Resumes

from PhD Scientific Seminar

Department Computer Systems and Technologies

Faculty Eletronics and Automation

Technical University Sofia branch Plovdiv,

19 March 2020

List of Authors

GEORGI ISKROV.....4
RADOSLAV FURNADZHIEV.....6
TEODORA MECHEVA.....8
VESELKA PETROVA-DIMITROVA.....10
TEODORA HRISTEVA.....12
VASIL TSVETKOV.....14
VASIL POPOV.....16
STEFAN LISHEV.....18
DIMITRE KROMICHEV.....20
BOYAN KATSARSKI.....22
MILENA ANGELOVA.....24
DONKA NESHEVA.....26
NIKOLAY NIKOLOV.....28

List of resumes

Research and implementation of blockchain-based network security.....4
Survey of methods and architectures for deployment and testing of containerized microservices 6
Modelling and analyzing the behavior of intelligent cognitive agents.....10
Deep Learning Model for Image classification.....12
Smart home presence and activity detection: from light control to security and fire alert system 14
Development of Mobile Robot Target Recognition and Following Behavior Using Deep Convolutional Neural Network and 2D Range Data.....16
Specialized measuring system for analysing thermal fields in hybrid systems and fluid streams18
A speed and accuracy focused approach to fpga based canny edge detection computations…20
Algorithm for collaborative P2P editing of a document stored as a dictionary and an index...22
A split-merge evolutionary clustering algorithm.....24
A machine learning approach to predict hypoglycemia occurrence for patients with type 2 diabetes.....26
An algorithm for automated transformation of the information from relational databases in json and xml.....28

Georgi Iskrov is PhD student in his first year at the Technical University of Sofia - branch Plovdiv, Faculty of Electronics and Automation (FEA). He is currently working as Network administrator at the Bulgarian company Solarity BG Ltd. His PhD research interests and activities are in the Internet security, prevent Network Level attacks, IPsec, SSL VPN.

RESEARCH AND IMPLEMENTATION OF BLOCKCHAIN-BASED NETWORK SECURITY

GEORGI ISKROV

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
iskrovgd@abv.bg

Initially, the Internet was conceived as an environment for discrete internal communication. No malicious actions were foreseen for the users. At present, the World Wide Web (WWW) has undergone unthinkable development, compared to the original intent of its creators. This has further added functionality providing protection. The need for transparency for consumers, the network level protection, led to the creation of the standard IPsec. The IPsec protocol was implemented primarily in the IP layer of the PCP/IP stack. IPsec uses multiple standard cryptographic technologies. It was originally designed to work with IPv6, but due to the urgent need for increased security, it has been redesigned and can also be applied in IPv4.

The purpose of the PhD work is to achieve higher security in the platforms used by the business. To make it easier feasible for companies to have increased security in accessing their Internet sites. The events of the past year - hacker attacks, theft of personal data from public servers, show unambiguously the need to increase security when working with the Internet.

The IPsec protocol is mainly used to organize VPN tunnels. In this case, the ESP and AH protocols work in tunneling mode. In addition, by setting up security policies in a certain way, the protocol can be used to create a firewall. The point of a firewall is that it controls and filters the packages that pass through it in accordance with the specified rules. A set of rules is set, and the screen looks through all the packages that pass through it. If the transferred packages fall under these rules, the firewall handles them accordingly. For example, it can reject certain packages, thereby interrupting unsafe connections. By setting up a security policy accordingly, you can, for

example, ban web traffic. To do this, it is enough to prohibit the sending of packages in which the messages of http and HTTPS protocols are invested. IPsec can also be used to protect servers, and all packages are discarded, except for the packages needed to perform the server functions correctly. For example, for a Web server, you can block all traffic, except for connections through the 80th port of the TCP protocol, or through the TCP 443 port when HTTPS is used.

Radoslav Furnadzhiev is 1st year PhD student at the Technical University Sofia branch Plovdiv, Faculty of Computer Systems and Technologies. He has experience with rapid development of enterprise applications for Continuous Integration & Deployment. His PhD research interests and activities are in exploring Orchestration and Coordination architecture patterns for achieving scalability and availability in modern containerized applications. Radoslav holds an MSc degree in Computer Science from the same university.

