GSoC 2021 Project Proposal

Organization : ML4SCI

Decoding quantum states through Nuclear Magnetic Resonance

MENTORS

Stephen Carr, Vesna Mitrovic, Chandrasekhar Ramanathan, Brad Marston, Charles Snider

AUTHOR

Anantha S Rao

Contents

1 Introduction and Student Information **1**

2 Abstract **1**

3 Technical Details **2**
3.1 Spin Echo in NMR 2
3.2 Modeling electronic and nuclear spins 2
3.3 Evaluation Task Results 3
3.3.1 Model Results 4

4 Proposed Deliverables **5**
4.1 Schedule of Deliverables 6

5 Other Information **8**
5.1 Why ML4SCI? 8
5.2 Relevant Background 8
5.3 Past Experience 8
5.3.1 Knowledge of Libraries and/or Technology stacks 9
5.4 Past Interaction with mentors 9
5.5 Other commitments 9

6 References **10**

1 Introduction and Student Information

- Name: Anantha S Rao
- Email: aanantha.s.rao@gmail.com, anantha.rao@students.iiserpune.ac.in
- GitHub: Anantha-Rao12
- Twitter: anantharao00
- Blog: https://anantha-rao12.github.io/
- Contact no: +91-9481079535 / +91-9665254659
- University: Indian Institute of Science Education and Research (IISER) Pune, India
- CGPA : 8.9/10 (*For the Semesters 1-5*)
- Major: Physics
- Minors: Data Science

2 Abstract

At low temperatures, many materials transition into an electronic phase which cannot be classified as a simple metal or insulator, and quantum phases of matter, like superconductors and spin liquids, are hard to study due to their fragile nature, making non-intrusive and indirect measurements important. Scientists hence use NMR (Nuclear Magnetic resonance) to probe these materials externally. NMR is an experimental technique mainly used in quality control and scientific research to determine the molecular structure, purity, and content of any sample. The GSoC project idea is to explore the connection between electronic phases and nuclei in these materials via simulations of NMR, analyze the time-evolution of nuclear spins in external magnetic pulses, and classify them using suitable Machine Learning models. We also aim to optimize an applied magnetic pulse sequence to best estimate a specific physical parameter from a given material.

3 Technical Details

3.1 Spin Echo in NMR

Magnetic resonance occurs in quantum systems when a magnetic dipole is exposed to two external magnetic fields: [1)]A static magnetic field with another, [2)]oscillating electromagnetic field. The oscillating field can then make the dipole transit between its energy states with a certain probability and when the frequency of the oscillating field leads to the maximum possible transition probability between any two energy states, magnetic resonance is said to have been achieved. In Nuclear Magnetic resonance, this fundamental phenomenon is used to measure various properties of the desired material. [1]
A common technique used to probe molecular properties in NMR is called the "Spin Echo". The nuclear spins are aligned, let loose, and then refocused, making a sharp peak, or "ech", of the original alignment. When the spins interact with each other through spin-spin coupling, the refocused echo in the magnetization can become highly distorted. This technique of studying the magnetization curve to determine electronic phases is used often by scientists and researchers. We want to use simulated Magnetization curves to train a Machine learning/ Deep Learning model to predict the model parameters and furthermore develop an algorithm to optimize the applied pulse sequence.

3.2 Modeling electronic and nuclear spins

Most materials can be classified by their electronic properties into three categories: metal, insulator, and semiconductor. This classification is based on a semi-classical description of the electrons in a crystal where the electrons are treated as a collection of classical particles, with energies that depend on their momentum in a way determined by the atomic structure of the crystal. These electrons interact with the nuclear spins of a material (by way of the hyperfine-interaction) and if the electron-nuclear coupling becomes strong enough (enhanced perhaps by a "quantum" electronic phase), a non-negligible two-step process can couple the nuclei with each other throughout the material. That two-step process is when a nuclear spin couples to an electron and changes its motion, and then that electron later "scatter" off

another nuclear spin elsewhere in the material. We represent this two-step scattering process by way of an effective spin-spin coupling between nuclei at position r_j and r_i, given by:

$$T_0(i,j) = \alpha \exp\left[\left(\frac{|r_j - r_i|}{\xi}\right)^2\right]$$

where α is the coupling strength and ξ is the coupling length. [2] Generally, α and ξ depend on the details of the nuclear-electron coupling and the quantum state of the electrons, but in our work, we will use them to discern if a spin-echo experiment can provide enough information to accurately "reverse engineer" these values from a single M(t) curve. Our simulations also include the dissipation of the nuclear spins wherein the spin information can be "lost" due to couplings with the environment. This occurs at a time scale $T_{decay} \approx \sqrt{\tau}$.

Our goal then will be to develop an ML model that accurately determines the above three variables (α, ξ, τ) from a single M(t) curve.

3.3 Evaluation Task Results

The jupyter-notebooks and results of the evaluation task are summarised in the GitHub repository:Anantha-Rao12/ML4SCI-NMR-evaltasks . In the evaluation task, we looked at several M-curves and built simple ML models to predict the parameters namely: (α, ξ, τ). Few observations from the evaluation task are:

1. All M-curves start with a high initial $|M|$ value and later drop until the 180 pulse that flips the spins.

2. Almost all the spins have low $|M|$ values between timestamp: (180-300) until a peak that corresponds to the spin-echo and then again the magnetization decreases. The peak is replaced by a trough for M-curves with strong spin-spin coupling.

3. We found that the distribution of $|M|$ during the 180 pulse and Spin echo phase is bimodal in nature and is amenable to classification studies (based on type of electronic interaction).

3.3.1 Model Results

1. The Magnetization curves dataset was split into (70-30)% as training and testing datasets. The training dataset was preprocessed and normalized with MinMaxScaler to speed up learning, while taking care of outliers. The result of training of various models are:

Table 1: Machine Learning Model results

Sr no	**ML model**	**Training Accuracy (%)**	**Testing Accuracy (%)**
1	Linear Regression	98.6	97.1
2	Ridge regression	98.6	97.1
3	RandomForest regression	98.1	86.1
4	KNN regression	82.4	72.9

2. We also looked at a simple example of regression with Artificial Neural networks with connected dense layers using the Keras API. Since, this exercise is part of a more general assessment task of predicting NMR parameters using Machine Learning, the model architecture and other statistical nuances were ignored. The percentage accuracy (90%) and mean-squared-error (6.20) of the model is nowhere near current industry standards, so we can implement certain steps to improve the model performance. Some of them are :

- Hyperparameter tuning with the Neural Network
  - We can vary the number of layers in the Dense network, the number of neurons in each layer and tune additional parameters like Batch Normalization, Max pooling, Dropout and analyse its implications on the final model. A more general approach can be to decrease the bias and variance of the model by employing more complex models and regularization techniques repectively.
- Regression after classification
  - Since the distribution of data points near the 180 pulse and the spin echo was observed to be bimodal, we can explore if the dataset can be partitioned and represented by two or more labels. This can be based on the low and large values

