

SAATVIGA SUDHAHAR

Email: saatviga.sudhahar@healx.io

Mobile: +447901093283

PROFILE SUMMARY

A well presented, self motivated and confident researcher who has extensive working experience as a research scientist in the areas of Natural Language Processing, Data Science and Machine Learning currently focusing on Knowledge graph Inference and Reasoning using Deep Learning. Having excellent research potential with an ability to actively lead research projects and contribute to the project's goals as well as with a proven publication track record.

EDUCATION

University of Bristol, UK

2010 - 2014

PhD Engineering Mathematics

Thesis title: Automated analysis of narrative text using network analysis in large corpora

Supervisor: Prof. Nello Cristianini (UoB)

Examiners: Dr. Tim Kovacs (UoB), Prof. Jonathan Lawry (UoB), Dr. Kalina Bontcheva (University of Sheffield)

Date of graduation : July 2015

Specialities: Natural Language Processing, Text Analysis, Data Science, Machine Learning, Computational Social Science

University of Colombo School of Computing, Colombo, Sri Lanka

2005 - 2009

B.Sc. in Information & Communication Technology (Special)

Thesis title: A Patient-Centric eHealth Initiative through Telemedicine and mCommunication

Supervisors: Dr. S.M.K.D. Arunatilae (UCSC), Mr. G. P. Seneviratne (UCSC)

Class: First class honors, GPA: 3.68 (out of 4)

EXPERIENCE

Healx Ltd, Cambridge

Feb 2019 - Present

Machine Learning Scientist

Cambridge, UK

- Evaluation and analysis of state-of-the-art knowledge graph completion (KBC) methods based on deep graph embeddings, re-inforcement learning and rule-based learning on a biomedical graph to improve accuracy of predictions
- Building novel graph reasoning methods based on multi-hop path extraction and ranking using learnt graph embeddings and transformer architectures.
- Working with end-users for active feedback on methods developed and with engineers and product team to deploy them.

Department of Computer Science, University of Bristol

Feb 2016 - January 2019

Research Associate in Machine Learning

Bristol, UK

- ThinkBig (ERC Advance Grant Funded)
 - Working on inferring implicit knowledge from text using probabilistic graphical models (Hinge-loss Markov Random Fields) for fact checking using Probabilistic Soft Logic (PSL).
 - Working on building Machine Learning models using LIWC (Linguistic Inquiry and Word Count) indicators to predict and understand user appeal of news readers for several outlets based on what they click and share.
 - Working on a series of projects involving natural language processing and text mining tasks, semantic parsing, sentiment analysis, entity extraction, triplet extraction, entity disambiguation and building knowledge graphs to study macroscopic patterns in news data.
 - Worked on building and optimizing statistical phrase based machine translation models to translate text in european languages. Implemented them as a web service to which multiple clients can connect. This was used to monitor global changes of public opinion towards specific topics/entities.

Department of Computer Science, University of Bristol*Research Assistant in Machine Learning*

Oct 2013 - Jan 2016

Bristol, UK

- CompLACS (EU FP7 Funded)
 - Implemented online classifiers on different relevance scoring functions: relevance to different news topics such as sports, politics etc and also appeal to readers of different news outlets such as BBC, NPR, Seattle Times etc.
 - Implemented a two-layer learning representation by combining modular adaptive modules into a single learning system where an intermediate representation of the data was learnt (hidden layer) by supervised online learning based on web streams. This was then used as an input to train another system, realising a combined architecture.
- ThinkBig (ERC Advance Grant Funded)
 - Worked on detecting macro-scopic and long-term cultural trends using text analytic tools and knowledge bases in 150 years of British news articles spanning the whole of the 19th and half of the 20th century.

University of Colombo School of Computing*Instructor*

Nov 2009 - Sep 2010

Colombo, Sri Lanka

- Conducted Lecture series and Tutorials in Software Engineering and Communication Technologies.
- Involved in Undergraduate Paper Marking for Bachelor of ICT and Bachelor of IT degree programs.
- Involved in the Post-implementation phase as a Developer in the eHealth project "Vidusuwa".