SURVEY OF METHODS AND ARCHITECTURES FOR DEPLOYMENT AND TESTING OF CONTAINERIZED MICROSERVICES

RADOSLAV FURNADZHIEV

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
rfurnadzhiev@tu-plovdiv.bg

Modern day world of software engineering is shifting from designing and developing large monolithic systems towards microservice based technologies in building enterprise applications. Microservice architectures distribute the application into small modules, each of which can be deployed and scaled independently of each other. Those design principles consisting of small collaborating services, each running in its own process and communicating with lightweight mechanisms, intend to overcome the drawbacks of monolithic architectures where all of the application's logic and data are managed in a single deployable unit. Along with this architectural approach, several deployment technologies have emerged, such as container-based virtualization and container orchestration solutions. These technologies allow to efficiently exploit cloud platforms, providing a high degree of scalability, availability, and portability for microservices. Each microservice could be deployed, updated, and scaled, without any impact on other microservices. Continuous Integration and Continuous Delivery play an important role in the development process of microservice based applications. Continuous Delivery principles provide process for fast release of changes to customers in a quick and sustainable way. On top of automatic deployment, microservices should be certified by functional, integration, performance, and security tests. Elaborate testing pipeline infrastructures are considered mandatory to enable

this form of end-to-end testing where the test suite of every service is executed to confirm there aren’t any regressions or breaking changes being introduced. Regression test is one of the tests used for ensuring reliability, which compares two system versions based on various metrics.

I intend to research microservices architectural design along with the various advantages, disadvantages of containerized microservices and automated methods of running tests in Continuous Delivery environment.

Teodora Mecheva is a PhD student in her 1st year at the Technical University of Sofia – branch Plovdiv. Her PhD research interests are Intelligent Transportation Systems methods and utilities. Teodora holds a MSc degree in Computer Science from the same university.

A SURVEY OF THE COMMUNICATION PROTOCOLS FOR VEHICULAR AD HOC NETWORKS

TEODORA MECHEVA

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
teodora.mecheva@tu-plovdiv.com

Modern vehicles are complex cyber physical systems. To realize complex goals of ITS (Intelligen Transportation Systems) they need to communicate each other (Vehicle to Vehicle) or to infrastructure (Vehicle to Infrastructure) forming Vehicular ad-hoc networks (VANETs).

Due to dynamic characteristics of VANETs – extraordinary volatile topology, tough time constraints, communication during high-speed traffic and high demand of security there are open issues about the technologies that meet the requirements of automotive applications.

Applications in ITS may be classified as safety and non-safety. Another classification differentiate safety, infotainment and traffic management application. They have different characteristics corresponding to different services in lower protocol layers.

Even there are number of approved standards in the area still remain issues and science is looking for better solutions.

One of the most widespread standards is the proposal of IEEE 1609 called WAVE (Wireless Access in Vehicular Environment). Alternative proposal is ISO's (Internet Standard Organization) architecture - CALM (Continuous Air interface for Long to Medium range) and C2C-CC (European Automotive Consortium).

Probably on MAC and physical layer the most approved protocol is IEEE 802.11b. On network layer there is still investigations and argues about routing and proliferation of the

messages.

The presented survey attempts to cover the most common VANET standards and to identify some of the tasks that may be subject to future research.

Veselka Petrova-Dimitrova is a PhD student in her second year at the Technical University of Sofia, branch Plovdiv, Computer Systems and Technologies Department. She received M.Sc. degree in “Computer Systems and Technologies” at Technical University of Ruse. She worked Her research interests are in analyzing the methods of knowledge representation and decision making algorithms, agent’s learning algorithms and their use in modeling rational behavior of intelligent agents.

