Dung Ngoc Thai

http://www.cs.umass.edu/~dthai
dthai@cs.umass.edu | (669)-233-3660

EDUCATION

UMASS, AMHERST
PH.D. COMPUTER SCIENCE
Sep 2016 - present

HCMUT
MS COMPUTER SCIENCE
Nov 2014 | HCM, VN
Cum. GPA: 3.9

SOCIALS

github.com/dungtn
linkedin.com/in/dung-thai
semanticscholar.org/Dung-Thai

COURSES

Reinforcement Learning
by Philip Thomas

Machine Learning
by Srihdar Mahadevan

Deep Learning
by Erik Learned-Miller

SKILLS

Programming Languages
Python • Java • C++
Matlab • Julia

Frameworks
Theano • Tensorflow • Keras
CUDA • NLTK

REFERENCES

Prof. Andrew McCallum , UMass
mccallum@cs.umass.edu
Prof. Nam Thoai , HCMUT
nam@cse.hcmut.edu.vn
Prof. Vu Dinh , HCMUT
vudda@uit.edu.vn

SUMMARY

I'm June, a PhD student at UMass Amherst, advised by **Professor Andrew McCallum** . I'm interested in various areas of Machine Learning (ML) and Natural Language Processing (NLP). My overarching research focuses on Case-Based Reasoning and its applications in open-domain and knowledge-base question answering.

RESEARCH EXPERIENCES

INFORMATION EXTRACTION SYNTHESIS LAB | GRADUATE STUDENT
Sep 2016 – present | UMass Amherst, US
Case-Based Reasoning (CBR) is an experience-based approach to solve new problems by adapting previously known solutions to similar problems. CBR is particularly useful for complex, compositional problems, such as question answering. I'm working on non-parametric, pre-train based CBR QA systems. My research also involves metric learning and learning of dense-retrieval models.

IBM | RESEARCH INTERN

Online 2021	**Question answering over incomplete knowledge bases** built a CBR system that utilizes KB Completion and pre-trained text representation for incomplete-KBQA.
Yorktown 2020	**Compositional question answering over knowledge bases** built a sketch-generator model jointly with end-to-end entity linking and relation extraction.

ADOBE INC. | RESEARCH INTERN

San Jose 2018	**Question answering on semi-structured tables** built a multi-heads attention model based on the Neural Programmer architecture.
San Jose 2017	**Variational autoencoder for semi-supervised question answering** learned an unsupervised question representation to improve the generalization of the supervised question answering model.

SELECTED PUBLICATIONS

EMNLP 2021	**Case-based Reasoning for Natural Language Queries over KBs** Rajarshi Das*, Manzil Zaheer*, **Dung Thai*** , Ameya Godbole*, Ethan Perez, Jay-Yoon Lee, Lizhen Tan, Lazaros Polymenakos, Andrew McCallum
NAACL 2021	**TABBIE: Pretrained Representations of Tabular Data** Hiroshi Iida*, **Dung Thai*** , Varun Manjunatha, Mohit Iyyer
REPL4NLP 2021	**Simultaneously Self-Attending to Text and Entities for Knowledge-Informed Text Representations** **Dung Thai** , Raghuveer Thirukovalluru, Trapit Bansal, Andrew McCallum
REPL4NLP 2021	**Knowledge Informed Semantic Parsing for Conversational QA** Raghuveer Thirukovalluru, **Dung Thai** , Mukund Sridhar, Shruti Chanumolu, Sankaranarayanan Ananthakrishnan, Andrew McCallum
CoNLL 2018	**Embedded-State Latent Conditional Random Fields for Sequence Labeling (Oral Presentation)** **Dung Thai** , Sree H. Ramesh, Shikhar Murty, Luke Vilnis, Andrew McCallum in SIGNLL Conference on Computational Natural Language Learning, Belgium
ICML 2017	**Low-rank hidden state embeddings for Viterbi sequence labeling** **Dung Thai** , Shikhar Murty, Trapit Bansal, Luke Vilnis, David Belanger, Andrew McCallum in 1st DeepStruct Workshop, in 34^{th} International Conference on Machine Learning , Australia.