all considered temperatures the $R^2$ score is $\approx$0.73 - 0.80, which indicates a good fit. However, the degree of the polynomial used for the fit is higher for the high temperatures compared to low temperatures. The orange datapoints indicate the train set formulations. Overall conductivity is strongly correlated with temperature as expected from Debye-Hückel-Onsager (DHO) theory\textsuperscript{27,28}. Consequently, we observe low conductivity for $r_{LiPF_6}$ >0.8 or <0.1. In general, we observe the maximum conductivity shifting towards higher conducting salt concentrations at higher temperatures as it was studied by Landesfeind et al.\textsuperscript{30,27,28}. The lowest overall measured conductivity is $1.94 \ mScm^{-1}$ at-30 °C showing a generally less pronounced dependence on $r_{PC}$ than on $r_{LiPF_6}$. This observation correlates with the concentration-conductivity relationship that is primarily dependent on conducting salt concentration\textsuperscript{30}.

Going from low to high temperatures, the system seems to allow for higher $r_{PC}$ and $r_{LiPF_6}$\textsuperscript{31,32} while yielding a high conductivity which is in good agreement with established theory. The model also seems to prefer little presence of PC at low temperatures for higher conductivity. Our finding is in line with Ding et al.\textsuperscript{33} who reported that from low to high temperatures, electrolyte formulation first enriches in EC\textsuperscript{34} then in PC which results in higher dielectric constant and consequently higher conductivity\textsuperscript{30,35}. A narrow global optimum at high temperatures at relatively high $r_{PC}$ of $\approx$0.3 , however lower conducting salt concentrations at 30°C and 50 °C can be observed. All but the -10°C optima exist near unsampled formulations. Based on the prediction of the trained model, 10 samples with highest predicted conductivity for each temperature were selected and reported to the experimentalists. The requested and considered formulations can be found in the https://github.com/BIG-MAP/electrolyte_optimization_one_shot_active_learning.
Figure 2: Trends in electrolyte conductivity at a) -30°C, b) -10°C, c) 30°C, d) 50°C. Orange data points represent the experimentally measured conductivity ($\sigma$) at the respective $r_{PC}$ and $r_{LIPF6}$. There is an overall incremental trend for higher $r_{PC}$ from -30°C to -10°C and a narrow optimum in electrolyte conductivity at higher $r_{PC}$ and lower $r_{LIPF6}$ at an unsampled formulation for high temperatures.

One-shot predictions and measurements

Utilizing the above polynomial models’ results, shown in Figure 2, a total of 100 electrolyte formulation (1 percentile) in an intentionally full exploitation mode were selected and suggested to the experimentalists by only communicating the formulation and the respective temperature, however not the predicted conductivity. The experimentalists randomly selected 24 temperature optima from the suggested ones. These selected formulations correspond to temperatures of -30 °C, 20 °C and 60 °C. Figure 3 compares the predicted and experimentally measured conductivity for the 24 selected formulations. There is a small deviation between the requested and actually measured formulations due to slight imperfections in the formulation process hence Figure 3 shows the conductivity prediction at the actually formulated composition. Inaccuracies occurring during the solid and liquid dispensing processes are technical in nature and are negligible given the fidelity assessment below. The error bar illustrates the conductivity error by reporting the maximum and minimum values among the repeated measurements.
The measured conductivities for the requested formulations are added as prior knowledge and the model is retrained. The regularization coefficient and polynomial degree are tuned subsequently using a Bayesian search optimization (See S5) and model uncertainty was incorporated using jackknife plus strategy (See S4).

**Figure 3:** Comparison between measured and predicted conductivity values at a) -30°C, b) -10°C, c) 30°C, d) 50°C sorted by the predicted value at the respective temperature. Orange points represent the mean values of measured conductivities with error bars relating to the min/max from repeated measurements. The high accuracy for low temperature predictions is best observed in the sorted predictions in a) where the lowest 8 predictions were the formulations predicted to be best at 60 °C, the intermediate 8 predictions were the optima for 20 °C and the top 8 are the actual -30 °C optima. Overall this suggests that there exists no globally optimal electrolyte and performance can vary by up to a factor of 2. Predictions for high temperatures are inaccurate suggesting a poor fit model.