MODELLING AND ANALYZING THE BEHAVIOR OF INTELLIGENT COGNITIVE AGENTS

VESELKA PETROVA-DIMITROVA

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
vesi_s_petrova@yahoo.com

The goal of the thesis is to model and study the behavior of intelligent agents (IA) and to investigate computational capabilities that enable IA to be involved in the governing, unstructured, and dynamically evolving social interactions that characterize the person. The research is aimed at exploring and analyzing methods for presenting knowledge and decision-making algorithms, training and self-study methods and using them to model the rational behavior of IA. Special attention is paid to the study of decision making in conflict situations when mixed emotions arise. The expected results will be scientifically applied in the field of artificial intelligence and intelligent systems.

At this stage of implementation, a prototype of Intelligent Smart Shopping System has been proposed. It includes technologies: text-to-speech, speech recognizing, Bluetooth low energy technology, holographic technology, information kiosks, picture exchange communication system, learning virtual agents. There are three Smart Shopping Cart Learning Agents in the process of implementation and research - Goal-based learning agent; Utility based learning agent; Personal learning agent.

The specific implementation of shopping agents can be installed and operate on every shopping cart in shopping malls or holographic screens. The following virtual IA training algorithms have been implemented: Reinforcement learning; decision k-d tree; identification tree. The utility-based agent uses a decision k-d tree to quickly find where (in which shop) the

customer is located according to his/her coordinates. It getting to the new promotion in the shopping mall according to user‘s shopping list and inform them. Reinforcement learning algorithm is used for the other Goal-based learning agent. The agent gets the shopping list from the customer and informs the customer about the sequence in which he/she can visit the shops to buy all needed goods. The personal utility-based agent makes the best identification tree according to the shopping data of each user. That way the agent knows their shopping habits and can suggest a shopping route and check if there are promotions on the goods from the categories, in which the customer could be interested.

The purpose of the thesis is to modify and improve these training algorithms and to investigate the IA's behavior when applied. It is intended to implement 3D visualization of training agents for smart shopping carts and prototype smart shopping carts. The system will use virtual reality technology. The implemented algorithms and models will be used for educational and scientific purposes in national and international Universities, as well as for practical implementation. The proposed smart system is expected to enhance the people with various communications capabilities to target and shop in major shopping malls according to their shopping list, promotions and habits. It will contribute to the creation of "Intelligence Centers" and is in line with "Innovative Strategy for Smart Specialization” thematic areas 1 and 2 "Informatics and ICT" and "Mechatronics and Clean Technologies" and more specifically Intelligent systems and appliances, "smart homes" - "smart cities", 3D digitization, visualization and prototyping will enhance the quality of training.

Teodora Hristeva has a master's degree in "Information technology" from Technical University Sofia, branch Plovdiv. She has more than 15 years of experience in software projects using variety of technologies with a special impact on Java and C#. 8 years of this period she spent in Barcelona, Spain working for multinational companies and the rest in Sofia and Plovdiv, again working for big firms and also in a start-up. Currently she is a PhD student in the Computer Systems and Technologies Department with topic "Parallelizing Deep Learning algorithms using graphic accelerators". Her interests are in big data processing and analyzing.

DEEP LEARNING MODEL FOR IMAGE CLASSIFICATION

TEODORA HRISTEVA

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
thristeva@gmail.com

Deep learning is entering in the everyday life of people in different forms. The reasons for this are the continuous development of computer systems, the increase of their computing power and the increase of data stored on electronic media. The main goals of developing self-learning models are to improve or replace existing methods for processing large amounts of information, to improve the services offered by different institutions, and generally to improve and facilitate the lifestyle of modern man. Machine learning can be used to detect complex relationships between a large set of input data, making it an appropriate method for solving a wide range of issues in different spheres.

Neural networks are information processing systems inspired by the biological neural networks that make up the human brain. These systems are taught by task execution without being explicitly programmed with task-specific rules. The main purpose of neural networks is to solve problems that the human brain can solve. Typical examples are computer vision, natural language processing, language translation, and so on.

We have designed a model to work with the classification of images in selected criteria

Architecture of the Model

Convolutional layer: This is the main layer in the building of convolutional neural networks. It shows the sequence of actions performed by the convolution layer. The kernel and bias blocks

represent the weights and deviations of the layer respectively. The activation function - ReLU - is performed on the sum. All the convolutional layers used include this activation function.

