Resumes

from PhD Scientific Seminar

Department Computer Systems and Technologies

Faculty Eletronics and Automation

Technical University Sofia branch Plovdiv,

04 December 2017

Editor: Nikolay Kakanakov

List of Authors

List of Topics

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 Cloud Network Architect and leads 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.

VSAN STRETCHED CLUSTER. METHODS FOR TRAFFIC OPTIMIZATION BETWEEN SITES.

DIMITAR GROZEV

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

Cloud? Yes we have cloud in our company….

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. Since it was born cloud design and implementation has evolved a lot. In this research, we will examine one of the latest cloud design based on HCI (Hyper Converged Infrastructure) and VSAN (Virtual Storage Area Network) implemented in a way that will give us site redundancy, ability to avoid disaster recovery or restore extremely fast from disaster. This is VSAN stretched cluster - virtual environment, where we can run mission critical applications at lower cost. Do we have site redundancy and disaster recovery site? This has been one of the most inconvenient question that I have ever heard. Maybe you are thinking why? Well, few words only, storage, storage replication, live storage replication…

VSAN stretched cluster has strict network requirements in order to work well. Problem that can be solved with a lot of bandwidth or with good optimization.

We will make a deep dive into site to site traffic, classify traffic types, monitor them using different technologies and suggest methods for optimization and better bandwidth utilization between sites (between hosts in vSAN stretched cluster). Stay with us to understand more!

Lilyana Boneva is a PhD student in her last year of PhD studies 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.

CLUSTERING TECHNIQUES FOR UPDATING DRUG SAFETY TOPICS

LILYANA BONEVA

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

Background: Pharmacovigilance is the practice of collection, analysis and prevention of adverse drug reactions (ADRs) induced by drugs. The detection of adverse drug reactions is performed using statistical methods and clusters of ADR terms from the MedDRA (Medical Dictionary for Drug Regulatory Activities) terminology. Standardized MedDRA Queries (SMQs) are groupings of terms assisting retrieval and evaluation of MedDRA-coded ADR reports worldwide. Each SMQ covers an important drug safety topic. MedDRA terminology is updated twice annually with new terms. Updating SMQs and creating new SMQs for topics not yet covered is a manual process performed by experts. Our research aims at automating the process of clustering new terms for drug safety topics.

Methods: The method explores the semantic distance between terms and applies clustering techniques. We investigate two approaches – One-pass Clustering Algorithm and Correlation Bi-Clustering Algorithm. In one-pass clustering individual terms are merged into existing SMQs based on semantic relatedness between terms. Alternative approach groups initially new terms in clusters and subsequently merges these clusters to SMQs. The initial set of clusters (SMQs) and clusters obtained on step 1 form a bipartite graph that is subject of bipartite correlation clustering (BCC) problem. The

bipartite graph is the input and output is a set of disjoint clusters covering the graph nodes. The objective of BCC is to generate a set of vertex-disjoint bi-cliques (clusters) which minimizes the symmetric difference between the input and the output set. Nodes from the set of SMQs are chosen randomly as ‘pivots’ or ‘centers’ and clusters are generated from their neighbor sets.

Existing SMQs are basis for evaluation of the accuracy of the algorithms and the fitness of the clustering output obtained to be used as a preprocessing phase assisting experts’ manual work.

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 working as software developer at IBS Bulgaria Ltd. Dimitar's research interests and activities are focused on image processing, computer vision and parallel computing. 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".

APPLICATION OF PARALLEL COMPUTING FOR SOLAR CORONA IMAGES PROCESSING

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 a 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.

Typical objectives of the methods of solar image processing are:

- image pre-processing to eliminate noise and, if necessary, restore images
- filtering images to highlight the structures and processes that take place in the solar corona
- automatic detection of events in the solar corona and evaluation of their characteristics

Progress in technologies for solar corona registration has led to improved spatial and temporal resolution of the data. Respectively increasing the amount of information that must be processed, both by increasing the resolution of the images and by reducing the intervals between the recording of two consecutive images. One of the biggest impulses for creating a systems for automatic detection of solar processes is SDO (Solar Dynamics Observatory). By providing about 1.5 TB of solar data each day SDO brings the scale of solar image processing to a next level of development. Due to the significant advances in the development of computer storage media and their

availability, as well as the possibilities for their processing, there has been an increase in the attention to automated and modular means of processing the sun images. One of the key approaches to implementing solar imaging software is building modular systems which will provide tools for event detection, classification and tracking. To develop an effective system which can automatically track and measure solar events, a hybrid approach should be used that combines image processing, optimization and algorithms.

