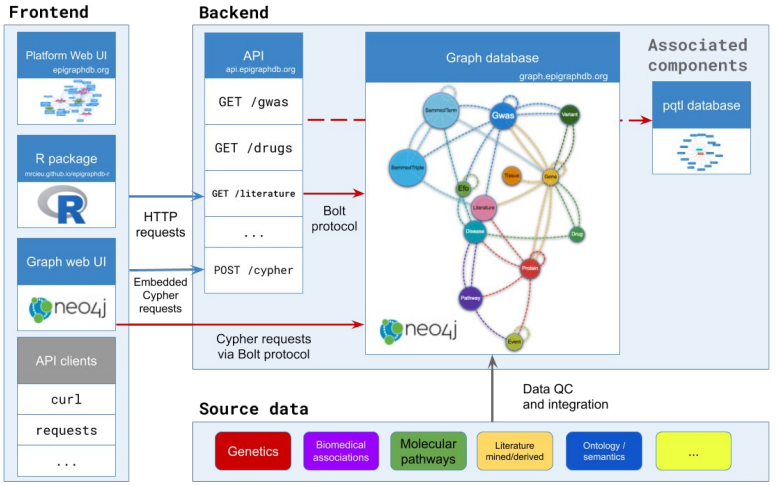
The Vectology platform, and phenotype mapping via ontology

2021-01-29 IEU data science talk

Yi Liu

whoami

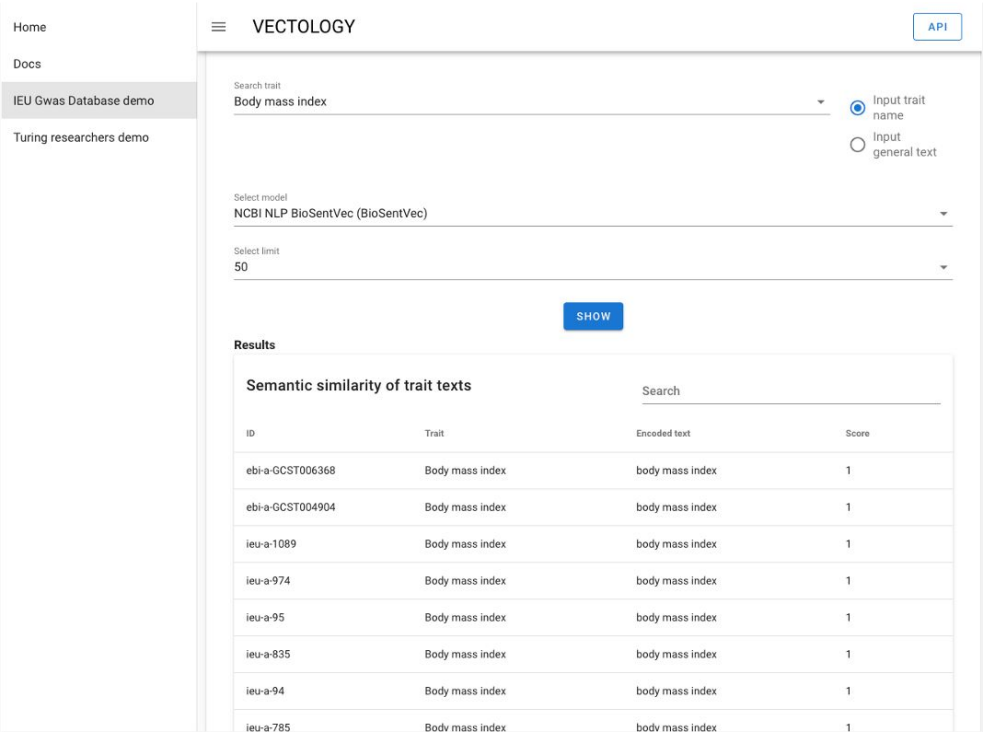
- Yi Liu, https://yiliu6240.github.io
- Senior research associate in health data science, IEU Programme 4
- EpiGraphDB https://epigraphdb.org development
- Currently working on natural language processing studies
- Register your software with https://mrcieu.github.io !
- Was co-organising the Bioinformatics reading group

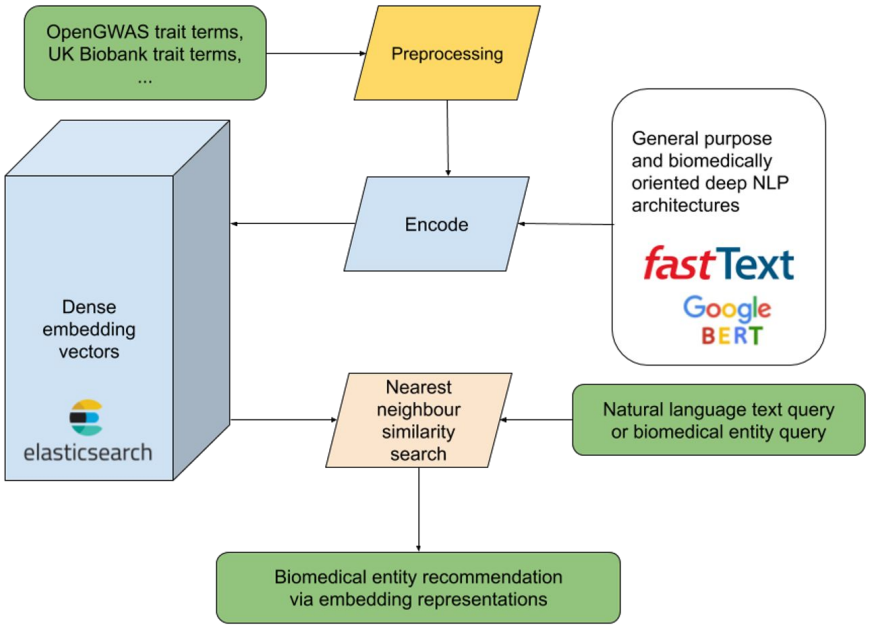


Outline

- The Vectology platform
  - Mapping biomedical entities via their vector representations
  - Mapping biomedical entities via ontology representation
- Technologies
  - Platform architecture
  - Finetuning a transformer model
  - Serving the model