Pooling layer: Only one action is performed in this layer and it is the application of the MaxPool operation. The layer is used to reduce the size of the input data. In this case, it is reduced 2 times on the "x" axis and 2 times on the "y".

Fully-connected layer: In the fully-connected the kernel and bias blocks are tensors with weights and layer deviations respectively. The MatMul operation performs a vector work between the kernel weight and the input data. The activation function ReLU is performed on the sum. The activation function is an addition to this layer. This is the last layer of the model, it calculates the assumptions made by the model. On their basis, the accuracy and probabilities of each assumption are calculated, and the values with which the weights of the model must be changed.

Dropout layer: The created pattern has only one Dropout layer. This layer is used to adjust the input data and is only used in model training mode. Dropout is a technique to reduce the chances of the model to adapt to training data and to achieve high results only (overfitting). Under this technique, part of the data is removed at random. Sets a ratio of how much of the data will be saved.

This architecture works well with the classification of images in selected criteria. The amount and quality of training and evaluation data is the most influential on the results achieved. When data quantity and quality are increased, better results are expected. Despite the limited number of training data, the model achieves approximately 77% accuracy on the data and a loss value of 0.75-0.95.

The techniques used to improve the initial results increase the accuracy of the projected model by 12% to 15%. This is mainly due to the reduction in the effect of overfitting. The greatest benefit to this is the artificial increase of training data, the reduction of the discovered characteristics during the training, the adjustment of the speed of training and the used weight optimizer.

Currently we are working on parallelizing this model and in the future we will have a more powerful network as deep learning and especially training convolutional neural networks can be time-consuming, even more so if big data sets are used. Parallelization is a good approach to reduce this cost.

Vasil Tsvetkov is a PhD student in his 2nd year at the Technical University of Sofia – branch Plovdiv. His PhD research interests and activities are embedded systems, Internet of Things, cloud based intelligent sensor systems. Vasil holds a MSc degree in Computer Science from the same university.

SMART HOME PRESENCE AND ACTIVITY DETECTION: FROM LIGHT CONTROL TO SECURITY AND FIRE ALERT SYSTEM

VASIL TSVETKOV

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
vaccc@abv.bg

In the recent years there is a significant increase in the use of electronics and networks that make daily tasks a bit easier. IoT has emerged as a leading field in the IT and the low cost of microcontrollers and small-sized sensors now is easier than ever to introduce a smart object in our home and office. There are many currently existing fields where infrared emission of the objects is measured, but with the increasing of the capabilities of the affordable microcontrollers and IR sensors it is possible to develop low-cost solutions for applications of infrared measurements in everyday environment. In a smart home network the IR sensor could be used as a presence device or temperature sensor.

The people detection, counting and localization are considered important and key control variables in smart sensing and modern HVAC on demand systems. Having the precision of a fraction of a degree Celsius and based on the size of the object it is possible to distinguish different objects like animals that could potentially falsely alert that there is an intruder.

In certain scenarios that are provided from the user the thermal image can tell us about potential fire hazards – for example if we have left the stove/toaster/heater on for too long, or even left it and gone out: if no other activity is spotted and the appliance is on it would be better to alert the owner for potential danger. Another scenario is when the homeowner is going to be absent for some time and any movement in the house should be considered as intrusion and the owner should be alerted as soon as possible.

My goal is to analyze the communication protocols, network architecture, microcontroller and sensor specifics and present the possible solutions for some of the most common cases in the Smart Home technology.

Vasil Popov is PhD student at the Technical university of Sofia – branch Plovdiv, Faculty of Electronics and Automation (FEA). He received B.Sc. degree in “Automation, Information and Control Systems” in 2014 and M.Sc. degree 2016. He has experience in developing PLC applications for manufacturing purposes. He is currently working as an Assistant Professor at the same university. The subject of his PhD is “New approaches for adaptation and training in controlling of intelligent robots”. His main research interests and activities are in the area of machine learning, deep learning, image processing, human-machine interfaces and adaptive control.

