Note: *The research proposals do not have to be one of the titles provided in this List, but it is recommended that the research is supported by experts available within the faculty of the Nazarbayev University School of Engineering and School of Science and Technology. Candidates may wish to generate their own titles. We also encourage the PhD candidates to contact potential Faculty Members for additional information on the following PhD topics.*

Prof. Sain Saginbekov sain.saginbekov@nu.edu.kz

1. Location Privacy Aware Reliable Communication Framework for Multisink Wireless Sensor Networks (WSNs)

With the availability of cheap sensor nodes, now it is possible to use hundreds of nodes in a Wireless Sensor Network (WSN) application. WSN applications are being used in a wide range of applications, including environmental, industrial, military, and health-care. WSNs are networks of small, battery operated sensor nodes that communicate over the radio. As sensor nodes are limited in battery lifetime and computational power, protocols developed for these nodes should be energy efficient. Usually WSNs are deployed with a single sink. However, there are several reasons for deploying WSNs with more than one sink. For example, one can deploy more than one sinks to prolong the lifetime of the network or to make it more reliable in cases when a node, a link or a sink fails. In some WSN applications the location privacy of source nodes and the sinks is important and therefore protocols used in the application have to prevent the location privacy. This research will focus on developing efficient reliable communication protocols that are aware of location privacy.

[1] Sain Saginbekov, Arshad Jhumka, Chingiz Shakenov: Towards Energy-efficient Collision-free Data Aggregation Scheduling in Wireless Sensor Networks with Multiple Sinks. SENSORNETS 2016: 77-86
[2] Sain Saginbekov and Arshad Jhumka. Many-to-Many Data Aggregation Scheduling in Wireless Sensor Networks with Two Sinks. Computer Networks 123: 184-199, 2017

Prof. Antonio Cerone antonio.cerone@nu.edu.kz

1. Formal analysis of interactive system using a cognitive architecture

Although interactive systems may appear to work correctly and safely when analysed in isolation from the human environment in which they are supposed to work, it is through the interaction between the computer and human components that critical errors emerge. This research project aims to extend a cognitive framework for modelling human automatic and deliberate cognitive processes and use it in combination with system or interface models to detect potential errors and analyse overall system properties. Modelling is carried out using rewriting logic and analysis is performed using an automated model-checking tool. The formal analysis will cover not only functional properties but also several categories of non functional properties, ranging from safety and security to usability and learnability.

[1] A. Cerone. A Cognitive Framework Based on Rewriting Logic for the Analysis of Interactive Systems. Software Engineering and Formal Methods (SEFM 2016), Vol. 9763 of Lecture Notes in Computer Science, Springer, 2016, pages 287-303.
[2] A. Cerone, Y. Zhao. Stochastic Modelling and Analysis of Driver Behaviour. Formal Methods for Interactive Systems (FMIS 2013), Vol 69 of ECEASST, 2013. https://journal.ub.tu-berlin.de/eceasst/article/view/965.
[3] A. Cerone, S. Connelly, P. Lindsay. Formal analysis of human operator behavioural patterns in interactive surveillance systems. Software and System Modeling 7(3): 273-286 (2008).

2. A multiscale, formal approach to ecosystem modelling

Human activities have important consequences on ecosystems and biodiversity and may break the conditions that guarantee the equilibrium and health of the ecosystem. This research project aims to define a general approach to the multiscale modelling and simulation of the dynamics of species that interact with their abiotic and biotic environment, in order to determine the appropriate policies for intervention and control that can restore and preserve equilibrium and health in the ecosystem. The approach will consider different levels of granularity and will integrate several ways of describing the ecosystem, ranging from visual representations to formal/mathematical descriptions. Various case studies will be carried out in the areas of migration, species reintroduction, and population and disease control. The project is carried out in collaboration with the University of Pisa and the GEOMAR Helmholtz Centre for Ocean Research Kiel.