Vectology platform, http://vectology.mrcieu.ac.uk





Vectology as a GWAS trait recommender

Coronary heart disease

Dataset: ieu-a-9

Download VCF | Download index | View report

PMID	23202125
Year	2013
Category	Disease
Sub category	Cardiovascular
Population	Mixed
Sex	Males and Females
ncase	63,746
ncontrol	130,681
Sample size	194,427
Number of SNPs	79,129
Unit	log odds
Priority	1

Top 30 related datasets

- ebi-a-GCST000998: Coronary heart disease
- ieu-a-7: Coronary heart disease
- ieu-a-6: Coronary heart disease
- ieu-a-8: Coronary heart disease
- ukb-d-I9_CHD: Major coronary heart disease event
- ukb-a-534: Diagnoses - main ICD10: I25 Chronic ischaemic heart disease
- ukb-d-I9_CHD_NOREV: Major coronary heart disease event excluding revascularizations
- ebi-a-GCST003116: Coronary artery disease
- ebi-a-GCST005194: Coronary artery disease
- ebi-a-GCST005195: Coronary artery disease
- ukb-b-16606: Diagnoses - secondary ICD10: I25.8 Other forms of chronic

Trait recommender via text vector representations

- Deep learning based trained models with biomedical domain focus
- Encode a trait to a high dimensional embedding (e.g. 1x768)
- Index these dense vectors in Elasticsearch
- Compute pairwise cosine similarity of two trait embeddings
- Recommend traits via nearest neighbour search

- Shortcomings:
  - Blackbox
  - Reliant on embedded knowledge representations

Model	BIOSSES	MEDSTS
BioSentVec (PubMed)	0.817	0.750
BLUEBERT (PubMed, base, uncased)	0.848	89.3

transformers

- seq2seq (“attention is all you need”), BERT, GPT-{1,2,3}, etc
- multi-headed self-attention, transformer encoders + decoders
- pretrained language models with massive natural language corpus
- finetuning for a downstream task
  - sentiment analysis (sequence classification)
  - named entity recognition, relationship extraction (token classification)
  - question answering
  - text generation
- BERT: transformer pretrained language model on Wikipedia + BookCorpus datasets
- BLUEBERT: further pretraining on BERT using biomedical datasets PubMed abstracts + MIMIC clinical notes

Trait recommender via ontology mapping

- Use task specific knowledge to reinforce entity representations
- Train (finetune) a biomedically general purpose Transformer model (BLUEBERT)
- Training task: learn the relative distance of ontology terms
- Inference: mapping traits as if they are new nodes in the ontology graph

Experimental Factor Ontology (EFO)

coronary artery disease

Search EFO | Search

http://www.ebi.ac.uk/efo/EFO_0001645 Copy

Narrowing of the coronary arteries due to fatty deposits inside the arterial walls. The diagnostic criteria may include documented history of any of the following: documented coronary artery stenosis greater than or equal to 50% (by cardiac catheterization or other modality of direct imaging of the coronary arteries); previous coronary artery bypass surgery (CABG); previous percutaneous coronary intervention (PCI); previous myocardial infarction. (ACC) [ ]

Synonyms: `coronary artery disease` `Artery Disease, Coronary` `Coronary Heart Diseases` `Diseases, Coronary Heart` `Coronary Arteriosclerosis` `Arteriosclerosis, Coronary` `Diseases, Coronary` `Coronary Atherosclerоses` `disorder of coronary artery` `CORONARY HEART DIS` `Atherosclerоses, Coronary` `Atherosclerosis, Coronary` `Arteriosclerоses, Coronary` `coronary disease` `Coronary Diseases` `Coronary Atherosclerosis` `Coronary Artery Diseases` `Coronary Disease` `coronary artery disease or disorder` `CORONARY DIS` `Artery Diseases, Coronary` `Disease, Coronary Artery` `Diseases, Coronary Artery` `CHD (coronary heart disease)` `Heart Diseases, Coronary` `CAD` `Disease, Coronary` `Coronary Artery Disease` `Disease, Coronary Heart` `Coronary Arteriosclerоses` `disease or disorder of coronary artery` `coronary heart disease` `disease of coronary artery` `CORONARY ARTERY DIS` `CHD - Coronary heart disease` `CHD` `Heart Disease, Coronary`

Tree view | Term mappings | Term history

Graph view | Reset tree | Show all siblings

- experimental factor
  - material property
    - disposition
      - disease
        - disease by anatomical system
          - cardiovascular disease
            - heart disease
              - coronary artery disease
            - vascular disease
              - arterial disorder
                - coronary artery disease
        - disorder by anatomical region
          - thoracic disease
            - heart disease
              - coronary artery disease