**Post-shot model refinement**

After one-shot active learning and the Bayesian hyperparameter tuning (described in the methods section), models are significantly improved. The predicted trends for low temperatures changed only marginally whereas the improvements for temperatures of 30°C and 50°C are significant as shown in Figure 4 (additional temperatures in S3). Together with the low temperature trends there is now a coherent trend observed suggesting higher \( r_{PC} \) and \( r_{LIPF_6} \) at higher temperatures for the temperature refined optimal conductivity\textsuperscript{29,33}. This highlights the significance of active learning in model refinement as we only added an additional 30% of data points to the dataset whilst qualitatively improving the model. Comparing these results to the symbolic regression model by Flores et al\textsuperscript{8} reveals significant differences in the mass ratios required for maximum conductivity.
The drastic improvement of the model becomes even more obvious upon plotting the temperature maxima with the spread of the top percentile as displayed in Figure 5. Before the learning shot, the optima followed no physically meaningful or interpretable trend whereas after adding the extra data the very fine trends in optima towards higher $r_{lip_{PC}}$ and slightly more $r_{PC}$ become obvious. Uncertainty quantification was performed using the jackknife plus strategy resulting in an average 95 % prediction uncertainty interval of $3 \cdot 10^{-3}$ mS/cm (See S4). However, the incorporation of the model agnostic prediction technique allows the measurement of aleatoric and epistemic uncertainty at any point. Comparing the results for electrolyte conductivity found by our one-shot active learning approach to literature such as Ding et al.$^{28}$ at 60 °C, and -30 °C suggests that the herein reported maxima corresponds to the global physical maximum conductivity in this system which is ≈12 mS/cm and 1.9 mS/cm. In another study, Landesfeind et. al.$^{29}$ indicate global maxima of 4.7 mS/cm, 7.6 mS/cm and 9.25 mS/cm at -10 °C, 20 °C, and 30 °C, respectively. Their results are compatible with our findings.

**Figure 4:** Trends in electrolyte conductivity after one shot active learning for optimizing the regressor at higher temperatures and understanding the underlying physicochemical properties of optimal electrolyte formulations. The selected formulations which were suggested from the previous model and used for conductivity determination, were added to the training set as an additional prior knowledge. Model parameters tuning and uncertainty measurement were implemented at this stage of active learning (See S4). Optimized conductivity trend for the a) 30 °C, and b) 50 °C. Trends for additional temperatures can be seen in the S3.
Figure 5: Trends of maximum conductivity a) before and b) after one-shot active learning and model optimization. Before introduction of the additional 24 electrolyte formulations, the trends are neither physically nor qualitatively interpretable. Each point corresponds to the mean conductivity value of the top percentile kernel. The amount required for other electrolyte solvents can be calculated as explained in the method section. The error bars represent the spread of $r_{\text{LIPF}_6}$ and $r_{\text{PC}}$ within the top percentile and not the uncertainty. Overall, higher $r_{\text{LIPF}_6}$ is needed at higher temperatures to reach the optima, with a minutely higher $r_{\text{PC}}$ from 20°C onward.

**Interactions and method fidelity**

Through the availability of a machine learning model that accurately and precisely predicts the trends in conductivity for all temperatures, an assessment of confounding inputs and method fidelity can be pursued. The model has two inputs: $r_{\text{PC}}$ and $r_{\text{LIPF}_6}$, and through the polynomial nature an analytical derivation is facile. The post-shot regularized polynomial equation for deriving conductivity ($\sigma$) from hyperparameter tuning is:

$$\sigma = c_0 + c_1 r_{\text{PC}} + c_2 r_{\text{LIPF}_6} + c_3 r_{\text{PC}}^2 + c_4 r_{\text{PC}} r_{\text{LIPF}_6} + c_5 r_{\text{PC}}^3 + c_6 r_{\text{LIPF}_6} + c_7 r_{\text{PC}} r_{\text{LIPF}_6} + c_8 r_{\text{PC}}^3 + c_9 r_{\text{LIPF}_6}^3$$ (eq.1)

i.e. a polynomial of degree 3 with the individual parameters shown in Table 1. Some coefficients change drastically with temperature whilst others barely change. Upon careful comparison to eq.1, one can see that those coefficients corresponding to a conducting salt-ratio-only term, scale almost exponentially whilst all others, i.e. solvent-ratio-only and solvent-conducting salt-ratio terms scale sigmoidal with temperature (See S6.b). These interaction coefficients allow for further research into the relationship governing the solvation shell properties upon electrolyte solvent variation.
Table 1: Polynomial coefficients incorporating ridge regularization after one-shot active learning for T= -30 °C to 60 °C