[1] S. Setiawan, A. Cerone and P. Milazzo. A Tool for the Modelling and Simulation of Ecological Systems Based on Grid Systems. Software Engineering and Formal Methods (SEFM 2015 Collocated Workshops: MoKMaSD), Vol. 9509 of Lecture Notes in Compd Simulation of Ecological Systems Based on Grid Systems. uter Science, Springer, 2015, pages 198-212.
[2] A. Cerone, M Scotti. Research Challenges in Modelling Ecosystems. Software Engineering and Formal Methods (SEFM 2014 Collocated Workshops: MoKMaSD), Vol. 8938 of Lecture Notes in Computer Science, Springer, 2014, pages 276-293.
[3] S. Setiawan, A. Cerone. Stochastic Modelling of Seasonal Migration Using Rewriting Systems with Spatiality. Software Engineering and Formal Methods (SEFM 2013 Collocated Workshops: MoKMaSD), Vol. 8368 of Lecture Notes in Computer Science, Springer, 2013, pages 313-328.

3. Process mining-based modelling

Process mining emerged in the field of business process management (BPM) as an innovative technique to exploit the large amount of data recorded by information systems. It supports the discovery of not only relations and structure in data but also control flow, which can be either visualised or analysed for conformance with an a priori model. This research project aims to investigate how effectively process mining could be used in the following contexts:

- the extraction of learning and skill acquisition models from open source software repositories;

- the analysis of online reviews and support forums to characterise the quality of a product;
- the decision-making process in emergency management.

Furthermore, process mining techniques produce a mere representation of the process behaviour rather than an actual model of the process. The project also aims at extending the scope of process mining to the discovery of an abstract, functionally structured model, which comprises a set of formal rules for generating the system behaviour (model mining). The potential PhD student will contribute to one of the application contexts above or to the extension of the process mining scope.

[1] Cerone A., Model Mining - Integrating Data Analytics, Modelling and Verification, Journal of Intelligent Information Systems, Springer, 2017, Online First, DOI: 10.1007/s10844-017-0474-3.

[2] Mukala P., Cerone A. and Turini F., An Empirical Verification of A-priori Learning Models on Mailing Archives in the context of Online Learning Activities of Participants in Free/Libre Open Source Software (FLOSS) Communities, Education and Information Technologies 22(6): 3207-3229 (2017), Springer, DOI: 10.1007/s10639-017-9573-6.

[3] van der Aalst W. M. P. Process Mining - Data Science in Action. Springer, 2016.

Adnan YAZICI (Computer Science, SST), adnan.yazici@nu.edu.kz

Title: An Energy-Efficient Fuzzy Clustering Algorithm using Deep Learning for Wireless Sensor Networks (WSNs)

Clustering is utilized for effective communication in Wireless Sensor Networks due to its efficiency in energy consumption. In these networks, clustering can be done either following a crisp approach or a fuzzy one. Fuzzy clustering methodologies are found to be superior to crisp clustering counterparts when the boundaries among clusters are uncertain. As a result of this, a significant number of studies have proposed fuzzy-based solutions to the clustering problem. Most rule-based fuzzy systems determine and tune the fuzzy rules and the shapes of the output membership functions by employing some field experts in trial and error processes; thus, a considerable amount of time is dedicated to obtain and tune these functions, and it is almost impossible or impractical to contrive a fuzzy system that possess the optimality property. In this PhD topic, we aim to study a new machine learning algorithm using deep learning to improve energy-efficiency and learning performance of the rule-based fuzzy clustering algorithms. Experimental analysis and evaluations of the proposed approach will be done to show that our approach performs and scales better than the other approaches used in comparison.

REFERENCES

1. S. A. Sert, H. Bagci and A. Yazici. ``MOFCA: Multi-objective fuzzy clustering algorithm for wireless sensor networks". Applied Soft Computing, Vol.30, pp. 151-165, 2015.

2. M. Lotfinezhad and B. Liang. ``Effect of partially correlated data on clustering in wireless sensor networks". Proceedings of the IEEE International Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON), Citeseer, pp. 172-181, 2004.