DEVELOPMENT OF MOBILE ROBOT TARGET RECOGNITION AND FOLLOWING BEHAVIOR USING DEEP CONVOLUTIONAL NEURAL NETWORK AND 2D RANGE DATA

VASIL POPOV

Control Systems Department
Technical University of Sofia – branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
vasil_popov@tu-plovdiv.bg

Recognition and tracking of dynamic objects play an important role in the development of service robots that have to co-exist with humans and other autonomous machines in a shared environment. It is also an important behavior for the mobile nodes within the structure of a robotized wireless sensor networks because they have to operate as a multi-agent formation. In this investigation an approach is proposed to use a deep learning convolutional neural network (CNN) and 2D LiDAR data for recognition and following of pre-specified dynamic objects. Here the notion following robot, implies an implemented method that makes possible to maintain a given relative positional relationship between the followed dynamic target and the robot. To be capable to exhibit a following behaviour the mobile robot has to be able to identify continuously the moving dynamic targets. The existing approaches allowing detection and tracking of dynamic targets using the sensors installed on a mobile platform can be broadly classified into two group. The model-free methods form the first group. The process of target detection in these methods relies on the motion cues. That is, they track a dynamic object over time and identify it using only the motion continuity. The advantage of these methods is that they do not impose restrictions on the shape or class of the detected object and do not require additional semantic information. Their drawback is that only moving objects can be detected. Potentially moving targets are thus

neglected. In its part related to the processing of 2D LiDAR data, the approach proposed in this investigation can be referred to the methods from this group. Laser range finders are commonly used for detection and tracking of moving targets thanks to their reliability, high sampling rate and wide field of view. The methods from the second group are model-based. The targets that these methods will have to detect should be known a priori by their shape or other features, generally described as parametric model. Most of them are image-based methods that are able to provide more information about the tracking objects including colour of clothing, height, face and skeletal information. The methods from this second group can detect potentially moving targets, but are restricted to targets described by an implemented parametric model. The second part of the proposed here approach, where the pre-specified target objects are recognized using a deep learning convolutional neural network can be considered as a model-based one.

The proposed approach is implemented using a KUKA youBot omnidirectional mobile platform equipped with an onboard USB camera and a 2D LiDAR. During the conducted experiments the designed deep learning neural network is able to recognize and localize an existing in the lab mobile robot iRobot Create, based on the images obtained from the onboard camera. The developed algorithm then localizes the recognized object and begins considering it as a potential dynamic target. The on-board 2D LiDAR can detect when the above target starts moving and will begin to track it. Meanwhile the robot will start following the selected target by controlling its own linear and angular velocity using the implemented PD controllers.

Stefan Lishev is PhD student at the Technical university of Sofia - branch Plovdiv, Faculty of Electronics and Automation (FEA). He received B.Sc. degree in “Computer Systems and Technologies” in 2008 and a M.Sc. degree in 2010. The subject of his PhD thesis has been “System for automation of scientific experiments with remote and mobile control” but now it is changed to “System with remote and mobile access for automation of thermal fields and fluid streams measurements”. He works now at the Department of Computer Systems and Technologies. His main interests are embedded systems, artificial intelligence, digital signal processing, Internet of Things. Stefan has five published articles in the field of his PhD thesis.

SPECIALIZED MEASURING SYSTEM FOR ANALYSING THERMAL FIELDS IN HYBRID SYSTEMS AND FLUID STREAMS

STEFAN LISHEV

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
stefan_lishev@tu-plovdiv.bg

Conducting research and experimentation requires controlled conditions to achieve the required precision, performance and safety. In cases where experiments are exploring slow processes, continuous measurements are required from days to months and storage of the obtained data. Prior work on the subject of the dissertation is design of monitoring system with remote and mobile control. It consists of a battery powered PIC microcontroller based logging module with capability for long-term work that connects through a wireless 802.15.4 connection to a server for storage and visualization of the received data. Server database access and experiment parameter control can be done via WEB page and a mobile Android application that allows convenient and flexible management and monitoring that meets current requirements and trends in the field. The system is designed for analyzing thermal fields in phase change materials storages in hybrid systems. It is also suitable for related applications like chemical and biochemical reactors, and columns, if the temperature is in the range of -20 ° to +125 °C. Data is stored in SD-card and can be transferred via USB interface to the PC by specialized software, where the thermal fields could be visualized.

