Few-Shot Learning with Similar Tasks via MOE for Mortality Prediction of Diverse Rare Diseases

Wu Haoxian (A0225199W), Chen Nan (A0224882Y), Wang Ruixiao (A0225545E), Ye Ruizhe (A0225250U)

Abstract

With the vast volume of Electronic Health Records(EHR) available online, the past few years have witnessed the profound impact of DL applying in healthcare application. EHR systems store data associated with each individual patients' medical chronological journey, including demographic information, diagnoses, medications, laboratory tests and results, medical images, clinical notes, and so on.

Considering the high dimensionality and temporality of medical journey, deep sequential model have been widely used for diverse applications, such as medical concept extraction, patient trajectory modelling, disease prediction, mortality prediction, to name a few.

Mortality prediction of diverse rare diseases using electronic health record (EHR) data is a crucial task for intelligent healthcare. However, data insufficiency and the clinical diversity of rare diseases make it hard for directly training deep learning models on individual data. Mortality prediction for these different diseases with insufficient samples can be viewed as a few-shot multi-task learning problem. Though different diseases have distinct medical inherence, common properties may be shared among some of them which can be considered as task similarity. By modeling both the task specificity and similarity, data insufficiency problem can be mitigated to make a more accurate prediction for patients with certain disease.

In our project, we leverage Mixture of Experts(MOE) to capture task similarity and specificity on deep sequential model for 24-hour earlier mortality prediction.

Dataset

eICU is populated with data from a combination of many critical care units throughout the continental United States between 2014 and 2015. For each patient, we select events occurred 24 hours earlier before unit discharge with counterpart features and arrange the events in the temporal order. We computed the frequency of each event among unit charging, remaining events with frequency more than 80%. The type of the events is listed below.

- lab event: MCHC, WBC x 1000, Hct, MCH, platelets x 1000, RBC, Hgb, RDW, MCV, bicarbonate, glucose, BUN, potassium, creatinine, calcium, sodium, chloride,
- vital sign: systolic, diastolic, noninvasivemean, sao2, heartrate, respiration

For each ICD (International Classification of Diagnose x) code in eICU, we calculate its sample size (i.e. the number of patients with this code). Then we select 251 ICD codes with less than 200 samples in eICU as rare diseases.

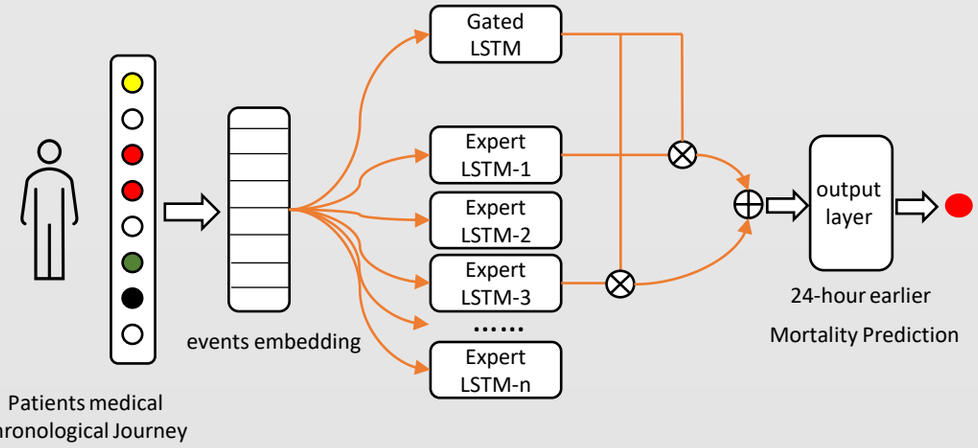
The description of 10 icdcodes with highest mortality ratio is summarized as the table below.

icd_string	mortality_ratio
neurologic|disorders of vasculature|increased intracranial pressure| with herniation syndrome	0.4174
renal|abnormality of urine quantity or quality|diabetes insipidus|central	0.3889
hematology|coagulation disorders|DIC syndrome	0.303
cardiovascular|ventricular disorders|cardiomyopathy|dilated|ethanol related	0.2857
pulmonary|disorders of lung parenchyma|interstitial lung disease| ILD associated with collagen vascular disease	0.2857
infectious diseases|head and neck infections|oral herpes	0.2857
transplant|s/p bone marrow transplant|s/p bone marrow transplant	0.28
pulmonary|pulmonary infections|pneumonia|opportunistic|fungal	0.2667
burns/trauma|trauma - CNS|fracture of skull|vault|closed	0.25
pulmonary|pulmonary infections|pneumonia|opportunistic|PCP	0.25

Methodology

We take LSTM as both gates and experts model. Individual chronological events are used by gated-LSTM to determine the latent disease property. Each expert-LSTM is a specific model for mortality prediction of one task. The weighted-sum of outputs of each expert are fed to the final output layer to generate the mortality signal in the next 24 hour.

The Overall model architecture is shown as the figure below.



Experiment

Within each icdcode, split train, test with ratio 4:1 for alive and expired samples respectively, 1/3 of train samples are further separated into validation.

The overall dataset is the combination of samples of all icdcodes for train, validation, test

Set learning-rate=0.0005, epoch_num=15, ran 5 times(seed 1024, 2048, 4096, 8192, 16384), dropout=0.3

For resource limitation, we set 200 as window size, and pad 0 at last if the length of chronological journey is shorter than 200.

Baseline model: treat all icds as one single task

LSTM, Transformer

MOE(10-experts) – LSTM

Assume there are 10 inherences among diseases, 10 tasks

Model	Metrics seed - [Test_loss, ROC_AUC, PRC_AUC]	Cost
MOE(10)-LSTM	1024 - [0.3029, 0.7640, 0.3341] 2048 - [0.2973, 0.7842, 0.3673] 4096 - [0.3047, 0.7641, 0.3732] 8192 - [0.3026, 0.7519, 0.3543] 16384 - [0.2949, 0.7712, 0.3829]	5s/step
Transformer	1024 - [0.3113, 0.7432, 0.2934] 2048 - [0.3193, 0.7478, 0.3053] 4096 - [0.3182, 0.7259, 0.2983] 8192 - [0.3113, 0.7292, 0.2841] 16384 - [0.3102, 0.7522, 0.3318]	230ms/step
LSTM	1024 - [0.3482, 0.6022, 0.1598] 2048 - [0.3443, 0.6536, 0.1792] 4096 - [0.3323, 0.6722, 0.2446] 8192 - [0.3388, 0.6405, 0.1843] 16384 - [0.3423, 0.6301, 0.1870]	610ms/step