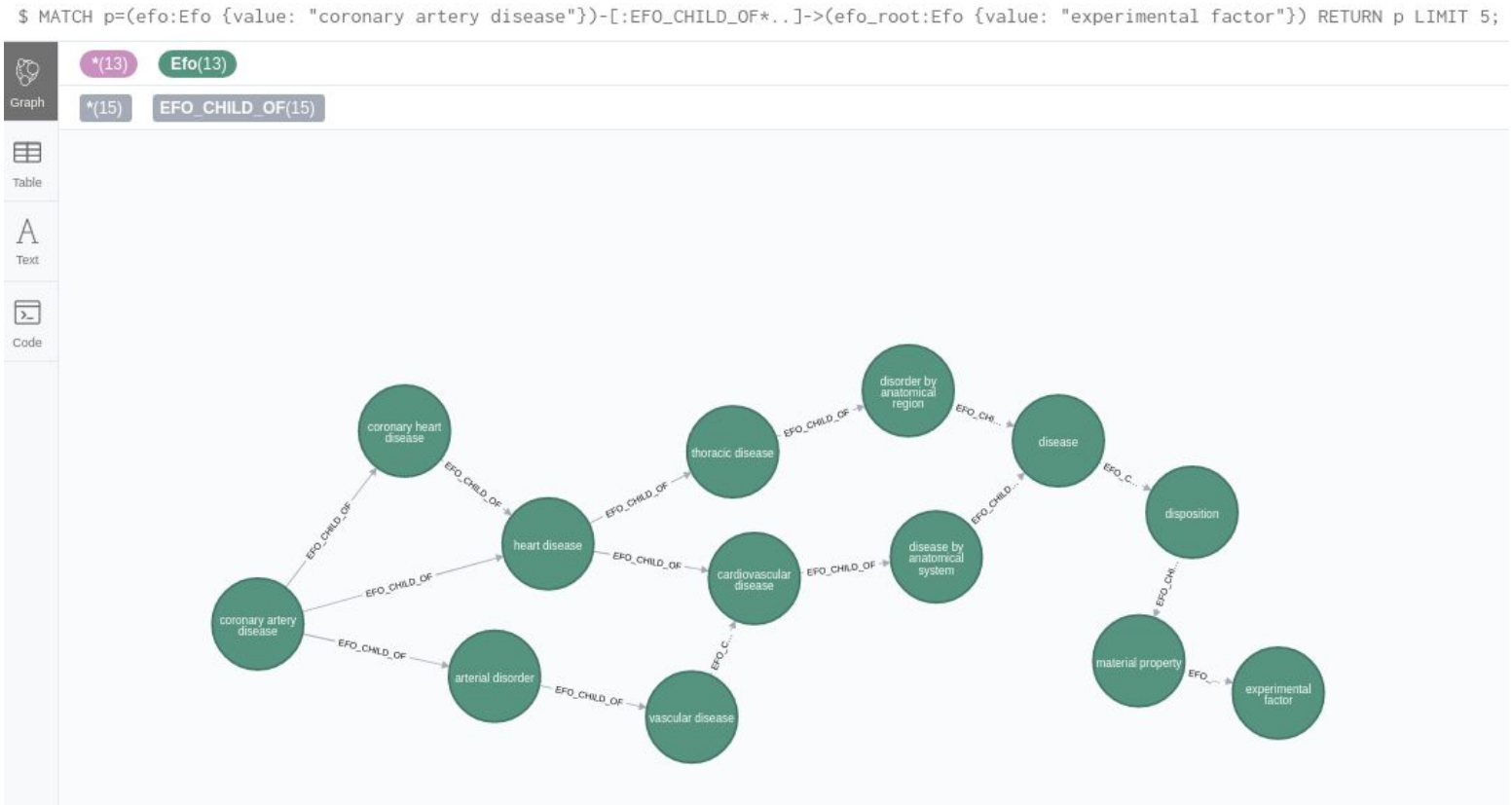
Term information

database cross reference

- ICD10:I25.10 (DOID:3393)
- UMLS:C1956346 (NCIT:C26732)
- GARD:0011944 (MONDO:equivalentTo)
- OMIM:610938 (http://www.ebi.ac.uk/ontology/webulous#OPPL_pattern)
- SNOMEDCT:53741008
- ICD10:I20-I25 (DOID:3393)
- ICD10:K76.1 (DOID:3393)
- ICD9:414.9 (DOID:3393)
- ICD10:I25.1 (DOID:3393)
- OMIM:608901 (http://www.ebi.ac.uk/ontology/webulous#OPPL_pattern)
- MSH:D003327

Experimental Factor Ontology (EFO)

https://epigraphdb.org



GWAS Catalog study, and its EFO annotation

	trait	efo_term
0	Crohn's disease	Crohn's disease
1	HIV-1 control	HIV-1 infection
2	Magnesium levels	magnesium measurement
3	HDL cholesterol	high density lipoprotein cholesterol measurement
4	LDL cholesterol	low density lipoprotein cholesterol measurement
5	Triglycerides	triglyceride measurement
6	Sudden cardiac arrest	sudden cardiac arrest
9	Orofacial clefts (maternal alcohol consumption...	alcohol consumption measurement, Orofacial cle...
10	HIV-1 susceptibility	HIV-1 infection
11	Age-related macular degeneration	age-related macular degeneration

