Predictive Modeling of the Severity/Progression of Alzheimer's Diseases

Why:

1. Classify the progression of AD
2. Create self-measured method for the convenience of patients
3. Point out the factors which impact the results will be determined.

What:

1. Dataset: NACC dataset
2. Label: CDRGLOB
3. Model 1: complete dataset with all patients and all columns
4. Model 2: complete dataset with all patients but no CDR and GDS score. (for the convenience of patients to measure at home), but I don't think GDS score should be removed
5. Model 3: only focus on those patients whose CDRGLOB is not that severe (remove patients with CDRGLOB 2 and 3). Also, remove CDR and GDS score.
6. Measurement: accuracy

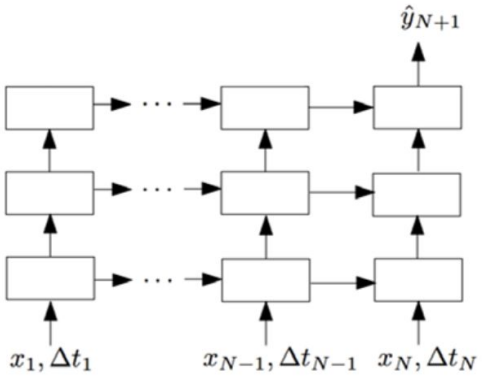
Early Detection Models for Persons with Probable Alzheimer's Disease with Deep Learning

Why:

1. Predict the AD's progression for the patients on early stages
2. Create self-measured method for the convenience of patients

What:

1. Dataset: NACC dataset
2. Preprocess: (1) since the model is for the early stage, so they only use those patients whose CDRGLOB is not that severe (remove patients with CDRGLOB 2 and 3). (2) Remove CDR and GDS score. (3) only keep focus on 17 more significant features about demographics, medical history and FAQ, listed in the paper. (4) remove those patients whose number of visits is less than 2. (But this model will lose the ability to do the early detect for those patients who only have one record).
3. model:



Predictive Modeling of the Progression of Alzheimer's Disease with Recurrent Neural Networks

Why:

1. Predict the AD progression of the next visit based on all the previous visit records
2. Consider uneven time interval between different visits when do the prediction

What:

1. Dataset: NACC
2. Label: CDRGLOB
3. Preprocess: (1) only keep those patients who are diagnosed as probable AD and have more than three visits. (2) filling the missing data: categorical: mode, continuous: mean, ordinal: median. (3) normalize the continuous data
4. Model: 2 layer LSTM. Loss function: Cross Entropy with L2 regularization.
5. Baseline model: LR, SVM, DT, RF.
6. Measurement: Accuracy, PPIA, SPIA

Conclusion:

1. For Baseline, aggregated features is much better than last visit or last k visits.
2. For LSTM, with time info is better than the model without time info
3. LSTM model outperforms all the baseline model.
4. CDR and FAQ data matters (I suspect that model can only cheat on CDR to predict CDRGLOB)
5. The features in CDR contributes more than those in FAQ.

Predicting Alzheimer's disease progression using multi-modal deep learning approach

Why:

1. Predict conversion from MCI to probable AD or to normal for those persons who are already MCI in the baseline visit (last visit in the history window)
2. Use multi-modal method to predict the progression

What:

1. Dataset: demographic, CSF, image, cognitive performance from ADNI
2. Task: predict MCI to AD or NC after δt time from the baseline visit (the last one in history window).
3. Model: GRU with all time series and multi-modal data.
4. Comparison: 1. Only use the baseline visit in the history window. 2. Only use single-modal data.
5. Measurement: Accuracy and AUC curve
6. Loss function: Cross-entropy with L1 regularization
7. Two steps in training: (1) predict MCI → AD or NC for GRU of each single-modal data to get the feature extractor for each kind of data. (2) fix GRU, and use AD-NC patients to pre-train the classifier LR to get the pre-trained latent threshold of AD and NC. (2) fix GRU and use LR to get the classifier to predict the final results (They cannot investigate the integrated information between different modal data).

Big Data Analytical Approaches to the NACC Dataset: Aiding Preclinical Trial Enrichment

Why:

Predict those ones who will convert from normal to MCI in the follow-up 4 years.

