



PhD Scientific Seminar

Computer Systems & Technologies Department

November 25th 2016, hall 2203

Contents

Dimitar Grozev	Private cloud traffic types, monitoring and methods for optimization
Dimitar Garnevski	Estimation of the speed and direction of solar prominence in sequence of frames
Lilyana Boneva	Semantic-aware Expert Partitioning
Donka Nesheva	An Overview of Cloud Computing in Healthcare
Nikolay Nikolov	Language processors for transformation between semi-structured data streams and relational databases
Milena Angelova	A Semantic Reasoning Engine for Lifestyle Profiling in Support of Personalised Coaching
Stefan Stoyanov	Sensor Data Processing - Key Factors for Choosing File Formats in Hadoop
Georgi Pajev	Smart Home - methods and utilities
Angel Chekichev	Feature extraction concerning the development of processes in objects, based on colour characteristics in images
Evgeni Jordanov	Search Engine Optimisation: Tools & Methods
Irina Kakanakova	Deep Learning Algorithm through Adaptive Contrastive Divergence
Editors:	Veselka Boeva, Nikolay Kakanakov

Dimitar Grozev is PhD student in his 5th year at the Technical university of Sofia - branch Plovdiv, Faculty of Computer Systems and Technologies. He is currently leading a small team of Network Engineers at VMware, responsible for network design and implementation of VMware Internal Private Cloud. His PhD research interests and activities are in exploring methodologies for traffic optimization and analysis in modern Data Centers. He is doing his research in university private cloud, built with VMware virtualization technology.

PRIVATE CLOUD TRAFFIC TYPES, MONITORING AND METHODS FOR OPTIMIZATION

DIMITAR GROZEV

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

Nowadays companies and organizations of all shapes and sizes already have or they are looking to deploy cloud implementations in order to increase IT efficiency, lower the costs and improve availability and disaster recovery. “Everything as a Service” that’s the final goal of modern Data Center transformation. Becoming more and more “cloudy” in the Data Center, we more and more gain of cloud benefits, until the day when service stops. That day we realize that we have to understand cloud traffic types and to monitor them. Cloud resources are easy to add, but not limitless. This is valid not only for networking, but for storage and computing. Adding resources is the easiest, but not always right approach to solve the problem. I know that from my own experience. We need to understand traffic nature. We have to ask questions:

- Is it user or management?
- Is it broadcast, unicast, multicast or BUM?
- Is it north-south or it is east west?
- Is it burst traffic or not?
- Etc...

All cloud technologies (virtualization, overlay, etc) are now against us. They hide the underlying processes.

In our research, we classify cloud traffic types, monitor them using different technologies and suggest methods for optimization. Stay with us to understand more!

Dimitar Garnevski is a graduate engineer with a master's degree in "Computer Science", which he has received from the Technical University Sofia, Plovdiv branch. He is currently a software developer at company IBS Bulgaria. He started his PhD in 2013 in "Automated processing and control of information". Subject of his work is "Expanding opportunities for image processing and research changes in the magnetic field of the solar corona".

ESTIMATION OF THE SPEED AND DIRECTION OF SOLAR PROMINENCE IN SEQUENCE OF FRAMES

DIMITAR GARNEVSKI

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

Observations of solar atmosphere reveals wide range of movements, from small local jets to ejections of coronal mass in global scale (Coronal Mass Ejection - CMEs). Recognition of these movements in solar corona and estimation of their characteristics is a key part of the space weather prediction process. CMEs or so called protuberance are solar activity and are the greatest formations in solar atmosphere.

The problems that are associated with the processing of images of the solar corona and in particular the CME can be divided into two groups: initial detection the event or prerequisites for it and analysing event. The initial detection is often done by people, in some cases using specialized software. The analysis of events is performed by astronomers by simply using software which performing only part of the activities.