Parallel implementation of image processing algorithms involves distribution of the operations on the images between several available computational resources. Below the resources used for image processing are multi-core processors on a computer system, a group of nodes (computers), multi-threaded computing devices in the GPUs. Regardless of the computing devices used, the main goal of assigning tasks is to maximize the use of available computing resources (processor, memory) as well as their even load.

From the point of view of performance in performing parallel calculations, the best performance is achieved by using environments that provide support for vector-based information processing devices as well as multi-threaded devices. These include technologies like CUDA, OpenCL and OpenACC. If we look at the possibilities of distribute calculations between several computational resources, we need to take into account MPI environments that allow for the distribution of the calculations between several nodes. By using parallel programming interfaces such as MPI in combination with technology such as OpenCL for massive parallelism, can be achieved very good results in processing a large number of images.

Stefan Stoyanov is PhD student 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

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).

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.

A different solution has appeared which is to scale out to more computers and create a distributed infrastructure of hundreds or even thousands of computers.

However the engineers may consider some alternative approaches to distribute the

processing work over time and to get rid of big part of the sensor logs in order to reduce the need of storage resource. Input data transformation can be considered in order to achieve high performance for queries against the data or to optimize in terms of storage.

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 on aggregates like COUNT, MAX, SUM or AVG on the same volumes of the data set compared to alternative approach involving centralized log aggregation intermediate data processing level.

Sensor data is supplied as input files in some of the most common formats for semi-structured data transmission over Internet - XML, JSON and CSV.

Milena Angelova is currently a Ph.D student in her 3rd year 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 "Advanced methods for big data analysis and modeling". Milena also is an assistant professor at the Department of Computer Systems and Technologies and a Full-Stack developer at Delta Source. Her research interests are focused on intelligent information access. This entails developing of retrieval techniques that support humans in dealing with massive volumes of data. She addresses her Ph. D. work on proposing of efficient clustering techniques for incrementally modeling a big volume of incoming data that making changes to the stored data frequently. Her primary research interests and activities are in the field of Information Retrieval Systems, Machine Learning, Semantic Web and Recommender Systems.

EVOLUTIONARY CLUSTERING TECHNIQUES

MILENA ANGELOVA

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

Nowadays, most of the organizations are constantly dealing with data that comes from their employees, customers, potential candidates or other external sources. They are currently looking for ways to use data to address their business problems. However, the amount of data being capture today is ever-increasing and extends beyond the storage and analyzing the capacity of traditional applications. In many practical applications such as expertise retrieval systems, the information available in the system database is periodically updated by collecting new data. By the more than 2 billion people and millions of enterprises living their lives and doing their work online, and by the millions of sensors and communicating devices sending and receiving data over the Internet. One of the challenges relates to an observation that in the real world data tends to change over the time. It is becoming impractical to re-cluster this large amount of available data every time. Practically, this will lead to seeking mechanisms for

continuous diagnostic of performance, and to able to adapt to changes periodically. Thus, we are focused on developing evolutionary clustering techniques that deal with a big volume of incoming data.