3. A.A. Abbasi and M. Younis. ``A survey on clustering algorithms for wireless sensor networks". Computer Communications, Vol.30, pp. 2826–2841, 2007.
4. F. Kuhn, T. Moscibroda and R. Wattenhofer. ``Initializing newly deployed ad hoc and sensor networks". Proceedings of the 10th Annual International Conference on Mobile computing and networking (ACM MOBICOM), pp. 260-274, 2004.

Adnan YAZICI (Computer Science, SST), adnan.yazici@nu.edu.kz

Co-Adviser: Sain Saginbekov (Computer Science, SST), sain.saginbekov@nu.edu.kz

Title: Handling Various Security Issues in Multimedia Wireless Sensor Networks (MWSNs) with Machine Learning

Multimedia Wireless Sensor Networks are resulted from efficient data gathering requirements occurring in indoor and outdoor environments. A great deal of WSNs operates by sensing the area of-interest (AOI) and transmitting the obtained data to a sink/(s) in order to be used in object detection, object classification, instance localization, or high level semantic information extraction processes. In this regard, security of raw and relayed data are both crucial and susceptible to malicious attempts considering the efficiency of the network. In this PhD thesis topic, we will study the effects of various security issues including selective forwarding attack taking place in the network layer and highlight possible results of successful attacks through experimentation. The study will also include how you can protect the sensitive data from various attacks. We will Machine learning techniques for both detecting the attacks and take an intelligent action to protect the system by predetermining the possible attacks. Analysis and evaluations will be done on selected multimedia sensor network information fusion architecture and obtained experimental results.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey”, Computer Networks, Vol.38, No.4, pp.393422, 2002.
[2] H. Zhang, X. Chu, W. Guo, and S. Wang, “Coexistence of Wi-Fi and Heterogeneous Small Cell Networks Sharing Unlicensed Spectrum”, IEEE Communications Magazine, Vol.53, No.3, pp.158-164, March 2015.
[3] C. Karlof and D. Wagner, “Secure routing in wireless sensor networks: attacks and countermeasures”, Ad Hoc Networks, Vol.1, No.2-3, pp.293315, 2003. [4] X.800 : Security architecture for Open Systems Interconnection for CCITT applications, http://www.itu.int/rec/T-REC-X.800-199103-I/e, 1991.

Prof. Bolatzhan Kumalakov bolatzhan.kumalakov@nu.edu.kz

1. Distributed management and refinement of sensor data

Autonomous machines generate massive data sets, which may be refined into knowledge about their performance. Academic literature provides numerous examples of how such knowledge is used to decrease carbon emissions, increase productivity and enhance managerial decision making. However, organizing decentralized exchange of such knowledge between geographically distributed machines remains a challenging task. Research is interested in designing and prototyping context aware, distributed knowledge management framework, that would let machines – located in different areas and environments – exchange knowledge, learn from it and optimize their individual performance.

[1] Kannisto P., Hastbacka D., Kuikka S. (2017) System architecture for mastering machine parameter optimisation. Computers in Industry, v. 85, p. 39-47

[2] Fountas S., Sorensen C.G., Tsiropoulos Z., Cavalaris C., Liakos V., Gemtos T. (2015) Farm machinery

management information system. Computers and Electronics in Agriculture, v. 110, p. 131-138

Perspective candidate will be expected to get familiar with machine learning and clustering algorithms, context recognition and data processing technologies. Although prior knowledge in the named areas is an advantage, it is not compulsory.