One of the main tasks is to determine the height of the observed event and the rate of progress over time. Stages of processing and analysis of the sequence of images containing CME are:

- initial filtering and preparation (clearing noise, lighting and contrast enhancement of the image)
- creating a map of the movement of the particles according to the directions on the

compass

- filtering the pixels against the desired directions - not considered particles which have a direction towards the interior of the solar corona
- creating a mask image
- in the processing of subsequent images are applied mask, in order to avoid treatment of areas that do not contain the requested information - the mask is increased in the direction from the solar corona into outer space with a percentage of their size, depending on the interval between the images
- from the generated map of the movements of the particles determine the direction and height of the event - from the solar corona of the highest limit of ejected solar substance
- from data obtained on the height of the protuberance and times of capturing solar corona can be made conclusions about the rate of change

Based on the established processing algorithm and the resulting data can be reviewed following guidelines for future development:

- developing and integrating algorithm to automatically detect emerging protuberance
- implementing a system to analysis of the relationship between the occurrence of the protuberance and the pulsation of the sun

Lilyana Boneva is a PhD student in her 4th year at the Technical University of Sofia – branch Plovdiv. The subject of her PhD thesis is “Intelligent approaches for extraction of knowledge from internet and other publicly available sources”. Lilyana holds a MSc degree in Computer Science from the same university. Her main research interests and activities are in the areas of artificial intelligence, bioinformatics and business intelligence. She is currently performing research dealing with data mining in the field of expertise retrieval and drug safety. Lilyana has one journal article and three papers in international conference proceedings.

SEMANTIC AWARE EXPERT PARTITIONING¹

LILYANA BONEVA

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
lil2@abv.bg

Expertise retrieval is not something new in the area of information retrieval. Finding the right person in an organization with the appropriate skills and knowledge is often crucial to the success of projects being undertaken.

We are presenting an algorithm for expert partitioning into groups based on degree of expertise similarity. Experts are represented with vectors of scores describing the experts' degree of relevance to a defined set of topics. Topics are formed from keywords extracted from individuals' public profiles and subsequently clustered into a set of topics. The Euclidean distance between each pair of vectors is finally calculated and the experts are clustered by applying a suitable partitioning algorithm.

Initial preprocessing step was done manually. The set of keywords used in constructing expertise profiles are gathered from Microsoft Academic Search, i.e., a researcher's scientific areas are defined by a list of keywords used in the profile page of the author.

1 V. Boeva, L. Boneva, E. Tsiporkova. Semantic-aware Expert Partitioning. G. Agre et al (Eds.): Artificial Intelligence: Methodology, Systems, and Applications - AIMS 2014, LNAI 8722, 13-24. Springer Int. Publishing Switzerland (2014)

For algorithm evaluation, a test collection from a scientific conference in the area of information technology in bio- and medical informatics is used. The lists of researchers presenting articles in each conference session are considered as relevant experts, thus, used as the ground truth to benchmark the results of the proposed clustering method.

Donka Nesheva is a 2nd year PhD student at the Technical University of Sofia – branch Plovdiv, Faculty of Electronics and Automation (FEA). She is currently working as an Assistant Professor at the same university. 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.

AN OVERVIEW OF CLOUD COMPUTING IN HEALTHCARE

DONKA NESHEVA

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

The healthcare sector as many other industries is facing the problems of high IT costs, need of more computational power, scalability, demand for interoperability, etc. Healthcare organizations are expected to provide new and improved patient-centered care capabilities at a lower cost. This is necessary to facilitate the adoption of electronic medical records (EMR), electronic health records (EHR), personal health records (PHR), and technologies related to integrated care, point-of-care access to demographic and clinical information, clinical decision support.

Cloud computing enhances the key requirements of the healthcare industry.

It enables on-demand access to computing and large storage facilities which are not provided in traditional IT environments.

It supports big data sets for electronic health records, medical images, sensor, genomic, and population data.

It improves the performance of data mining and analysis techniques in order to retrieve new medical knowledge and insights. Patient data analysis in real time ensures that the care provider makes better, faster and more accurate decisions.

It facilitates sharing EHRs among authorized physicians and hospitals in various geographic areas, providing more timely access to life-saving information and reducing the

need for duplicate testing.

However, despite the significant advantages of using cloud computing there are some concerns that impede its adoption: privacy and security, reliability, interoperability.