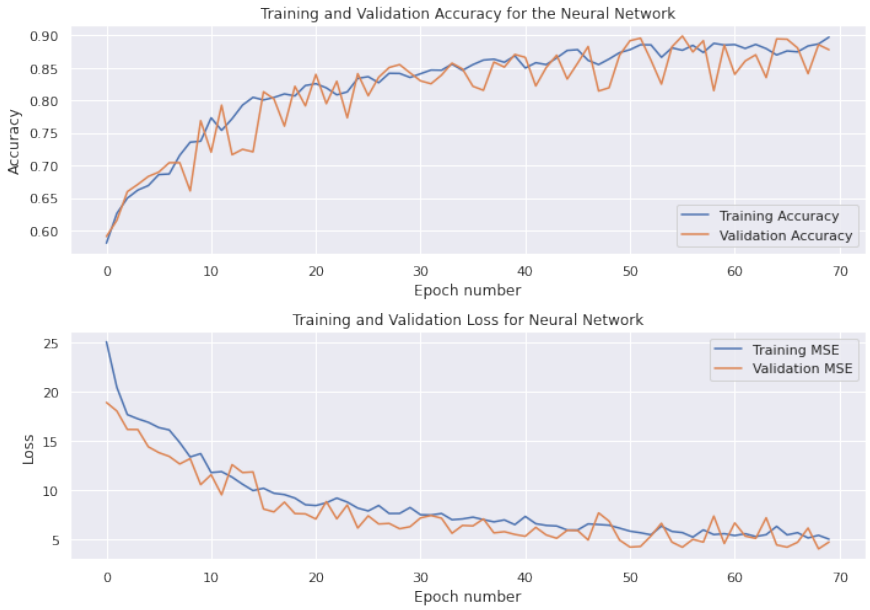

Figure 1: Results of Training a simple Dense Neural Network for parameter prediction of Magnetization curves [Link to the notebook]

of the Spin echo at around timestamp 300-350 and its correlation with the initial 180 pulse. Regression might work well with such a dataset where separate models are built for both/all the cases.

4 Proposed Deliverables

1. Implement a classification model that determines the type of electronic interaction based on only the time-dependent curve. Determine the sensitivity of classification to noise.

2. Utilize a neural network to predict the strength, range, and dissipation parameters of a given magnetization curve.

3. Develop an algorithm which optimizes an applied pulse sequence to best estimate a specific physical parameter from a given material.

4.1 Schedule of Deliverables

1. **Community Bonding Period: June 1, 2021 - June 7, 2021**
   During the community bonding period, I will look into the relevant literature useful for the project and brush up on relevant physics-based concepts. I have basic knowledge of NMR and will spend this time understanding the nuances of M-curves and producing minimal working code to motivate the methods under development. I will also try as much as possible to get to know the ML4SCI developer community, interact with them and head start a great journey.

2. **Week 1**
   Work on developing a classification model to determine the type of electronic interaction. In the evaluation task, it was observed that the distribution of average magnetization near the 180 pulse and spin echo timesteps is bimodal. This serves as a caveat to bin each M-curve as high/low and then start with a simple logistic model for classification.

3. **Week 2**
   Adding noise during training can make the training process more robust and reduce generalization error. Here, I plan to work on analysing the effect of noise on prediction. One possible way to mitigate noise errors would be to work with the fourier transform of the signals and training the model in fourier space after noise-removal. We will quantify and report sensitivity to noise using metrics like AUC, ROC and confusion matrices.

4. **Week 3**
   Optimize and build other classification models (K-NearestNeighbours, Decision-Trees, Random Forest, Support Vector Machines etc) and benchmark the performance of each model and effect of noise on them.

5. **Week 4**
   Work on developing and optimizing (hyperparameter tuning) an artificial neural network to predict the strength, range, and dissipation parameters of a given magnetization

curve. We continue work from the evaluation task and will use techniques like Transfer learning [3] to attain state-of-the-art accuracies and prediction-levels.

6. **Week 5**
   Complete all previous tasks. This is a buffer week for any unprecedented delays. Publish blog post. Prepare for Phase 1 Evaluation.

Phase 1 Evaluation

7. **Week 6**
   Work on developing an algorithm which optimizes an applied pulse sequence to best estimate a specific physical parameter from a given material. Here we will look at the layers of a neural network and analyse the gradients or important features of the model to determine what aspects of the experimental signal are most important.

8. **Week 7 and 8**
   Work on integrating the three proposed methods and test them. Complete documentation for newly built methods, verify results, fix bugs (if any) and write additional unit tests.

9. **Week 9 and 10**
   Complete Jupyter Notebook Tutorials for all the proposed methods and modifications. Publish blog posts and prepare for Phase 2 Evaluation.

Phase 2 Evaluation

10. **Post GSoC and Future Work**
    After the proposed 10-week timeline, I would love to start implementing any additional features and contribute to ML4SCI even after GSoC and given an opportunity, would love to participate in scientific research with the mentors.