The second part of the system is for studying of the velocity field in the work chamber of the

wind tunnel ULAK -1. The measurements are done using five-hole probe in a plane located behind the subject of study. In measuring the magnitude and direction of the air flow velocity in low speed wind tunnels the probes with a spherical end portion are used. The five-hole probe is applied in this case. Central hole was located at the tip of the probe, the four side holes were symmetrically located at 45° with respect to the frontal probe axes X. For automation of the measurements and storage of data is designed special software, written in the language of LabView.

The new direction of research is Internet of Things. The focus is mainly on developing and testing new protocols and algorithms for communication, as well as investigating existing ones between embedded devices that use different types of communication – wired and wireless. Different scenarios are going to be test via simulating different kinds of networks of devices with the help of Network Simulator 2 and similar tools. After doing simulations the promising solutions will be tested as real systems.

Dimitre Kromichev is an engineer holding a Master of Science Degree in Computer Systems and Technologies. He is an assistant professor at the Department of Marketing and International Economic Relations, University of Plovdiv.

A SPEED AND ACCURACY FOCUSED APPROACH TO FPGA BASED CANNY EDGE DETECTION COMPUTATIONS

DIMITRE KROMICHEV

Department of Marketing and International Economic Relations
University of Plovdiv
24 Tzar Asen Street, Plovdiv 4000, Bulgaria
dkromichev@yahoo.com

Conducting research and experimentation requires controlled conditions to achieve the John Canny's algorithm for contour detection is a precise and reliable image processing technology. Its FPGA implementation faces two problems: speed and accuracy.

In the publications on Canny there are two general assessments: 1) Canny's mathematics is too complicated to be implemented in FPGA without approximation; 2) Canny is too slow and, therefore, inappropriate for real-time implementation. The objective of this work is to design and explore a new organization of FPGA based Canny computations guaranteeing optimal speed on the basis of total mathematical accuracy.

Explored is the impact of the entire set of approximations on detected contours' precision. This requires the development of a set of tools and methodologies. Both synthetic and real life images are used to prove the strong negative impact of approximations on precision.

Defined and explored are six capital parameters of FPGA based Canny computations focused on maximum performance: pipelining, memory, parallelism, number of uses of a single pixel, integer arithmetic, input data width. Designed is a set of computational algorithms which combine 100% mathematical accuracy with optimal speed. The algorithms' performance is evaluated on a comparative basis with respect to maximum clock frequency and minimum number of clock cycles required to compute an accurate result. Presented are the utilized resources. The following Intel (Altera) FPGA platforms are used for the tests: 130 nm – 28 nm Cyclone I – V and 130 nm – 28 nm Stratix I – V.

Designed and explored is an entirely new approach to the organization of computations in each of the five Canny modules. Experimentally proven is the upper limit of frequency of operation in FPGA based Canny. Speed capabilities are explored on a comparative basis. For the purpose, developed is a specific methodology encompassing Register transfer level modeling, testing detected contours' precision and analyzing the working FPGA based Canny using real-life images. Tested and proven is the capability of the proposed organization of computations to guarantee positive slack for every single mathematical operation.

The relevant conclusions to the study are being drawn. References are being updated.

Boyan Katsarski is PhD student in his third year at the Technical university of Sofia - branch Plovdiv, Faculty of Electronics and Automation (FEA). Boyan holds a MSc degree in Computer Science from the same university. He is currently working as an assistant at the Department of Computer Systems and Technologies. His PhD research interests and activities are in Distributed Computing, Convergence of State in Partitioned Distributed Systems, Collaborative Editing Systems, Conflict-free Replicated Data Types and Operational Transformations algorithms.

ALGORITHM FOR COLLABORATIVE P2P EDITING OF A DOCUMENT STORED AS A DICTIONARY AND AN INDEX

BOYAN KATSARSKI

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
katzarsky@gmail.bg

Most of the CRDTs for text editing are based on atoms that are characters. This is a problem because meaning is contained not within a single character but rather within words, phrases and sentences. Assigning blame for badly formed phrases is difficult since one word may be edited by couple of users. A better solution may be using words or phrases as atoms and improve methods like LOGOOT.