Integration and interoperability are key components in the healthcare ecosystem. They ensure reliable exchange of commonly understood information to facilitate coordinated patient care. Different participants (physicians, primary care providers, lab workers, scientists) have different terminologies which require standardization and interoperability.

Data maintained in a cloud may contain personal, private or confidential healthcare information that requires the proper safeguards to prevent disclosure, compromising or misuse.

Nikolay Nikolov 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 System architect and software developer at the Bulgarian company Dieselor Ltd. His PhD research interests and activities are in the Language processors of formal languages and Parsing of semi-structured data.

LANGUAGE PROCESSORS FOR TRANSFORMATION BETWEEN SEMI-STRUCTURED DATA STREAMS AND RELATIONAL DATABASES

NIKOLAY NIKOLOV

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

In recent years, XML became the preferred format for encoding and exchange of data, especially on Internet. On the other hand most of the information is stored in relational databases. Hence, the need to transfer data between relational databases and XML appears.

The cases of transfer are:

- From relational database to XML
- From XML to relational database
- From one relational database into another

Usually, a specialized software is used to do this transfer. As a result, this software must be rewritten each time, when the structure of the database changes. This is slow and expensive. The complexity of the task increases additionally by transferring information between databases with different structure.

The purpose the doctoral work is to create an universal language processor (a software automaton) which can transform data from relational databases into semi-structured data streams and vice-versa, on the basis of external rules, written in a specialized language. Hence, when the parameters of data transformation are changed, only

the description of these rules will be changed, not the software.

For the purpose, the XSD language shall be extended with additional syntax constructs, which describe the data mapping (relational/XML and vice-versa). Then, the XSD rules shall be translated into productions of a context-free grammar, whereafter these productions will be processed by using the LR parsing method.

The LR parsing method is a table-driven method that uses a bottom-up strategy. The LR method manages its own stack, which has two sections. The first one stores the symbols from the input tape and the results of reductions. The second one stores the states, the push-down automaton (PDA) goes through, during the parsing process. To determine these states, the corresponding sets of LR(0)-Items must be generated. The already determined states are written into a table, which is used to control the work of the language processor, mentioned above.

Milena Angelova is a PhD student in her 2nd year at the Technical University of Sofia – branch Plovdiv. The subject of her PhD thesis is “Advanced methods for big data analysis and modelling”. Milena's research interests are focused on intelligent information access. This entails developing of retrieval techniques that support humans in dealing with massive volumes of data. Her main research interests and activities are in the field of Information Retrieval Systems, Machine Learning and Semantic Web.

A SEMANTIC REASONING ENGINE FOR LIFESTYLE PROFILING IN SUPPORT OF PERSONALIZED COACHING

MILENA ANGELOVA

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

LifeStyle is the state where sports, physical activity, social interaction and psychological well-being are promoted to prevent several chronic diseases including cardiovascular disease, diabetes, hypertension, obesity, depression and osteoporosis. The goal is to entice and support people to change their behaviour.

This research has been performed within a large ARTEMIS EU project WITH-ME². The aim of this research is to develop a sound methodology enabling the profiling of coachees that are being supported by a lifestyle coach. This profiling is to be useful for the professional coaches to personalize their approach for helping people acquire and maintain a healthier lifestyle. For the coachee profiling, the known relevant factors for a healthy lifestyle are gathered, e.g., sleep duration and activity level. Based on coachee input in a questionnaire, to be filled in with support of the coach, a qualitative profile is generated in terms of food and liquids consumption, physical activity, personality and cardiovascular risk. The modelling and reasoning environment is realized with semantic technologies. Semantic Web is a concept that

² This research has been conducted as a part of my MSc thesis project during a three month internship visit at The Collective Center for the Belgian technological industry, Sirris, Brussels. In addition, the obtained results are published in S. Mikolajczak, T. Ruetten, E. Tsiporkova, M. Angelova, V. Boeva, A Semantic Reasoning Engine for Lifestyle Profiling in Support of Personalised Coaching. GLOBAL HEALTH 2015 (Nice, France) 79-83.

enables better machine processing of information on the Web, by structuring documents written for the Web in such a way that they become understandable by machines. This can be used for creating more complex applications (intelligent browsers, more advanced Web agents), global databases with the data from the Web, reuse of information.