Epic Lanka (Pvt) Ltd*Associate Software Engineer*

Feb - Aug 2008

Colombo, Sri Lanka

- Worked as a developer in Java and VB.NET.
- Contributed for the development of the Work Force Management System (WFM) for Sri Lanka Telecom (SLT).
- Contributed for the development of a secure File Transfer System for Nations Trust Bank (NTB), Sri Lanka.

SKILLS

- Programming Languages: Java, Python
- Databases: SQLServer, MySQL, NoSQL (MongoDB), Postgres, Neo4j
- Distributed Computing: Hadoop MapReduce
- Operating Systems: Linux (Ubuntu, Fedora, CentOS), Windows
- ML Libraries: Scikit-learn, Tensorflow, PyTorch
- Data processing: numpy, numba, pandas, scipy
- NLP Frameworks: Huggingface, Spacy, GATE, OpenNLP, Stanford CoreNLP, NLTK etc
- ML Frameworks: Spell, Comet
- Graph processing: DGL, PytorchBigGraph, LibKGE, PyKEEN
- Statistics: Matlab
- Network Analysis & Visualisation: Cytoscape, Gephi, JUNG, InfoViz, Prefuse
- Statistical Machine Translation: Moses
- Language Modelling Frameworks - IRSTLM, SRILM
- DevOps: Jira, Gitlab, Jenkins, Sonarqube, Nexus repository manager
- Cloud: AWS, GCP

PROFESSIONAL QUALIFICATIONS

- Sun Certified Java Professional (SCJP), certified under the J2SE 1.4 platform
- Certified with Merit in UNIX,C,C++ offered by SSI Education, India
- Certified with Credit in Oracle 9i offered by the Mercantile Institute of Information Technology (Pvt) Ltd, Sri Lanka

PUBLICATIONS

Journal Papers

- Lansdall-Welfare T., Sudhahar S., Thompson J., Lewis J., FindMyPast Newspaper Team & Cristianini N. (2017) Content analysis of 150 years of British periodicals. PNAS: doi:10.1073/pnas.1606380114
- Jia S., Lansdall-Welfare T., Sudhahar S., Carter C. & Cristianini N. (2016). Women Are Seen More than Heard in Online Newspapers. PLoS ONE 11(2): e0148434. doi:10.1371/journal.pone.0148434
- Sudhahar, S., Veltri, G. A., & Cristianini, N. (2015). Automated analysis of the US presidential elections using Big Data and network analysis. *Big Data & Society*, 2(1), 2053951715572916
- Sudhahar, S., Fazio, G.D., Franzosi, R. & Cristianini, N. (2013). Network analysis of narrative content in large corpora. *Natural Language Engineering*, 21(1), pp. 81-112.
- Sudhahar, S., & Cristianini, N. (2013). Automated Analysis of Narrative Content for Digital Humanities. *International Journal of Advanced Computer Science*, 3(9).
- Vatsalan, D., Arunatilake, S., Chapman, K., Sudhahar, S., & Abeywardhana, C. (2012). eClinics Integration Techniques for Clinical Information Systems Moving in to a National Network. *Sri Lanka Journal of Bio-Medical Informatics*, 2(4), pp. 130-143.