What:

1. Dataset: NACC
2. Preprocess: cognitively normal at baseline visit, and have three or more visits
3. Measurement: Specificity and ROC-AUC
4. Model: SVM, LR, RF
5. **Question: what the hell is baseline visit??**

Early Diagnosis of Alzheimer's Disease Using Informative Features of Clinical Data

Why:

Diagnose AD at the early stage, which is considered as a classification problem.

What:

1. Dataset: clinician diagnosis and clinical judgment columns in NACC
2. Pipeline:

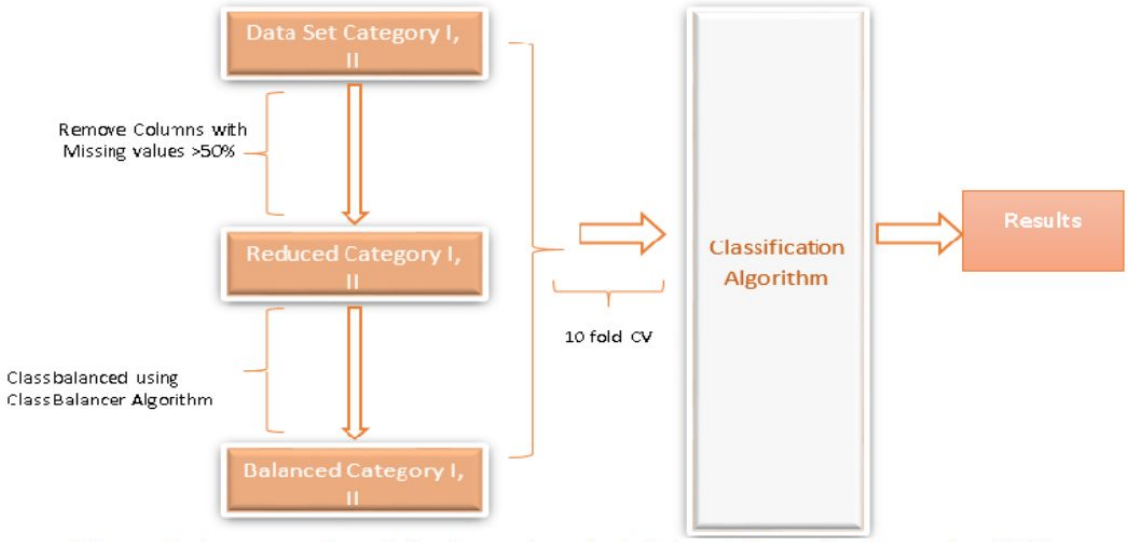


Figure 1. A proposed model using categorical data set for early diagnosis of AD.

3. Measurement: Accuracy, precision, AUC and training time
4. Conclusion: Naive Bayes is the classification model for this problem.

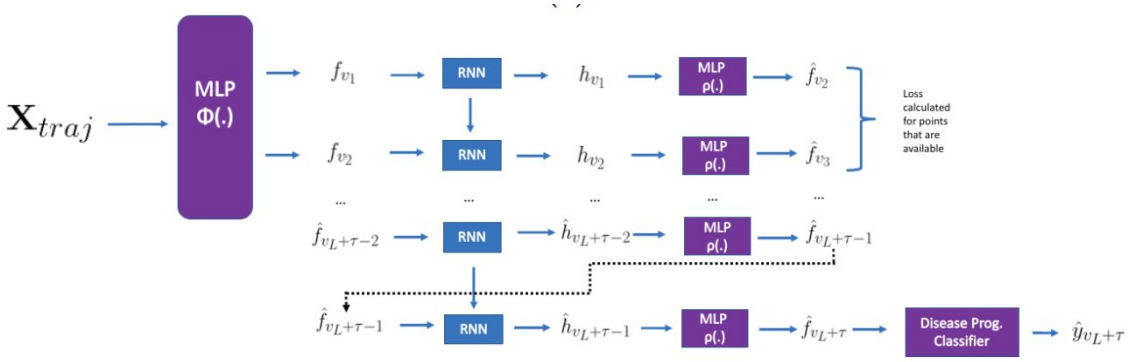
FLARe: Forecasting by Learning Anticipated Representations

Why:

1. Predict the cognitive status as AD, MCI, or normal for the $t = \tao$ later.
2. There are two drawbacks for the traditional RNN-concat model: 1. In the time interval between the last visit in the trajectory and the to-be-predicted time, it cannot take the time information into account for the forecasting. 2. The irregular time interval between different visits in the trajectory is not considered.

