#1 solution: *ape-MPNN*

Approach keywords: MPNN, Attention, ACSF

Introduction

This project was done in Nanjing University, specifically by the collaboration between two interdisciplinary members, one at the School of Chemistry and Chemical Engineering, Institute of Theoretical and Computational Chemistry (ITCC) and the other at the Department of Computer Science and Technology, so our team consists of both Machine Learning background and domain knowledge.

Feature selections

The design of molecular representations is a key problem in chemistry and materials virtual screening. We choose our representations mainly according to ref [1,2,3] as show below. For atom features, we include information of atom type, *chirality, formal charge, partial charge, ring sized, hybridization, hydrogen bonding, aromaticity, van der waals radius, valence of out shell, radius of bond and node degree* with total size of 38 and *bond type, topological path length, shortest path bonds, atoms in the same ring, graph distance, extended distance and geometric distance* for bond features, respectively. We also apply the Atom-centered Symmetry Functions (ACSFs) [4] to our later model.

Model selections & Overall architecture

We made three major boosts on leaderboard, corresponding to three key models of our work.

(1) ape-MPNN: we designed a new neural network architecture, namely Attention Pooling Embedded MPNN *(ape-MPNN),* with the inspiration of the visual attention mechanism [5] to absorb the enriched heterocyclic chemical features and extract quantum chemistry knowledge to fill the gap between simple structure information and the target electronic structure properties. The inner GRU block of *MPNN* was replaced with LSTM.

(2) *ACSF regularized ape-MPNN:* Based on ape-MPNN, we introduced new descriptors ACSF [4] as the regularizer to the model to pay attention to the more accurate symmetric atomic environment attributes [6], thus to dive deeper into the subtle chemistry understanding of atom bonding, interatomic potentials and long term effects of graph node neighbors. This model help us to be the top 1 solution for a long time.

(3) Weighted ACSF regularized ape-MPNN: In this period, we assigned different loss weight to different molecules according to their heavy atom number, so the heavier molecules should gain a higher weight since all the test data are bigger atoms.

Computation resource & Hyper-parameter tuning:

We only have one GTX1050Ti for features processing and early models and one RTX2080 Ti for later models. Due to very limited computing resources, we did not make too much concerns on tuning the parameter, but add more chemistry knowledge into the neural network with better architecture design to make ends meet.

Training details:

All models are trained using ADAM optimizer with the default parameter setting. We trained the ape-MPNN (attention pooling embedded MPNN) model from random parameters with decreased learning rate step by step, which takes about 4 days for a RTX 2080 Ti GPU. Due to computation limitation, we train the ACSF-regularizer and weighted-learning model from pre-trained ape-MPNN model of 120 epochs with a learning rate of 5e-4 for 100 epochs. We obtain our final model by ensembling all the three models through a weighted mean method with the help of valid dataset.

What didn’t work

1 Dropout: Wherever we employ the dropout and whatever small the dropout rate it is, the model will blow up and fail to converge.
2 Sequentially combining GAT and Transformer: Intuitively those models should help a lot based on the MPNN to achieve higher score, however they failed.
3 Add global features with META-NN: We used to believe that global features will help the MPNN model to capture some information of the molecule for improvement, but they don’t.

Acknowledgements

We thank the stimulating discussions with Professor Jing Ma and Professor Yanwen Guo for instruction.

Reference:

1. Kearnes, Steven , et al. Molecular graph convolutions: moving beyond fingerprints. Journal of Computer-Aided Molecular Design 30.8(2016):595-608
2. Chen, Chi , et al. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. (2018).
3. Chen, Benson , R. Barzilay , and T. Jaakkola . Path-Augmented Graph Transformer Network. (2019).
4. J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, J. Chem. Phys. 134 (7) (2011) 074106.
5. Kevin Xu, et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. (2015)
6. Abolghasemi, Pooya, et al. Pay attention! - Robustifying a Deep Visuomotor Policy through Task-Focused Attention. (2018).