Method

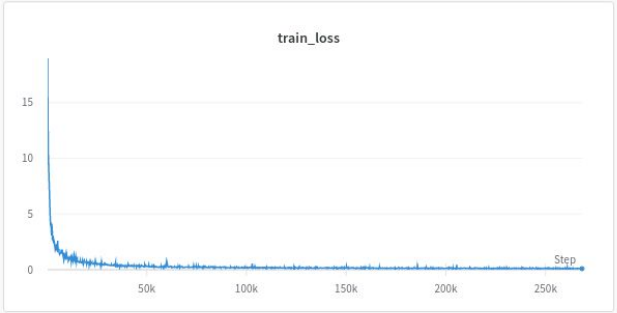
- Process EFO graph into a networkx graph
- Attach GWAS Catalog traits as new EFO nodes
- For every node:
  - pairwise distance as number of steps via shortest path
  - self-distance with its synonyms as step 0
  - (not done yet) pairwise synonym distance
- Uniform sample 30%
- Type 1: regression of distance scores
- Type 2: classification of distance scores
  - 0 - 4 asis
  - 5+ squashed to 5

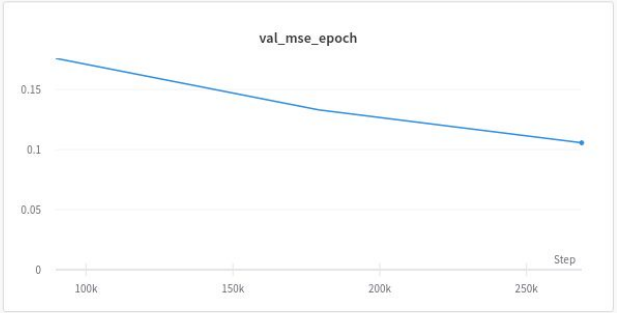
Training task

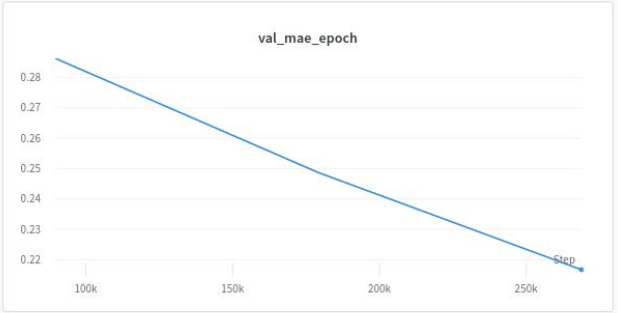
```
text: [CLS] dorsolateral prefrontal cortex functional measurement [SEP] arthroderma cajetani [SEP]
target: 9.0     pred: 9.029166221618652

text: [CLS] primary fanconi syndrome [SEP] diabetic retinopathy [SEP]
target: 6.0     pred: 5.206423282623291

text: [CLS] deafness - genital anomalies - metacarpal and metatarsal synostosis [SEP] organic [SEP]
target: 13.0    pred: 13.291205406188965

text: [CLS] cirrhosis [SEP] ring chromosome 21 [SEP]
target: 7.0     pred: 7.4355149269104

text: [CLS] chest pain [SEP] spontaneous mutation [SEP]
target: 11.0    pred: 11.058844566345215

text: [CLS] hcc0630 [SEP] nci - h209 [SEP]
target: 2.0     pred: 1.9326502084732056

text: [CLS] holocarboxylase synthetase deficiency [SEP] genetic disorder [SEP]
target: 3.0     pred: 3.068690061569214

text: [CLS] maleate [SEP] 1 - naphthylacetic acid [SEP]
target: 3.0     pred: 2.883960723876953

text: [CLS] anterior spinal artery syndrome [SEP] hand - foot - genital syndrome [SEP]
target: 3.0     pred: 2.9395782947540283

text: [CLS] gm17157 [SEP] pc - 6 [SEP]
target: 2.0     pred: 1.8309314250946045
```

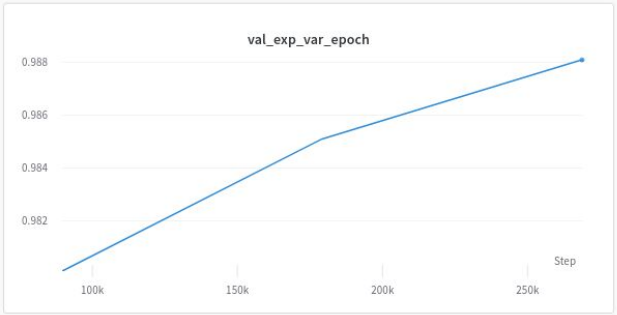
Results

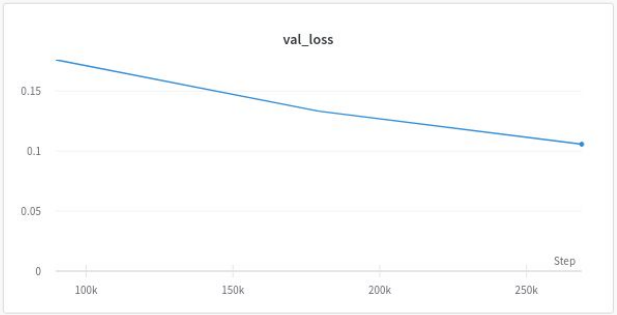
- Only need a few finetuning epochs, as expected
- Small sample suffice
- Mk1 (regression): explained variance 0.98, MSE 0.11, MAE 0.22
- Mk2 (classification): accuracy 0.99, precision 0.99, recall 0.99