Prof. Kok-Seng Wong (Computer Science, SST) kokseng.wong@nu.edu.kz

Title: A Collaborative Framework for Distributed Privacy-Preserving Machine Learning

The environment in which a machine-learning (ML) algorithm operates has driven recent progress in machine learning. In a classical machine-learning system, it involved a single program running on a single machine. With the rapid development of network technologies, it is now common to deploy the machine learning system with distributed architecture design. However, there is growing recognition that such deployment causes privacy concerns to the data owners and model owners. For instance, the goal of knowledge extraction from a large amount of distributed data collides with the privacy of individuals. On one hand, data owners can extract useful patterns from collected data using these technologies. On the other hand, these technologies can become a threat for individual privacy as well as sensitive private patterns (i.e., privacy of model). If a data owner does not trust the machine learning system, he or she may provide false data or no data at all. This can cause the machine-learning model susceptible to accuracy problem (e.g., producing an over fitted model that will be inaccurate). In this project, our aim is to integrate privacy-preserving techniques into existing machine learning algorithms such as classification, clustering, decision tree and regression analysis. Our solution attempts to reach a desirable balance between the goals of machine learning tasks (data utility) and protection of sensitive data (data privacy) at different stages for different players (data owner, model owner, and user).

[1] Yee Jian Chew, Kok-Seng Wong, and Shih Yin Ooi, Privacy protection in machine learning: The state-of-the-art for a private decision tree, in Security and Authentication: Perspectives, Management and Challenges. 2017. p. 13-39.
[2] Yee Jian Chew, Shih Yin Ooi, Kok-Seng Wong, Ying Han Pang and Seoung Oun Hwang, "Evaluation of Black-Marker and Bilateral Classification with J48 Decision Tress in Anomaly based Intrusion Detection System", Journal of Intelligent & Fuzzy Systems (ISSN: 1064-1246), 2018. 08 (First Online).

Prof. Vasileios Zarikas vasileios.zarikas@nu.edu.kz

1. Bayesian Networks for engineering applications

The only mathematically consistent methodology to drive rational decisions is the framework of Bayesian reasoning. New techniques for the problem of assigning probabilities to large Conditional Probability Tables (CPT) will be developed. Automatic ways to fill the CPTs from fuzzy rules or incomplete databases will be structured [1]. Furthermore, the utilization of Dynamical Bayesian networks in various engineering decision problems will be explored. Special focus will be given to apply Bayesian reasoning in renewable energy applications such as wind energy, photovoltaic energy or smart grids[2]. Requirements: mathematical and programming skills in C

[1] Vasilios Zarikas, E. Papageorgiou, Expert Systems,
Volume 32, Issue 3, 1 June 2015, Pages 344-369
[2] Mónica Borundaa, O.A. Jaramillo, Alberto Reyes, Pablo H. Ibargüengoytia, Renewable and Sustainable Energy Reviews 62 (2016) 32–45.

2. Regulatory framework for Artificial Intelligence (AI)

Last years, there is a debate about how to regulate and in what extend AI, [1,2,3]. A state regulatory framework seems necessary to avoid harm. However, regulations and laws are also a slow-moving tool which may create several problems in the development of technology. Furthermore, regulation experiences political interference and public unreasonable fear.
AI is a very fast-moving area of research and unwise regulations may set a barrier to innovation and future developments. AI can propose important solutions to vehicle safety, improved productivity, health system etc. Nevertheless, AI experts forecast also potential significant risks when true AI systems will be developed capable to develop new non trivial knowledge (mathematical theorems, patents etc.). Elon Musk has urged U.S. governors to regulate AI "before it's too late". AI systems will eventually overtake the ability of humans to understand future Science. The latter means that humans will not be able anymore to follow the reasoning of an AI system decisions, assuming the best case that AI will provide explanations (which is not always true i.e. Neural networks).
The regulatory framework will be developed in a mathematical consistent way using Bayesian networks and relevant mathematical utilities that will represent the involved ethical policies and the rival importance of the proposed AI solution.

[1] Point: Should AI Technology Be Regulated?: Yes, and Here's How *By Oren Etzioni .Communications of the ACM, December 2018 (Vol. 61, No. 12)*
[2] Counterpoint: Regulators Should Allow the Greatest Space for AI Innovation *By Andrea O'Sullivan, Adam Thierer Communications of the ACM, December 2018 (Vol. 61, No. 12)*
[3] Regulating Artificial Intelligence Systems, Harvard Journal of Law and Technology, By Mathew Scherer, Volume 29, Number 2, Spring 2016