5 Other Information

5.1 Why ML4SCI?

I believe that Python is currently the juggernaut in science and engineering and the current leader in ML, AI, Pharmaceuticals, etc. ML4SCI being the organization dedicated to using machine learning in the basic sciences, I believe that as a Physics major (and Data Science minor), my background and interest in the fields of machine learning, artificial intelligence, and quantum mechanics can complements well with this GSoC summer project. Since understanding quantum states is a central theme in many areas of research (condensed matter physics, atomic and molecular physics, quantum computing, etc), I believe that the broader impact of my work with the additional feature of using AI will greatly help researchers and practitioners to explore a plethora of other phenomena and moreover significantly simplify their work. Lastly, as an advocate for open-source, it would indeed be an absolute honor to contribute these tools for open science and society.

5.2 Relevant Background

5.3 Past Experience

I have a reasonably good background in machine learning and quantum mechanics. I am an advanced Python user with 3 years of experience in working with the design, development, and deployment of Machine learning and Deep Learning models and currently work as an undergraduate researcher at the Complex Systems and Quantum Information Lab at IISER Pune, India. I have previously contributed to many open source organizations, namely:

1. JuliaDynamics : ChaosTools.jl (Contributor to Version : 1.26) (#165)
2. iGEM Software : IISER-Pune-India (Main contributor and Maintainer)

and am the author of many more open-source projects on machine learning, bioinformatics and quantum information. At iGEM (International Genetically Engineered Machine), I developed DeLeMa-detect (Deep Learning for Malaria Detection) and PACMal (Peptides

Against Cerebral malaria), a comprehensive suite of solutions against malaria. Currently, I am working on developing quantum-machine-learning models to understand many-body localization and quantum chaos under Prof.MS Santhanam at IISER Pune. I also have basic knowledge of Nuclear Magnetic Resonance and its associated experimental details from a previous project with Prof TS Mahesh at IISER,Pune. I am confident that my skills and passion will allow me to uphold the status of the work expected by me.

5.3.1 Knowledge of Libraries and/or Technology stacks

As for the familiarity with technology stacks, I have extensively used Google's Tensorflow framework for projects related to deep learning. Keras and scikit-learn are my go-to Deep/ Machine learning libraries. Numpy is the library that I have used most in my projects. It is a library that I often use for school work as well (almost every week). Since all the above mentioned libraries are written in Python, I am fluent with Python programming in general. I think I will have to put in extra effort during the project to learn about the experimental aspects of NMR (which I am more than happy to) since the project will certainly require a strong knowledge base in that area as well.

5.4 Past Interaction with mentors

I have been in contact with Dr.Carr via email shortly after ML4SCI was announced as one of the participating organizations. Since I was interested in a project at the intersection of quantum physics, machine learning and because I have some previous experience with NMR, I emailed the mentors about my interest to integrate machine learning to understand the quantum phases of matter. Dr.Carr has advised me well during the evaluation tasks and also provided adequate feedback on my draft proposal.

5.5 Other commitments

My third-year final exams get over by the end of May and post-that, I will be able to spend 25-30 hours per week throughout the summer on this project. I am crediting an experimental

physics course at my university in the summer that requires me to spend 3-5 hours/week, but the course has not been confirmed yet due to the uncertainties with the pandemic situation in India. A reduction in working hours in a week because of unavoidable conditions will be surely compensated later in the following weeks. In such circumstances, I will communicate with my project mentor Dr.Stephen Carr well in advance. For project catch-up, I will be available via Skype/Zoom in Indian Standard Time Zone, GMT +5:30.

6 References

References

[1] Claude Cohen-Tannoudji, Bernard Diu, and Franck Laloe. Quantum mechanics. vol. 1. 4. rev. upd. ed.; quantenmechanik. bd. 1. 2009.

[2] Piotr Tekely. Malcolm h. levitt. spin dynamics: basics of nuclear magnetic resonance. john wiley & sons, chichester, uk, 2001, 686 pp. price£ 34.95. isbn 0-471-48921-2, 2002.

[3] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.