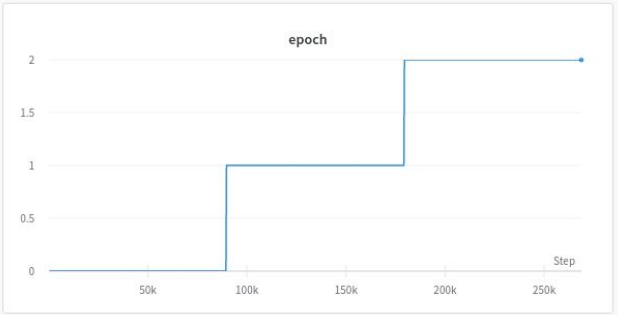












Results - the good

	trait	efo_term	pred	target	diff
0	Malignant mesothelioma	mesothelioma	1.000238	1.0	0.000238
1	Metabolite levels	metabolite measurement	0.999371	1.0	0.000629
2	Tyrosine levels	tyrosine measurement	1.000991	1.0	0.000991
3	Butyrylcholinesterase levels	butyrylcholinesterase measurement	0.998593	1.0	0.001407
4	Hypertension (SNP x SNP interaction)	hypertension	1.001603	1.0	0.001603
5	Optic disc area	optic disc area measurement	0.998011	1.0	0.001989
6	Esophageal cancer (alcohol interaction)	esophageal carcinoma	1.003343	1.0	0.003343
7	Obsessive-compulsive disorder or autism spectr...	obsessive-compulsive disorder, autism spectrum...	0.995708	1.0	0.004292
8	Vestibular neuritis	vestibular neuronitis	0.995653	1.0	0.004347
9	Nonalcoholic fatty liver disease	non-alcoholic fatty liver disease	1.004780	1.0	0.004780
10	Pit-and-Fissure caries	pit and fissure surface dental caries	0.995210	1.0	0.004790
11	White matter lesion progression	white matter lesion progression measurement	1.004839	1.0	0.004839
12	Prostate cancer (SNP x SNP interaction)	prostate carcinoma	0.993228	1.0	0.006772
13	Large artery stroke (TOAST classification)	large artery stroke	0.993190	1.0	0.006810
14	Pulse pressure (dietary potassium intake inter...	pulse pressure measurement, dietary potassium ...	0.992917	1.0	0.007083
15	Schizophrenia or cigarettes per day (pleiotropy)	schizophrenia, cigarettes per day measurement	1.008028	1.0	0.008028
16	Hypersomnia (HLA-DQB1*06:02 negative)	hypersomnia	0.991550	1.0	0.008450
17	Prostate cancer (early onset)	prostate carcinoma	0.991139	1.0	0.008861
18	Hemoglobin concentration	hemoglobin measurement	1.012157	1.0	0.012157
19	Niacinamide levels	niacinamide measurement	0.987777	1.0	0.012223

Results - the bad

	trait	efo_term	pred
10	HIV-1 susceptibility	HIV-1 infection	2.531984
12	Mean forced vital capacity from 2 exams	vital capacity	2.653558
26	Response to lithium treatment in bipolar disorder	response to lithium ion	2.525187
27	Adiposity	obesity	3.215511
32	Alcohol consumption (transferrin glycosylation)	alcohol drinking	3.374798
35	Cutaneous nevi	nevus	2.636189
49	Fat intake	energy intake	5.139570
57	Primary biliary cholangitis	biliary liver cirrhosis	4.314609
63	Homeostasis model assessment of insulin resist...	HOMA-IR	4.880254
64	Homeostasis model assessment of beta-cell func...	HOMA-B	4.789968
71	Response to citalopram treatment	response to antidepressant	2.578714
77	Tanning	suntan	4.601673
88	Chronic kidney disease	creatinine measurement, cystatin C measurement...	5.243328
99	Response to hepatitis C treatment	chronic hepatitis C virus infection	3.759204
101	Hoarding	obsessive-compulsive disorder	3.219840
107	Cholesterol, total	total cholesterol measurement	3.113816
112	Lp (a) levels	apolipoprotein A 1 measurement	2.776184
114	Amyotrophic lateral sclerosis (SNP x SNP inter...	sporadic amyotrophic lateral sclerosis	2.512428
119	Warfarin maintenance dose	response to anticoagulant	2.632805
140	Blue vs. green eyes	eye color	2.756887

Results - the really bad

```
trait: Orofacial clefts (maternal alcohol consumption interaction)
efo_term: alcohol consumption measurement, Orofacial clefting syndrome, cleft lip
distance: 11.0

trait: Body mass in chronic obstructive pulmonary disease
efo_term: chronic obstructive pulmonary disease, body mass index
distance: 11.0

trait: Breast cancer (survival)
efo_term: event free survival time, survival time, breast carcinoma
distance: 13.0

trait: End-stage coagulation
efo_term: factor VIII measurement, von Willebrand factor measurement, coagulation factor measurement
distance: 11.0

trait: Colorectal cancer (environment interaction)
efo_term: smoking status measurement, colorectal cancer, alcohol consumption measurement, overweight body mass ind
ex status
distance: 13.0

trait: Orofacial clefts
efo_term: Orofacial clefting syndrome
distance: 12.0

trait: Type 2 diabetes (dietary heme iron intake interaction)
efo_term: type II diabetes mellitus, dietary heme iron intake measurement
distance: 12.0

trait: Lung cancer (asbestos exposure interaction)
efo_term: lung carcinoma, asbestos exposure measurement
distance: 11.0

trait: Oral cancers (chewing tobacco related)
efo_term: matrix metalloproteinase measurement, mouth neoplasm
distance: 11.0

trait: Age at smoking initiation in chronic obstructive pulmonary disease
efo_term: chronic obstructive pulmonary disease, smoking initiation
distance: 11.0

trait: Type 2 diabetes and 6 quantitative traits
efo_term: insulin measurement, glucose measurement, type II diabetes mellitus
distance: 11.0

trait: Waist Circumference - Triglycerides (WC-TG)
efo_term: triglyceride measurement, metabolic syndrome
distance: 13.0
```

