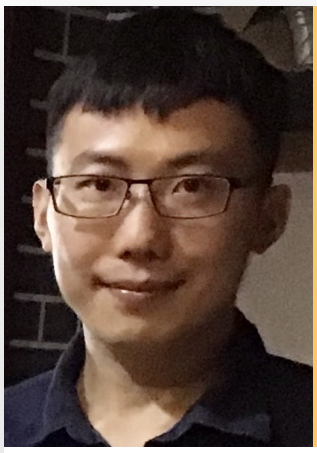

JINGSHU LIU

Data scientist

jingshu.liu@ls2n.fr

France

https://www.linkedin.com/in/jingshu-liu-6ba9b679

https://github.com/jingshu-liu

https://bitbucket.org/stevall/data-crawler

869436889

CAREER SUMMARY

- Currently working at Easiware-Dictanova as a data scientist while preparing my Ph.D. in natural language processing advised by Emmanuel Morin.
- Doing research on NLP and machine learning, focusing on cross-lingual applications and sequence modeling with transfer learning using pre-trained language models.
- With hands-on capability on machine learning and deep learning, broadly interested in applied machine learning for industrial scenarios and distributed system learning.

WORK EXPERIENCE

Data scientist/Machine learning — Easiware

2017 - Present

- Built from scratch a bilingual neural network based word and phrase embedding mapping framework in Java with Deeplearning4j-0.91.
- Implemented topic model pipelines using clustering on pre-trained unified phrase embeddings.
- Provisioned sparse matrix support and other mathematical optimizations in Nd4j-0.91.
- Designed and built an encoder-decoder framework for sequence modeling with Pytorch-1.2. Fully campatible in CPU and GPU mode which runned in OVH cloud server using the manage tools openstack and nvidia gpu cloud.
- Incoporated pre-trained Transformer based language models into our neural networks for real life scenarios.

Achievements:

- Improved the bilingual multi-word and single-word lexicon induction by an average of 22 points in MAP on client data.
- The new topic modeling system replaced the existing rule-based topic modeling system.

Environments:

Java · Python · Pytorch · Deeplearning4j · Keras · Scikit Learn · OpenStack-Docker

Natural Language Processing Intern — Dictanova

2016

- Implemented a term extraction and Aspect Based Sentiment Analysis pipeline for simplified and traditional Chinese language in Java with UIMA architecture and ElasticSearch storage.
- Improved Chinese language preprocessing (POS-tagging) for FNLP. Meanwhile, added an innovative Chinese lemmatizer for reduplicated words.
- Data cleaning and visulization using Pandas and R.

Achievements:

- Achieved state-of-the-art results on Aspect Based Sentiment Analysis on Semeval2016.
- Improved the term extraction accuracy by 50%.

Environments:

Java · UIMA · Deeplearning4j · Python · R · ElasticSearch

Natural Language Processing Intern — LLF

SKILLS & TOOLS

Backend

Java8
Python3
C
Perl
PHP

Machine Learning Framework

Deeplearning4j
Pytorch
Keras
R
Tensorflow
Torch

Others

HTML · CSS · MySQL · LaTeX · HDFS Hadoop · Git · Unit Test · Agile · Gradle · Lua · Neo4j · ElasticSearch

OTHER PROJECTS

Poem bot
Built and trained in Java a peom bot which can generate the next second part of a couplet given the first one.

Hackthon CafData 2015
Built from scratch in 48h a waiting time prediction

2015

- Collaborated with researchers in Duel Project on humain dialogue classification.
- Annotated sentiment analysis corpora with Brat.

Environments:

Java-Corenlp Perl Python Numpy Brat

EDUCATION

PHD candidate in NLP — University of Nantes

2017 - Present (Expected to graduate in January, 2020)

Thesis title: Unsupervised cross-lingual representation modeling for variable length phrases.

- Unsupervised bilingual phrase alignment.
- Monolingual sequence modeling with RNN, CNN, LSTM and modern Transformer based architecture.
- Bilingual word embedding.
- Data augmentation/selection for low-resource scenario.

Results:

- Improved state-of-the-art results on phrase synonymy by almost 33% on low-resourced specialized domain corpora.
- Achieved state-of-the-art results on bilingual word mapping.
- Proposed a new tree-free graph based neural network for encoding short sequences including single-words. It outperformed state-of-the-art results on unsupervised bilingual phrase mapping by an average of 8.8 points in MAP while holding a comparable results for the single-word subset.

Master in NLP — University of Paris Diderot

2014-2016

Notable courses:

Machine learning; Statistics; Algorithm; 1st order logic; Text mining.

BS in Applied Mathmatics — University of Paris Diderot

2012-2014

Notable courses:

Linear algebra; Mathmatical analysis; Java programming; C programming; Probability theory; HTML/CSS/PHP; MySQL

PUBLICATIONS

- Alignement de termes de longueur variable en corpus comparables spécialisés
  Jingshu Liu, Emmanuel Morin, Sebastián Saldarriaga
  TALN2018
- Towards a unified framework for bilingual terminology extraction of single-word and multi-word terms
  Jingshu Liu, Emmanuel Morin, Sebastián Saldarriaga
  Coling2018
- Continuous phrase representation learning with wrapped context prediction
  In preparation
- A unified and unsupervised framework for bilingual phrase alignment on specialized comparable corpora
  Jingshu Liu, Emmanuel Morin, Sebastián Saldarriaga, Joseph Lark
  Arxiv
- From unified phrase representation to bilingual phrase alignment in an unsupervised manner
  In preparation

system in Python based on the data given by la Caf in a hackthon competition.

Gounki game

Implemented a Gounki game in C.

Sheep and wolf evolution game

Implemented an evolution game in Java with a minimal UI.

Recipe website

Built a recipe website with Mysql and PHP which was hosted in the campus network of Paris Dederot University. A student can register to find others who can teach him the recipes he wants to learn.

VOLUNTEER EXPERIENCE

Custom layer for Deeplearning4j

Implemented a custom layer for Deeplearning4j (before alpha version) and the pull request was merged into the main project.

Machine Learning Meetup

Talk on nantes machine learning meetup 2019.

Liaison manager

Responsible for the communication between the team of Groupe Edmond de Rothschild and the host city for Extream Sailing Series 2011 in Qingdao.

Interpreter

Interprater for Tianhui (SARL) at China Import and Export Fair in Guangzhou, 2010.

LANGUAGE

Chinese (Native)

English (Professional)

French (Professional)

INTERESTS

Badminton, Basketball, Running

Language, History

Board & Video game