Conference Papers

- Sudhahar, S., Pierleoni, A., & Roberts, I. (2019, November). Reasoning Over Paths via Knowledge Base Completion. In Proceedings of the *Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)* (pp. 164-171).
- Sudhahar S. & Cristianini N. (2018) Detecting Shifts in Public Opinion: a big data study of global news content. *Symposium on Intelligent Data Analysis (IDA 2018)*. (pp. 316-327). Springer, Cham.
- Bindris N., Sudhahar S. & Cristianini N. (2018) Fact Checking from Natural Text with Probabilistic Soft Logic. Accepted at *Symposium on Intelligent Data Analysis (IDA 2018)*.(pp. 52-61) Springer, Cham.
- Lansdall-Welfare, T., Sudhahar, S., Thompson, J. & Cristianini, N., 2017, October. The Actors of History: Narrative Network Analysis Reveals the Institutions of Power in British Society Between 1800-1950. In *International Symposium on Intelligent Data Analysis* (pp. 186-197). Springer, Cham.
- Dzogang, F., Lansdall-Welfare, T., Sudhahar, S., & Cristianini, N. (2015). Scalable Preference Learning from Data Streams. *24th International Conference on World Wide Web Companion* (pp. 885-890). International World Wide Web Conferences Steering Committee.
- Lansdall-Welfare, T., Sudhahar, S., Veltri, G. A., & Cristianini, N. (2014). On the coverage of science in the media: A big data study on the impact of the Fukushima disaster. *IEEE International Conference on Big Data*, pp. 60-66.
- Sudhahar, S., Lansdall-Welfare, T., Flaounas, I., & Cristianini, N. (2012). Quantitative Narrative Analysis of US Elections in International News Media. *Internet, Politics, Policy 2012 (IPP2012): Big Data, Big Challenges Conference*, 20-21 September 2012, Oxford, UK.
- Sudhahar, S., Lansdall-Welfare, T., Flaounas, I., & Cristianini, N. (2012). ElectionWatch: detecting patterns in news coverage of US elections. *13th Conference of the European Chapter of the Association for Computational Linguistics*. Avignon, France, April 2012, pp. 82-86. *Association for Computational Linguistics*.
- Sudhahar, S., Franzosi, R., & Cristianini, N. (2011). Automating Quantitative Narrative Analysis of News Data. *2nd Workshop on Applications of Pattern Analysis (WAPA)*, Castro Urdiales, Spain, September 2011, pp. 63-71.

- Sudhahar, S., Vatsalan, D., Wijethilake, D., Wickramasinghe, Y., Arunathilake, S., Chapman, K., & Seneviratna, G. (2010, February). Enhancing Rural Healthcare in Emerging Countries through an eHealth Solution. *Second International Conference on eHealth, Telemedicine, and Social Medicine, 2010. ETELEMED'10*. pp. 23-28. IEEE.
- Vatsalan, D., Arunatileka, S., Chapman, K., Senaviratne, G., Sudahar, S., Wijetileka, D., & Wickramasinghe, Y. (2010, February). Mobile technologies for enhancing eHealth solutions in developing countries. *Second International Conference on eHealth, Telemedicine, and Social Medicine, 2010. ETELEMED'10*. pp. 84-89. IEEE.
- Sudhahar, S., Vatsalan, D., Wickramasinghe, Y., Wijethilake, D., Arunathilake, S., Chapman, K., & Seneviratna, G. Vidusuwa: A Mobile Telemedicine Solution for Patients in Emerging Countries. *eAsia 2009 Conference*, Colombo, Sri Lanka.
- Vatsalan, D., Arunathilake, S., Chapman, K., & Seneviratna, G., Sudahar, S., Wickramasinghe, Y., Wijethilake, D. Enhancing eHealth using m-Communications in Developing Countries. *eIndia eHealth 2009*, Hyderabad, India.

Other Reports

- Sudhahar, S. (2015) Automated Analysis of Narrative Text using Network Analysis in Large Corpora. PhD Thesis.
- Flaounas, I., Sudhahar, S., Lansdall-Welfare, T., Hensiger, E., & Cristianini, N. (2012) Big Data Analysis of News and Social Media Content.

AWARDS

- Frontier Prize Award for the best paper at the Sixteenth International Symposium on Intelligent Data Analysis (IDA 2017), London.
- University of Bristol Postgraduate Research Scholarship Award for my PhD studies in the University of Bristol
- NBQSA 2010 Silver Award for Project ViduSuwa
- e-Swabhimani 2009: The National Best e-Content Award in Sri Lanka under eHealth and Environment Category for the project "ViduSuwa".
- Manthan Award SouthAsia 2009: award under eHealth category for the project "ViduSuwa"
- eAsia 2009 Speaker Award, Dec. 2009 for the presentation done under eHealth Track

REFERENCES

Available upon request