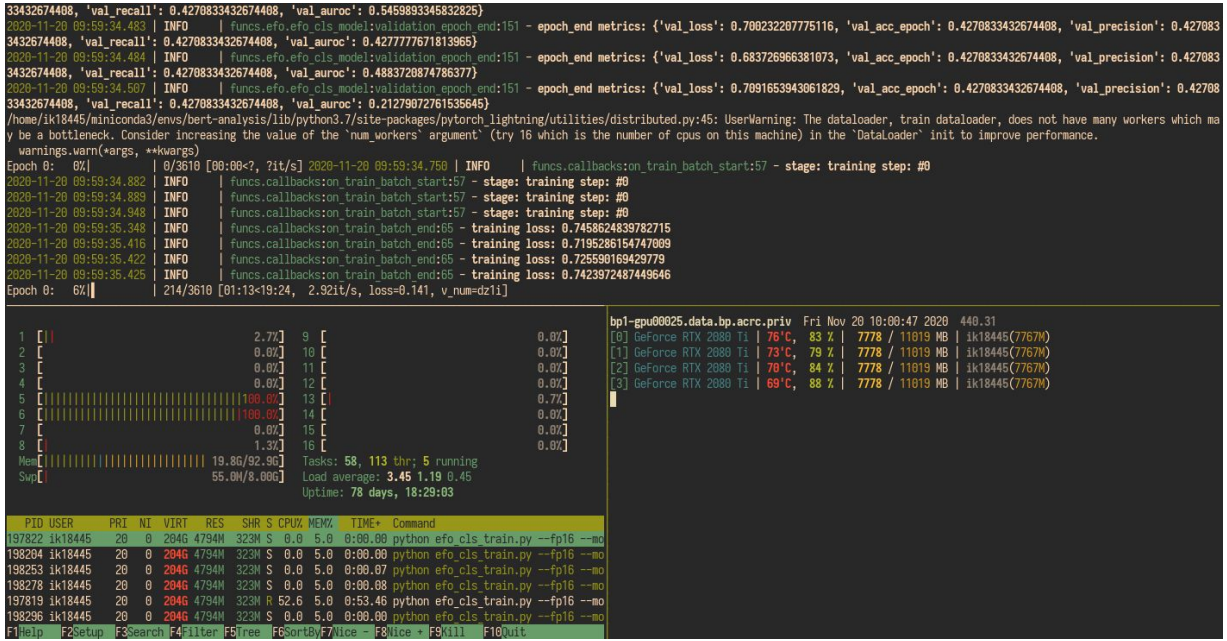
1.0	1973
2.0	1693
3.0	1132
-0.0	400
4.0	341
5.0	245
10.0	154
11.0	132
9.0	129
6.0	95
7.0	92
8.0	90
12.0	73
13.0	49
-1.0	17
14.0	11
16.0	1
15.0	1

Architecture of the Vectology platform

- Processing
  - python
- Model training
  - pytorch lightning, huggingface transformers, wandb
- Web services
  - Orchestration: docker-compose
  - Frontend (http://vectology.mrcieu.ac.uk): Vue.js, Vuetify
  - Backend API (http://vectology-api.mrcieu.ac.uk): FastAPI
  - Recommender: Elasticsearch
  - biosentvec serving: Django
  - pretrained bert serving: bert-as-service https://github.com/hanxiao/bert-as-service
  - inhouse model serving: torchserve

Model training

UoB bluepebble cluster



- 4 finetuning epochs
- Paralleisation
  - 1 node, 4 GPUs
  - distributed data parallel
- half-precision (16bit) computation
- One cycle learning rate scheduling
- Data loader
  - RAM 90G
  - Custom SQL loader
  - Tokenization on-the-fly

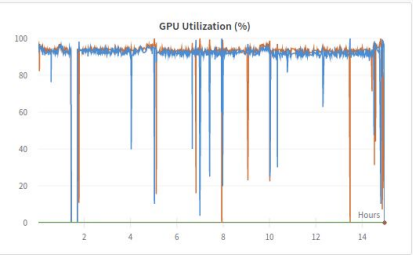
huggingface transformers and pytorch lightning

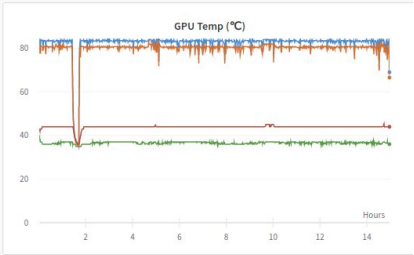
- huggingface transformers

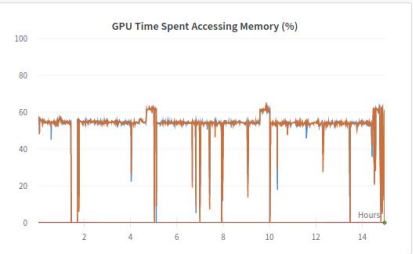
- pytorch lightning
  - keras style building blocks (trainer, data module, model)
  - much better hackability (monitoring, scheduling, etc)
  - builtin support for
    - mixed floating point, parallelised training, etc
    - lr finder, early stopping, etc
    - 3rd party plugins: wandb, mlflow