Stefan Stoyanov is PhD student in his 3rd year at the Technical university of Sofia - branch Plovdiv, Faculty of Electronics and Automation (FEA). He is currently leading a small team of Software Engineers at the Bulgarian company IT Advanced Ltd. His PhD research interests and activities are in the exploring of methodologies for performance analysis on big data processing. He is performing research dealing with machine and sensor data executing data processing on multi-nodes Hadoop cluster.

METHODOLOGY IN ANALYSIS OF SENSOR DATA PROCESSING
PERFORMANCE. FRAMEWORKS FOR BIG DATA STORAGE AND PROCESSING

STEFAN STOYANOV

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
stefan.stoyanov@it-advanced.com

Working with big data sets has become increasingly common in many areas.

Nowadays, in some areas, data growth has reached the point where a single relational database is not enough. This big data phenomenon has first time appeared in areas like meteorology, sensor data analytics, Internet search, biological research, genomics, finance, and many more. The usual answer to data growth problems has been to scale up and put more storage and processing power in a single machine, but current computer architectures could not keep up with the growth of storage and processing needs. An alternative approach has appeared which is to scale out to more computers and create a distributed infrastructure of hundreds or even thousands of computers.

Sensor Data

A sensor is a device that measures a physical quantity and transforms it into a digital signal. Sensors are always on, capturing data at a low cost, and powering the “Internet of Things.” The process used to extract data from multiple sources, transform it to fit some analytical needs, and load it into a data warehouse for subsequent analysis, a process known as “Extract, Transform & Load” (ETL). The engineers may consider some input data

transformation in order to achieve high performance for their queries against the data or to optimize in terms of storage. Different log aggregation tools may be in help as well.

Apache Hadoop has emerged as the de facto standard for managing Big Data. In very simple terms, Hadoop is a set of algorithms (framework) which allows log aggregation, storing huge amount of data on Distributed File System HDFS and MapReduce processing with Apache Hive as a query processing tool.

In our research we investigate how Apache Hadoop performs with the same volume of the data set, supplied to Hadoop as input files in different data formats: XML, JSON and CSV (which are some of the most common formats for semi-structured data transmission over Internet), applying or not for intermediate storage data compression and data transformation to data serialization binary formats like Apache Avro and RCFile (Record Columnar File) or ORC (Optimized Row Columnar).

Georgi Pazhev is PhD student in his 1st year at the Technical university of Sofia - branch Plovdiv, Faculty of Electronics and Automation (FEA). He is currently working in Seeburger AG as Java developer. His PhD research interests and activities are in design and implementation of Smart devices used in building and in particular smart homes.

SMART HOME - METHODS AND UTILITIES

GEORGI PAJEV

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

The article presents the methods and devices for building systems for automation type smart home, including methods, technologies and concepts for their design. There are presented the already established in this area platforms and architectures offered by the major producers as Amazon, Bosch, Intel, Samsung, open source hardware and the smart home developed in the Technical University, based on two gateways using open source hardware platforms.

The smart home is a system for creation of distributed automation which consist different types of networks. Basically in the case of connection of heterogeneous networks are used gateways, which can provide the transparent work of the application. One possible and low-cost realization of the smart home is the use of open source hardware platforms such as Olinuxino, Arduino and his varieties (Pinguino, Olimexino 328 etc.). The University developed smart home is based on wire-less (Bluetooth, IRDA, Wi-Fi, ZigBee) sensor networks and consists of two gateways – control and executive. The control gateway is created by the hardware platform Olinuxino A13 and the executive – by the Arduino compatible hardware platform Olimexino 328. The communication between these two types of gateways is created by the CNDEP. The executive gateways consist sensors and actuators, by which the executive gateways gets, process or transmit data and sends it as reply to the control gateway. The control gateway consists rules, defined by the user on the base of which different actions must be executed by the executive gateways. Every one of them should be configured by the control gateway according to the sensors and actuators, which are connected to the executive

gateway. The control gateway sends the configurations for each of the executive gateways, which describes the actions which must be executed.