These word-atoms form a dictionary that will enable inverse document indexing right from the start. The inverse indexing will enable built-in search of the edited text. The document structure will use dictionary indices instead of the words themselves – making the document itself shorter.

The dictionary can be partitioned in sections owned by the users. Since a section is owned only by one user – it can be garbage collected only by the owner-user and not requiring a consensus or a lock protocol.

Adding or deleting a character in a word will introduce temporary noise in the dictionary since the currently edited word will change. But this will affect only the private section of the user-dictionary. If the word was not owned by the editing user the new word is created in the user-owned partition of the dictionary.

Milena Angelova is currently a Ph.D student at the Department of Computer Systems and Technologies of the Technical University of Sofia - branch Plovdiv. She received a B.Sc. degree in Automation and Control Systems in 2013 and an M.Sc. degree in Computer System and Technologies in 2015 from the same university. Her Ph.D. thesis supervisors are Prof. Veselka Boeva from the Technical University of Sofia - branch Plovdiv and Elena Tsiporkova from Sirris (Belgium). The subject of Milena's Ph.D. thesis is "Intelligent Approaches for Profiling, Analysis and Localiza4on of Exper4se from Public Online Sources." Her research interests are focused on knowledge discovery and information retrieval. This entails developing of retrieval techniques that support humans in dealing with massive volumes of data. In her Ph. D. work she has proposed efficient clustering techniques for modeling and analysing of streaming data. Her primary research interests and activities are in the field of Information Retrieval Systems, Machine Learning, Data Mining, Semantic Web and Recommender Systems.

A SPLIT-MERGE EVOLUTIONARY CLUSTERING ALGORITHM

MILENA ANGELOVA

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv Tsanko
Dyustabanov 25, 4000 Plovdiv, Bulgaria
mangelova@tu-plovdiv.bg

Most of the organizations are continually dealing with data that comes from their employees, customers, potential candidates or their external sources. Applications as expertise retrieval and healthcare decision support systems adapt their databases with available information periodically. Typically, the stored data is partitioned in disjoint subject categories. The existing clustering techniques as incremental and data stream algorithms present some challenges. For example, incremental clustering algorithms receive data elements one at a time and then re-cluster the data all of the data from scratch. Data stream algorithms lead to scalability issues and memory limitations. Practically, it is becoming impractical to re-cluster the massive amount of data whenever a new portion of data arrives. The drawbacks of the existing methods motivated us to focus on developing evolutionary clustering technique that can adapt the newly incoming data without a need to re-clustering the existing clustering.

In this work, we propose evolutionary clustering techniques that are suited for applications

affected by concept drift. We have developed a bipartite correlation clustering technique that can be used to adapt the existing clustering solution with new arriving data. The suggested technique is supposed to provide the flexibility to compute clusters on a new portion of data collected over a time period and to update the existing clustering solution by computed a new one. For example, some clusters will be transformed by splitting their elements among several new clusters and others will be updated by merging them with ones from arriving data. The proposed algorithm, entitled Split-Merge Evolutionary Clustering, is evaluated and compared to another bipartite correlation clustering clustering PivotBiCluster and incremental evolving clustering Dynamic split-and-merge algorithm on four different test data sets.

For future work, we are planning to extend this study with new case studies and to evaluate both algorithms on richer data sets that are being updated in different time periods.

**Based on: V. Boeva, M. Angelova, E. Tsiporkova. A Split-Merge Evolutionary Clustering Algorithm, 11th International Conference on Agents and Artificial Intelligence ICAART 2019 (Prague, Czech Republic, February 19-21, 2019).*

Donka Nesheva is a PhD student at Sofia Technical University – branch Plovdiv, Faculty of Electronics and Automation (FEA). She was an Assistant Professor at the same university for several years. She has experience in developing Health Information Systems (HIS), Electronic Health Records (EHR) and mobile applications. Her main research interests and activities are in the area of health informatics and in particular exploring methodologies for storing and analyzing patient health data in cloud environments.