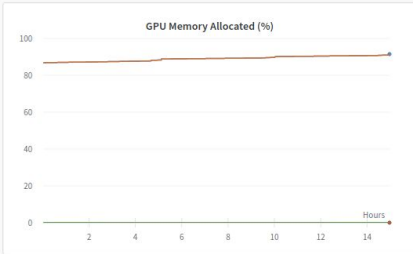
wandb (weights and biases https://wandb.ai)

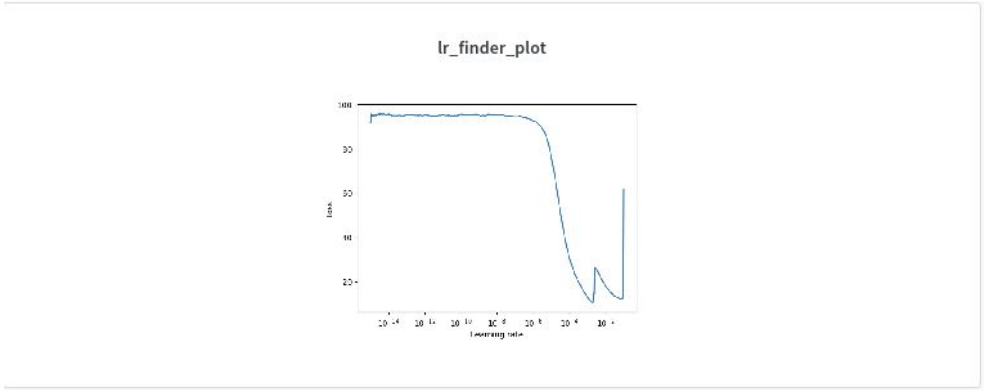
- metrics and artifacts logging
- system monitoring











lr_data

Search

Page 1 of 19

lr	lr_scientific	loss
1e-15	1.0000e-15	91.80097961425781
1.0710412066840061e-15	1.0710e-15	93.59448843291301
1.1084325481163935e-15	1.1084e-15	95.81126867114496
1.1471292664151318e-15	1.1471e-15	95.89981470858879

simple parsing

- hyperparameter definition and parsing in one place!

```
> python bert_ner_conll2003.py --help
usage: bert_ner_conll2003.py [-h] [--learning_rate float] [--weight_decay float] [--adam_epsilon float] [--init_learning_
rate [float]]
                             [--use_lr_scheduler] [--num_steps_per_epoch [int]] [--model.model_name str] [--data_module.m
odel_name str]
                             [--max_tokenization_length int] [--train_batch_size int] [--eval_batch_size int] [-j int] [-
n] [--num_train_epochs int]
                             [--seed int] [--gpus int] [--overwrite] [--fp16] [--accelerator [str]]

optional arguments:
  -h, --help            show this help message and exit

TransformerModelHparams ['model']:
  TransformerModelHparams(learning_rate: float = 0.0003, weight_decay: float = 0.01, adam_epsilon: float = 1e-08, init_le
arning_rate: Union[float, NoneType] = None, use_lr_scheduler: bool = False, num_steps_per_epoch: Union[int, NoneType] = N
one, model_name: str = 'bert-base-uncased')

  --learning_rate float, --model.learning_rate float
  --weight_decay float, --model.weight_decay float
  --adam_epsilon float, --model.adam_epsilon float
  --init_learning_rate [float], --model.init_learning_rate [float]
  --use_lr_scheduler, --model.use_lr_scheduler
  --num_steps_per_epoch [int], --model.num_steps_per_epoch [int]
  --model.model_name str

DataModuleHparams ['data_module']:
  DataModuleHparams(model_name: str = 'bert-base-uncased', max_tokenization_length: int = 128, train_batch_size: int = 64
, eval_batch_size: int = 64, num_workers: int = 2)

  --data_module.model_name str
  --max_tokenization_length int, --data_module.max_tokenization_length int
  --train_batch_size int, --data_module.train_batch_size int
  --eval_batch_size int, --data_module.eval_batch_size int
  -j int, --num_workers int, --data_module.num_workers int

TrainerHparams ['trainer']:
  TrainerHparams(dry_run: bool = False, num_train_epochs: int = 1, seed: int = 42, gpus: int = 1, overwrite: bool = False
, fp16: bool = False, accelerator: Union[str, NoneType] = None)

  -n, --dry_run, --trainer.dry_run
  --num_train_epochs int, --trainer.num_train_epochs int
```

```python
def create_parser() -> ArgumentParser:
    parser = ArgumentParser()
    parser.add_arguments(ModelHparams, dest="model")
    parser.add_arguments(DataModuleHparams, dest="data_module")
    parser.add_arguments(TrainerHparams, dest="trainer")
    return parser


def process_args(args: argparse.Namespace, data_module):
    args.model.num_steps_per_epoch = get_num_steps_per_epoch(
        data_module=data_module, args=args.trainer
    )
```

```
 <N>  ~/1/+/ml-things/lightning/src/bert_ner_conll2003.py  19:0 23%       LF UTF
```

```python
@dataclass
class ModelHparams:
    learning_rate: float = settings.LEARNING_RATE
    weight_decay: float = settings.WEIGHT_DECAY
    adam_epsilon: float = settings.ADAM_EPSILON
    init_learning_rate: Optional[float] = None
    use_lr_scheduler: bool = flag(False)
    num_steps_per_epoch: Optional[int] = None


@dataclass
class TransformerModelHparams(ModelHparams):
    model_name: str = settings.TRANSFORMER_MODEL_NAME_BASE


@dataclass
class TransformerDataModuleHparams:
    model_name: str = settings.TRANSFORMER_MODEL_NAME_BASE
    max_tokenization_length: int = settings.MAX_TOKENIZATION_LENGTH
    train_batch_size: int = settings.BATCH_SIZE
    eval_batch_size: int = settings.BATCH_SIZE
    num_workers: int = field(settings.NUM_WORKERS, alias=["-j", "num_workers"])
```