Another decision for realization, which requires hardware with more resources, is the voice assistant Alexa, created by Amazon. This assistant recognizes voice commands by the user and gives an opportunity for creation of new voice commands (skills) using web - service technology REST and the programming languages Java, Python and JavaScript. Intel offers its solution to develop a smart home by their technology referred as Internet Of Things (IoT). This technology is a concept for computer network of physical objects (devices, vehicles, buildings and other things), having embedded electronic devices to interact with each other or with the outside. Many producers of smart devices such as Samsung use this technology.

Angel Chekichev graduated from the Technical college in Smolyan, obtaining an associate degree with two majors: “Computer technics and technology” and “Computer and communication systems”. In 2012, he obtained a Master's engineering degree from the Technical university of Varna in the speciality of “Computer technics and technology”, majoring in “Telecommunications networks and systems”.

Since 2014, he has been enrolled in the extra-mural PhD programme of the Technical university of Sofia in the Faculty of electronics and automatics of the Plovdiv branch. His PhD research interests and pursuits are concerned with the extraction of characteristics of the development of processes in samples captured digitally based on the colouring characteristics of the images. His other interests include: digital processing of signals; signals and systems; computer graphics and modeling.

He is currently a member of the staff and teaches at the Technical college in Smolyan which is a branch of Plovdiv University „Paisii Hilendarski“.

FEATURE EXTRACTION CONCERNING THE DEVELOPMENT OF PROCESSES IN OBJECTS, BASED ON COLOUR CHARACTERISTICS IN IMAGES

ANGEL CHEKICHEV

Computer Systems and Technologies Department
Technical University of Sofia-branch Plovdiv
Tsanko Dyustabanov 25, 4000 Plovdiv, Bulgaria
ma8721@abv.bg

The dissertation's subject is related to medical diagnostics and is based on the examination of digital images.

Contemporary methods for compiling and processing digital images necessitate the exploration of new means and approaches, aimed at a thorough analysis in order to obtain higher quality of the digital imaging which will lead to the identification of the most appropriate methods for their visualisation.

The dissertation deals with the changes observed in samples of biological material captured digitally. In certain cases these changes are manifested in changes of the geometrical

shape and size, and in other cases – in changes of the internal structure and colouring. A comparison with a standard sample is performed in order to establish the deviations; however, to eliminate possible bias, automation of the analytical process is required. A lot of the issues in this field are identical to the issues related to capturing and analysis of random images, namely:

- different conditions when capturing similar objects;
- establishing suitable points of localisation when determining the picture and image sizes and creating a topography;
- extracting characteristics and creating attribute space;
- formulating comprehensive criteria for the identification of an object.

The existing solutions to these problems in areas outside the scope of medical diagnostics prove to be an appropriate foundation for the comparison and testing of methods which would be suitable for use within the context of medical diagnostics as well. The combination of their use with methods, which do not create images and are based on another type of signals, would present the opportunity to obtain better results when standardising samples. A key element in this regard is the creation of an attribute space, which the present dissertation addresses, dealing with the extraction of characteristics of the biological processes in the samples captured based on the colouring characteristics of the images.

Evgeni Yordanov is a PhD student in his 1st year in TU Sofia, branch Plovdiv, Faculty of Electronics and Automation (FEA). He has a Bachelor's degree in Telecommunications Engineering and a Master's degree in Computer Business Informatics Engineering from TU Sofia. He then continued with a post-graduate certificate in the field of Content Strategy from NorthWestern University - Medill School of Journalism, Media and Integrated Marketing Communications in the United States of America. His research interests are based in the field of work that he is in Digital Marketing, thus the topics of his doctorate are in the field of marketing automation, search engines and in specifics - the optimization and advancements of specific results within the search engine result pages.

SEARCH ENGINE OPTIMISATION: TOOLS & METHODS

EVGENI JORDANOV

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

The “Search Engine Optimization: Methods and Tools” presentation is regarding the topic of Search Engines – their history, way of operations as well as methods and tools used in this day of age in order to achieve a well optimized result in the search engine result pages (SERPs).

