Lucas Sterckx

lucassterckx@gmail.be | +32 (0)9 33 14979
http://lusterck.github.io | @lusterck
Technologiepark Zwijnaarde 15, 9052 Ghent, Belgium

RESEARCH STATEMENT

I am a doctoral student at Ghent University finalizing my PhD in natural language processing. In the past I have worked on multiple information extraction problems in collaboration with major Belgian media companies and startups. My more recent research focuses on sequence-to-sequence modeling and neural language models for creative writing and song lyrics.
I am passionate about education, communication, data science for social good and mental health and the intersection of machine learning and music. While at Ghent University I guided practical sessions in basic and advanced Python programming, engineering projects and master theses. I strive to keep students motivated and interested while helping them to develop their ideas.
Currently, I am looking for new opportunities in industry or academia to collaborate with experienced researchers and work on challenging natural language processing problems.

RESEARCH

GHENT UNIVERSITY - IMEC | GHENT, BE
IDLab, Department of Information Technology
2013 – 2017 | PhD Student

- Topic Models [1, 2]
  During my M.Sc. thesis, I studied topic modeling in large corpora of song lyrics. My research was on better, automatic evaluation of topic models and measuring their interpretability. For this, I invented a new evaluation measure which showed higher correlation with human evaluation than standard methods.
- Knowledge Base Population [3, 4, 5, 6]
  I was the developer of the Ghent University's knowledge base population system for two participations in the 2014 and 2015 English Slot Filling tracks at the Text Analysis Conference (TAC) benchmarks organized by NIST. In 2015, this system placed 2^{nd} out of 12 systems. My technique for relation extraction used low dimensional embeddings of relation patterns to filter distantly supervised training data and train better supervised classifiers.
- Keyphrase Extraction [7, 8, 9, 10]
  I was researcher for the keyphrase extraction track of STEAMER, a government funded project in collaboration with some of Flander's biggest media providers. In this context, I was a visiting researcher at the University of North Texas to work with prof. Cornelia Caragea and her group, to research methods to counter noise in training data for keyphrase extraction. The developed techniques were made commercially available via a spin-of company of the research group.
- Automated Annotation Systems [11]
  Recently, I have been working on models to automatically annotate text, difficult to understand or interpret by readers, with explanations or simplification. For this, I use techniques from the field of neural machine translation. I developed this idea while being a visiting researcher at the Machine Intelligence Lab at Cambridge University under the supervision of prof. Bill Byrne

EDUCATION

GHENT UNIVERSITY
2011 - 2013 | Ir. MSc in Computer Science Engineering
Ghent | Graduated with honors

VRIJE UNIVERSITEIT BRUSSEL
2007 - 2011 | Ing. MSc in Electronics and Information Technology Engineering
Brussels | Graduated with honors

TEACHING

- Teaching assistant for the introductory course on programming in Python for first-year engineering students
- Supervising student projects for data mining and engineering courses and giving introductory lectures on machine learning and natural language processing.
- Supervising master theses on machine learning and natural language processing.
- Private tutoring of undergrad students for math and statistics courses.

REVIEWING

EACL 2017
EMNLP 2017
Workshop on Learning to Generate Natural Language at ICML 2017
StarAI Workshop at UAI 2017

AWARDS

2011 - Monetary Award for Outstanding Thesis
2015 - FWO Travel Scholarship for Short Stay Abroad
2017 - FWO Travel Scholarship for Long Stay Abroad

LANGUAGES

PROGRAMMING

Experience with:
Python • Shell • Java • C++ • JavaScript • Matlab • R

Machine Learning Frameworks:
Pytorch• Tensorflow • Keras • Torch • Lasagne • MXNet • Scikit-Learn

SPOKEN & WRITTEN

Fluent:
English, Dutch
Intermediate:
French
Elementary:
German, Spanish

PUBLICATIONS

[1] L. Sterckx, "Topic detection in a million songs," 2013. Diss. MSc in Comuter Science Engineering.

[2] L. Sterckx, T. Demeester, J. Deleu, L. Mertens, and C. Develder, "Assessing quality of unsupervised topics in song lyrics," in ECIR, vol. 8416 of Lecture Notes in Computer Science, pp. 547–552, Springer, 2014.

[3] M. Feys, L. Sterckx, L. Mertens, J. Deleu, T. Demeester, and C. Develder, "Ghent University-IBCN participation in TAC-KBP 2014 slot filling and cold start tasks," in 7th Text Analysis Conference (TAC 2014), pp. 1–10, 2014.

[4] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Using active learning and semantic clustering for noise reduction in distant supervision," in 4th Workshop on Automated Base Construction at NIPS 2014 (AKBC-2014), pp. 1–6, 2014.

[5] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Ghent University-IBCN participation in the TAC KBP 2015 cold start slot filling task," in 8th Text Analysis Conference, 2015.

[6] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Knowledge base population using semantic label propagation," Knowledge-Based Systems, vol. 108, pp. 79 – 91, 2016. New Avenues in Knowledge Bases for Natural Language Processing.

[7] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Topical word importance for fast keyphrase extraction," in WWW (Companion Volume), pp. 121–122, ACM, 2015.

[8] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "When topic models disagree: Keyphrase extraction with multiple topic models," in WWW (Companion Volume), pp. 123–124, ACM, 2015.

[9] L. Sterckx, C. Caragea, T. Demeester, and C. Develder, "Supervised keyphrase extraction as positive unlabeled learning," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 1924–1929, Association for Computational Linguistics, 2016.

[10] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Creation and evaluation of large keyphrase extraction collections with multiple opinions," Language Resources and Evaluation, pp. 1–30, 2017.

[11] L. Sterckx, J. Naradowsky, B. Byrne, T. Demeester, and C. Develder, "Break it down for me: A study in automated lyric annotation," in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. –, Association for Computational Linguistics, 2017.

REFERENCES

I was supervised by and worked closely together with the people listed below.

PROF. CHRIS DEVELDER
IDLab, Dept. of Information Technology
Ghent University – imec
Technologiepark Zwijnaarde 15, 9052 Gent, Belgium
chris.develder@ugent.be

DR. THOMAS DEMEESTER
IDLab, Dept. of Information Technology
Ghent University – imec
Technologiepark Zwijnaarde 15, 9052 Gent, Belgium
thomas.demeester@ugent.be

PROF. CORNELIA CARAGEA
Computer Science
Kansas State University
Manhattan, KS 66502, USA
ude.etats-k@aegaracc

PROF. BILL BYRNE
Information Engineering Division
Cambridge University
Trumpington Street, Cambridge, CB2 1PZ, UK
bill.byrne@eng.cam.ac.uk