torchserve

```
model-store
    bluebert_efo
        config.json
        pytorch_model.bin
    bluebert_efo.mar
    densenet161-8d451a50.pth
    densenet161.mar
    .gitignore
    .gitkeep
scripts
    bluebert_efo
        serve_example.py
        serve.py
    densenet
    .gitkeep
Dockerfile
README.md
requirements.txt
<N'>  % torch-serve  1:2 All

torch-serve:
  build:
    context: ./torch-serve
    dockerfile: Dockerfile
  restart: unless-stopped
  ports:
    # inference API
    - ${TORCH_SERVE_PORT_0:-7580}:8080
    # management API
    - ${TORCH_SERVE_PORT_1:-7581}:8081
    # metrics API
    - ${TORCH_SERVE_PORT_2:-7582}:8082
  volumes:
    - ./torch-serve/model-store:/home/model-server/model-store
    - ./torch-serve/scripts:/home/model-server/scripts


<N>  /d/vectology-dev/docker-compose-dev.yml  43:13 Bot
```

```python
class BertEfoSeqClassifierHandler(BaseHandler, ABC):
    def __init__(self):
        super(BertEfoSeqClassifierHandler, self).__init__()
        self._batch_size = 0
        self.initialized = False

    def initialize(self, ctx):
        """
        Init and load model.
        """
        self.manifest = ctx.manifest
        self.metrics = ctx.metrics
        logger.info(f"Manifest: {self.manifest}")
        properties = ctx.system_properties
        logger.info(f"properties: {properties}")
        self._batch_size = properties["batch_size"]
        self.device = torch.device(
            "cuda:" + str(properties.get("gpu_id"))
            if torch.cuda.is_available()
            else "cpu"
        )

        model_dir = Path(properties.get("model_dir"))
        # num_labels =1 -- regression
        config = AutoConfig.from_pretrained(
            str(model_dir / "config.json"), num_labels=1
        )
        self.model = AutoModelForSequenceClassification.from_pretrained(
            str(model_dir), config=config
        )
        self.model.eval()
        self.tokenizer = AutoTokenizer.from_pretrained(BASE_MODEL_NAME)
        logger.info(
            "Transformer model from path {0} loaded
successfully".format(model_dir)
        )

        self.initialized = True

    def preprocess(self, request_data: List[Any]):
        """Tokenize inputs"""
        logger.info(f"Received data: {request_data}; {type(request_data)}")
        data = RequestData(**request_data[0])
        encodings = self.tokenizer(
            data.body.text_1,
            data.body.text_2,
            truncation=True,
            padding="max_length",
            max_length=MAX_LENGTH,
            return_tensors="pt",
        )
        return encodings

    def inference(self, input):
        with torch.no_grad():
            output = self.model(**input)["logits"].reshape(-1).tolist()
        return list(output)

    def postprocess(self, inference_output) -> List[Dict]:
        logger.info(f"inference_output: {inference_output}")
        output = [{"output": inference_output}]
        return output
```

```python
    def preprocess(self, request_data: List[Any]):
        """Tokenize inputs"""
        logger.info(f"Received data: {request_data}; {type(request_data)}")
        data = RequestData(**request_data[0])
        encodings = self.tokenizer(
            data.body.text_1,
            data.body.text_2,
            truncation=True,
            padding="max_length",
            max_length=MAX_LENGTH,
            return_tensors="pt",
        )
        return encodings

    def inference(self, input):
        with torch.no_grad():
            output = self.model(**input)["logits"].reshape(-1).tolist()
        return list(output)

    def postprocess(self, inference_output) -> List[Dict]:
        logger.info(f"inference_output: {inference_output}")
        output = [{"output": inference_output}]
        return output


_service = BertEfoSeqClassifierHandler()


def handle(data, context):
    try:

        if not _service.initialized:
            _service.initialize(context)

        if data is None:
            return None

        data = _service.preprocess(data)
        data = _service.inference(data)
        data = _service.postprocess(data)

        return data
    except Exception as e:
        raise e
```

torchserve

```
|1 test.py|
import requests

url = "http://localhost:7280/predictions/bluebert_efo"
text_1 = [
    "Metabolite levels",
    "tyrosine levels",
    "Hypertension (SNP x SNP interaction)",
]
text_2 = [
    "metabolite measurement",
    "tyrosine measurement",
    "hypertension",
]

r = requests.post(url, json={"text_1": text_1, "text_2": text_2})
r.raise_for_status()
print(r.json())
```

```
In [1]: import requests

In [2]: url = "http://localhost:7280/predictions/bluebert_efo"

In [3]: text_1 = [
   ...:     "Metabolite levels",
   ...:     "tyrosine levels",
   ...:     "Hypertension (SNP x SNP interaction)",
   ...: ]

In [4]: text_2 = [
   ...:     "metabolite measurement",
   ...:     "tyrosine measurement",
   ...:     "hypertension",
   ...: ]

In [5]: r = requests.post(url, json={"text_1": text_1, "text_2": text_2})

In [6]: r.raise_for_status()

In [7]: print(r.json())
{'output': [0.9993720054626465, 1.0009912252426147, 1.0016032457351685]}

In [8]:
```

The Vectology platform, and phenotype mapping via ontology

2021-01-29 IEU data science talk

Thank you for listening

Yi Liu