A MACHINE LEARNING APPROACH TO PREDICT HYPOGLYCEMIA OCCURRENCE FOR PATIENTS WITH TYPE 2 DIABETES

DONKA NESHEVA

Computer Systems and Technologies Department
Sofia Technical University - branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
donka.nesheva@gmail.com

Diabetes Mellitus is a chronic disease characterized by the inability of the pancreas to control blood glucose concentrations. In order to avoid serious complications, patients must continually monitor their blood glucose levels and adjust their medication doses to reach levels close to normal. Blood glucose levels that are too low (hypoglycemia) lead to complications including weakness, confusion, dizziness, sweating, shaking, and, if not treated in time, seizures, coma, or death. Severe hypoglycemia is associated with insulin-dependent diabetes and may occur more frequently as metabolic control approaches normal.

A prediction model that can notify patients for upcoming changes in their blood glucose levels would allow them to take preventive actions. This would not only improve overall blood glucose control, but could impact patients safety.

The majority of studies use time series regression models to predict future blood glucose values. This concerns patients getting blood glucose measurements every few minutes using continuous glucose monitoring (CGM), usually ones with Type 1 diabetes (T1D). The task is more complicated for patients with Type 2 diabetes (T2D) which typically submit 1 to 2 blood glucose records per day.

We approached the task as a classification problem evaluating the risk of hypoglycemia

occurrence for patients with T2D. We trained a model using decision tree classification algorithm to predict a hypo incident in the next 7 days. The training data set came from de-identified patient data of patients with T2D, collected in a platform for remote patient monitoring. Patients had their data submitted by means of mobile application or cellular blood glucose meter.

We used a set of input variables that is a combination of glycemic, medical and personal data. A glycemic profile based on the glucose readings for the last 7 days and statistical data for the previous 30 days from a given day was set for every patient. In addition, we used sex, age, HgA1C, creatinine, weight, height, BMI, LDL, eGFR, BP, recently occurred complications, medication class and dosage.

We validated the model in a clinical trial with real patients and the prediction accuracy reached over 96%. We deployed the prediction model in a remote patient monitoring system. The providers are being notified for patients at risk of hypoglycemia, and they can recommend treatment changes and inform patients to prevent incidents in time.

Further work should be focused on using other machine learning algorithms like random forest or SVM and comparing the performance. A direct patient notification for risk of hypo event should be implemented too.

Nikolay Nikolov is PhD student in his fourth year at the Technical university of Sofia - branch Plovdiv, Faculty of Electronics and Automation (FEA). Nikolay holds a MSc degree in Computer Science from the same university. He is currently working as an assistant professor at the Department of Computer Systems and Technologies. His PhD research interests and activities are in the Language processors for formal languages, Database systems and Parsing of semi-structured data.

AN ALGORITHM FOR AUTOMATED TRANSFORMATION OF THE INFORMATION FROM RELATIONAL DATABASES IN JSON AND XML

NIKOLAY NIKOLOV

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
n.nikolov@tu-plovdiv.bg

There are many different decisions of transformation from relational data to JSON or XML format, but they are either ad hoc decisions or decisions, which required additional (in some case very detailed) description of the transformation process.

The proposed algorithm is a universal decision for automated transformation from relational databases into semi-structured data (JSON or XML). The aim is to be created a tool, which will be able to transform the data, retrieved from relational database with arbitrary SQL query into semi-structured data (JSON or XML). This tool must be able not only to “pack” data into the required output format. It must be able to recognize the hidden “hierarchical nature” of the “flat” relational data and automatically transform this kind of data into JSON or XML format, preserving existing relationships between them. This transformation must be done without (or with very minimal) intervention from the user.

The algorithm consists of three stages:

1. An analysis of the input SQL query.
2. Shredding the input SQL query into several “atomic” SQL queries.
3. Output document generation.

The most important part of the algorithm - Analysis, must answer the following questions:

1. From which relational tables will be retrieved the information?
2. Which fields from relational tables are involved?
3. What are the relationships between the involved relational tables: which table is Master table and which table is Detail table. Which fields are involved?