<table>
<thead>
<tr>
<th>T (°C)</th>
<th>c \cdot 10^{-3}</th>
<th>c_0</th>
<th>c_1</th>
<th>c_2</th>
<th>c_3</th>
<th>c_4</th>
<th>c_5</th>
<th>c_6</th>
<th>c_7</th>
<th>c_8</th>
<th>c_9</th>
</tr>
</thead>
<tbody>
<tr>
<td>-30</td>
<td>1.0</td>
<td>-0.3</td>
<td>5.6</td>
<td>-2.3</td>
<td>-9.7</td>
<td>-0.6</td>
<td>1.4</td>
<td>0.6</td>
<td>6.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-20</td>
<td>1.2</td>
<td>-0.3</td>
<td>8.9</td>
<td>-3.2</td>
<td>-13.6</td>
<td>-0.7</td>
<td>1.9</td>
<td>0.8</td>
<td>5.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>1.4</td>
<td>1.0</td>
<td>12.1</td>
<td>-3.9</td>
<td>-17.1</td>
<td>-0.3</td>
<td>2.4</td>
<td>0.9</td>
<td>6.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.5</td>
<td>0.4</td>
<td>15.9</td>
<td>-0.1</td>
<td>-4.9</td>
<td>-0.4</td>
<td>3.3</td>
<td>0.9</td>
<td>7.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.5</td>
<td>1.2</td>
<td>20.4</td>
<td>-0.9</td>
<td>-5.9</td>
<td>-0.6</td>
<td>4.2</td>
<td>0.7</td>
<td>8.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.6</td>
<td>1.7</td>
<td>25.0</td>
<td>-1.4</td>
<td>-6.8</td>
<td>-0.3</td>
<td>5.1</td>
<td>0.5</td>
<td>10.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.6</td>
<td>2.7</td>
<td>29.7</td>
<td>-3.2</td>
<td>-6.9</td>
<td>-0.4</td>
<td>5.6</td>
<td>0.0</td>
<td>12.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1.6</td>
<td>3.0</td>
<td>35.1</td>
<td>-3.2</td>
<td>-7.3</td>
<td>-0.7</td>
<td>6.2</td>
<td>-0.6</td>
<td>14.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.6</td>
<td>3.2</td>
<td>41.3</td>
<td>-3.2</td>
<td>-7.6</td>
<td>-0.7</td>
<td>6.4</td>
<td>-0.7</td>
<td>17.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>1.6</td>
<td>3.0</td>
<td>47.4</td>
<td>-2.5</td>
<td>-6.9</td>
<td>-0.7</td>
<td>6.6</td>
<td>-1.8</td>
<td>21.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A long lasting debate of how precise the electrolyte formulation needs to be can be answered using this model. An error propagation estimation can be done when the gradient of a function and the uncertainty of the underlying input is known. From the herein reported measurements, we know the uncertainty of the conductivity and we can easily calculate the gradient of the conductivity w.r.t. the formulation. Here, we take the median uncertainty of the conductivity measurements (\(\Delta \sigma_{\text{exp}} = 3.5272 \cdot 10^{-4} \text{ mS/cm}\)) and divide them by the largest gradient of conductivity w.r.t. to formulation (both uni- and bivariate) at every temperature (Figure 6) to obtain a conservative estimate of the maximally allowed formulation error (eq.2) that would be on the same order like the measurement noise. Unsurprisingly one can have larger errors in solvent-to-co-solvent ratios as in conducting salt-to-solvent ratios. Interesting however is that an error of about 10% in the solvents is acceptable for most temperatures. Dosing of the conducting salt should however be as precise as possible as at high temperatures the error should not exceed 1.5%.

\[
\left( \alpha \left( \frac{\partial \sigma}{\partial r_{PC}} \right)^2 + \beta \left( \frac{\partial \sigma}{\partial r_{LiPF_6}} \right)^2 \right)^{0.5} \cdot e_{r_{PC}, LiPF_6} \leq \Delta \sigma_{\text{exp}} \quad \text{for} \quad \alpha, \beta \in \{(0, 1), (1, 0), (1, 1)\} \quad \text{(eq.2)}
\]
Conclusion
This study shows the utility of active learning to improve model accuracy and precision on complex data with few examples. The pre-shot model significantly underfit the data such that obtained trends did not follow a physically meaningful trend. After one-shot active learning, the discovered model produced smooth cross temperature optima even though it was not trained with inter temperature data. Obtained trends in the optima suggest that for low temperatures, the conducting salt concentration should be minimized whilst for higher temperatures the salt concentration should be increased. These findings are in good agreement with the DHO theory due to the activation energy being on the same order of magnitude as the temperature activation. We find that a globally optimally conducting electrolyte does not exist as those optimal at low temperatures perform poorly at high temperatures. Those electrolytes optimized for near room temperature show ≈20% less conductivity at low and about half the conductivity at high temperatures compared to the range optimized formulations. Through the availability of an easily derivable model, we can discuss electrolyte solvent-conducting salt interactions and find mostly sigmoidal or exponential temperature trends hinting at two different mechanisms. The derivable model also allows an elucidation of maximally allowed formulation errors which lie at ≈10% for solvents and 1.5% for conducting salts at most temperatures. Through the conservative choice of a low degree polynomial model due to scarce data availability we were able to obtain optima and interpretable insights translatable to existing physicochemical laws such as the DHO theory.

Acknowledgements
This work contributes to the research performed at CELEST (Center for Electrochemical Energy Storage Ulm-Karlsruhe) and was funded by the German Research Foundation (DFG) under Project ID 390874152 (POLiS Cluster of Excellence). This project received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957189.

References