What:

1. Dataset: ADNI, using demographics, cognitive test score, and image features
2. Model:



3. Model illustration: (1) for each time step, there are different multilayer encoders MLP for different data categories. Then the input for RNN is the concatenation of output of different MLP. (2) After RNN, the MLP is to get the hidden representation of the next step, and the unsupervised axillary loss is applied between the predicted hidden representation and the true representation of the next step. (3) in the forecasting procedure, since there is no true representation for the next step, then it feeds the hidden representation into RNN directly. (4) finally, there is a multi-layered classifier for the classification for the visit $t=\tao$.
4. Other techniques used in this paper: data augmentation for both training and test dataset, they set an additional parameter T as the number of visits in the trajectory.

Personal Thoughts:

1. Maybe we can use the label for each visit in the history to do the supervised learning
2. How to encode different time interval between different visits in the trajectory? (1) concatenate time info (2) refer to the positional encoding as in Transformer paper.
3. In this paper, they used ADNI dataset, whose time interval between different visits are fixed, 6 months. I think NACC dataset is more challenging since it is different and cannot specify the fixed time interval for the forecasting. So we need a stronger time information encoder to solve this problem.

ShortFuse: Biomedical Time Series Representations in the Presence of Structured Information

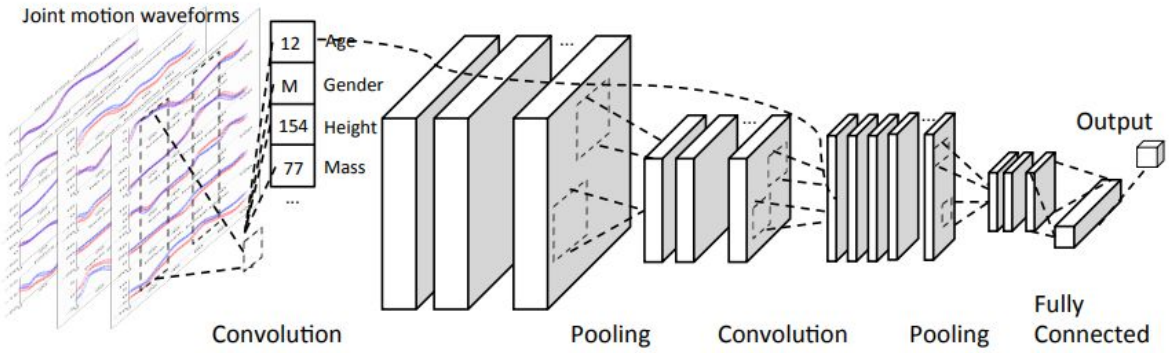
Why:

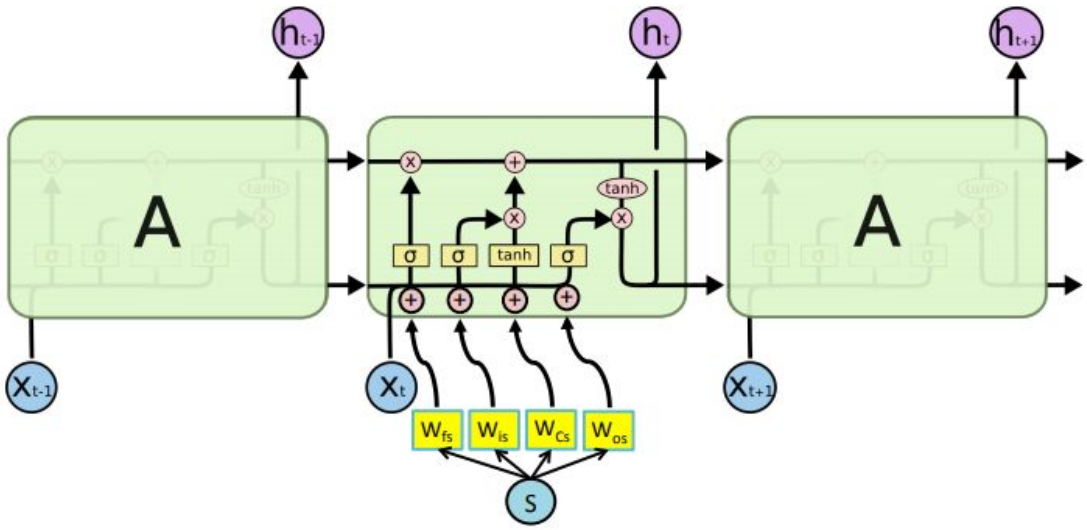
1. Structure covariants information is important for accurate prediction along with the time series data.
2. Since covariants stay the same during different time steps, it should not introduce different parameters for different time steps.
3. It is useful to use the structure covariants to guide the feature extractor for the time series data