A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages (SERPs). The information may be a mix of web pages, images, and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler.

Web search engines get their information by web crawling from site to site. The "spider" checks for the standard filename robots.txt, addressed to it, before sending certain information back to be indexed depending on many factors, such as the titles, page content,

JavaScript, Cascading Style Sheets (CSS), headings, as evidenced by the standard HTML markup of the informational content, or its metadata in HTML meta tags.

Indexing means associating words and other definable tokens found on web pages to their domain names and HTML-based fields. The associations are made in a public database, made available for web search queries. A query from a user can be a single word. The index helps find information relating to the query as quickly as possible.

Some of the techniques for indexing, and caching are trade secrets, whereas web crawling is a straightforward process of visiting all sites on a systematic basis.

Search engine optimization (SEO) is a methodology of strategies, techniques and tactics used to increase the amount of visitors to a website by obtaining a high-ranking placement in the search results page of a search engine (SERP) -- including Google, Bing, Yahoo and other search engines.

It is common practice for Internet search users to not click through pages and pages of search results, so where a site ranks in a search results page is essential for directing more traffic toward the site. The higher a website naturally ranks in organic results of a search, the greater the chance that that site will be visited by a user.

SEO helps to ensure that a site is accessible to a search engine and improves the chances that the site will be found by the search engine. SEO is typically a set of "white hat" best practices that webmasters and Web content producers follow to help them achieve a better ranking in search engine results.

Some of the most useful are provided by the search engines themselves. Search engines want webmasters to create sites and content in accessible ways, so they provide a variety of tools, analytics and guidance. These free resources provide data points and unique opportunities for exchanging information with the engines.

Irina Kakanakova received her MSc in Computer Systems and Technologies from Technical University of Sofia - branch Plovdiv. She is a PhD candidate at the Technical University of Sofia. Her PhD research interests and activities are in coping with ill-structured problems. The research includes exploring new, more effective algorithms for dealing with unstructured data and schemes for combining heterogeneous data sources.

DEEP LEARNING ALGORITHM THROUGH ADAPTIVE CONTRASTIVE DIVERGENCE

IRINA KAKANAKOVA

Technical University of Sofia
irinanicolova@gmail.com

There is an increasing interest about using unstructured data sources like images, text, sound and video nowadays. Deep learning algorithms dramatically improved the effectiveness of dealing with such sources. These algorithms extract high-level, complex abstractions as data representations through a hierarchical learning process. Complex abstractions are learnt at a given level based on relatively simpler abstractions formulated in the preceding level in the hierarchy, and the learning is executed in a greedy manner.

There are two key “ingredients” that provided the success of deep learning algorithms: hierarchical structure building blocks and learning procedure used. The building blocks of the hierarchies used in deep learning process are structures called Restricted Boltzmann Machines (RBM). RBM is a type of stochastic neural network and Markov Random Field, with the restriction that their neurons must form a bipartite graph: a pair of nodes from each of the two groups of units may have a symmetric connection between them; and there are no connections between nodes within a group. The learning procedure is an algorithm called Contrastive Divergence. In contrast with Maximum Likelihood algorithm, which deals with whole input data space to compute the gradient of model parameters, Contrastive Divergence uses a reconstruction of the input data space. The reconstruction used is obtained by hitting input data with Monte Carlo Markov Chain (MCMC) according to current model parameters one or more times. The latter computational procedure is much more effective, and the results in most cases are satisfactory.

It is proved that the longer MCMC is, the better results are obtained, but efficiency decreases. Usually the learning procedure is executed several times with different MCMC

step number to find the most suitable for the problem being solved. More sustainable method to obtain the optimal ratio computational cost – model quality by a Contrastive Divergence learning procedure with variable number of steps is proposed.

Main problem for developing such algorithm is the evaluation of the current state of the learning process, because the Contrastive Divergence learning procedure doesn't optimize cost function directly. Metric based on gradient (size, dispersity) and the depth of the learning process (except cost) is used.

Experiments on different datasets were held. They produced different learning curves. In 80% of the experiments was indicated improvement in time needed to achieve specified accuracy.