What:

1. Two tasks: (1) Forecasting osteoarthritis progression (2) Predicting the outcome of surgery in patients with cerebral palsy

2. Proposed models: shortfused CNN, shortfused LSTM and latefused CNN





3. Model illustration: (1) short fuse CNN: add structured covariants as an additional channel for the input time series data, note that for along the temporal axis, the attached structured covariants are the same to guarantee that there is no additional parameter. (2) short fuse LSTM: the structured covariants control three gates and output feature generation. (3) latefuse: features from CNN on time series data and features from CNN on structured data just merge before the classifier.
4. Conclusion: (1) structured covariant data helps the prediction (2) early fuse model is better than the later fuse model.

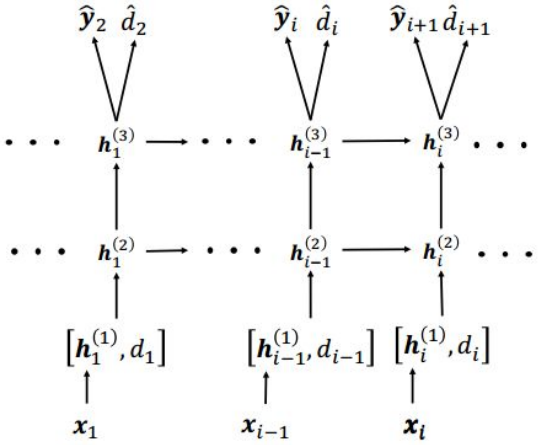
[Doctor AI: Predicting Clinical Events via Recurrent Neural Networks]

Why:

1. Predict the physician diagnosis and medication order of the next visit based on the previous records
2. Predict the time interval between this visit and next visit

What:

1. Dataset: (1) private Sutter Health dataset (2) MIMIC.
2. Data process: (1) only keep those patients with more than two visits (2) group ICD codes into broader categories.
3. Model:



4. Model illustration: (1) the paper used two-layer GRU as sequential modeling. (2) the input into each time step is the concatenation of features and the time gap between this visit and the next one. (2) the output for each time step is one for the predicted label at **NEXT** time and one for the predicted time gap between two visits. (3) between the raw input and the GRU, there is a linear layer to project it onto a latent space with lower dimension. (4) the classification layer is two linear layers, one for the label, one for the time gap. There is a ReLU to guarantee the output time is positive.(5) the loss function is cross entropy loss for the label and MSE for the visit time. (6) dropout is used in the classifier and different GRU layers. L2 regularization is also used.
5. Model pretrain: they pretrain the linear layer between the raw input and the GRU using Skip-Gram model.
6. Measurement: top 30 @ recall for the classification and R^{2} for the time prediction.
7. Ablation studies: (1) explore the performance vs. visit times (2) explore the transferability of the model.
8. Conclusion: (1) RNN outperforms MLP and LR (2) RNN with Skip-Gram pre-training is better than the linear embedding.

Personal Thoughts:

1. What about replacing the predicted time task with diagnosis task?
2. What about feeding the output of each status representation into the next step?
3. What about using attentional model for it?
4. What about using Transformer model for it?