Incremental clustering methods process one element at a time. Typically, they only store a small number of elements, as a constant number. Due to increasing volumes of data, it is no longer possible to keep all the data in memory at the same time. For example, sensors with small memory are not able to keep a big amount of data so new techniques for data stream clustering should be proposed. Another challenge for incremental clustering methods is a scanning and an adapting the changes to stored data. Most of the algorithms iterate over the data multiple times while they integrate the streaming data. Namely, evolutionary clustering techniques discussed herein can evaluate the clusters on a new incoming data over a defined time and update the current data with new one more efficiently. In the current work, we propose and study three different evolutionary clustering algorithms and evaluate them by using data from PubMed repository. PubMed is one of the largest biomedical repositories which accessing the Medline database of references and abstracts on life science and biomedical topics primarily. Our three evolutionary clustering techniques can be split into two categories: a Partitioning-based and two graph-based clustering algorithms (PivotBiCluster and Merge-Split PivotBiCluster). To integrate two clusters into a single clustering solution, the Partitioning-Based algorithm uses a merge schema, where the cluster centers of the available clusters are considered. Subsequently, the cluster centers are divided into groups by applying some clustering algorithm, and the clusters whose centers belong to the same group are merged to obtain the final clustering. However, the Partitioning-based clustering algorithms needs prior knowledge about the optimal number of clusters. This operation will cost too much time for the algorithm to iterate over the data. However, the other two graph-based clustering algorithms do not require such information to provide a good clustering solution. PivotBiCluster and Merge-Split PivotBiCluster are obtained by merging clusters from both sides of the graph, i.e., some of the existing clusters will be updated by some of the computed new ones. However, existing clusters cannot be split by PivotBiCluster even the corresponding correlations with clusters from the newly extracted data elements reveal that these clusters are not homogeneous. This was motivation task for us to propose a Merge-Split algorithm. Merge-Split is able to analyze the correlations between corresponding clusters and to merge them in a single cluster. Otherwise, if the corresponding clusters do not have a correlation, the algorithm will split the cluster elements among several new clusters.

For future work, we aim to evaluate these tree algorithms in different application domains and online data sources.

Donka Nesheva is a 3r d 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.

CLOUD-BASED DECISION SUPPORT SYSTEM FOR DIABETES MANAGEMENT

DONKA NESHEVA

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

Noncommunicable chronic diseases have become the leading causes of mortality and disease burden worldwide. One of the widely spread problems are hypertension, heart disease, dyslipidemia, and diabetes. When they are not well controlled chronic diseases can cause other chronic disease or expensive complications. For example diabetes is the major cause of kidney failure, blindness, and non-traumatic leg amputations and a leading cause of stroke and heart disease. In 2015, estimated 30.3 million people of all ages—or 9.4% of the U.S. population had diabetes. In Bulgaria the number is around half of a million.

Although good metabolic control can reduce the risk of diabetes complications and associated costs, many patients have glucose levels above goal, and hypoglycemia often complicates treatment. The goals for management of diabetes are well defined, effective therapies are widely available, and practice guidelines have been disseminated extensively. Despite such advances, health care providers often do not initiate or intensify therapy appropriately.

Limitations in managing chronic diseases are often due to clinical inertia—failure of health care providers to initiate or intensify therapy when indicated. Clinical inertia is due to at least three problems: overestimation of care provided; use of “soft” reasons to avoid intensification of therapy; and lack of education, training, and practice organization aimed at achieving therapeutic goals.

In order to overcome clinical inertia in 2015, prof. Lawrence S. Phillips, MD and

team has developed a new approach of diabetes management. Based on their work we developed a cloud-based decision support system designed to incorporate 4 strategies which have been shown to be effective in improving glycemic control: personalized management based on individual patients' clinical markers and medications; facilitated transmission of patient data to clinicians; patient education and engagement; implementation of diabetes management guidelines and algorithms.

Patients are engaged to record daily glucose values and medications through mobile applications. Data is captured and stored into a private cloud warehouse, where the system runs algorithms to evaluate metabolic control and recommend changes in management as needed; the algorithms are designed to mimic the decision making of an experienced en-docrinologist. Healthcare team reviews recommended changes and send them back to the patient's mobile application.

The results of conducted trials shows significantly lower levels of A1C values than levels at baseline at +3, +6, +9, and +12 months. Despite the improvement in A1c levels, recorded hypoglycemia occurrences were reduced to less than 1 episode per month.

Nikolay Nikolov is PhD student in his second 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 System architect and software developer at the Bulgarian company Dieselor Ltd. His PhD research interests and activities are in the Language processors for formal languages, Database systems 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.bg

The purpose of 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.

Since there is a huge variety of possible XSD definitions, there may be cases in which the productions of corresponding contex-free grammar will cause ambiquity during the LR analysis.

Up to now some sample grammars, which cause ambiguity in the SLR, LALR and full LR methods for parsing tables generation has been studied. For each one of these cases the common form of the grammar, which causes the particular type of ambiguity has been defined. Some possible solutions of the outcoming "reduce-reduce" and "shift-

reduce" conflicts have also been analyzed and evaluated.

Evgeni Yordanov is a PhD student in his 2nd 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.Now he's involved heavily with the startup ecosystem, and lecturing on the topics of digital marketing and automation in several academies and universities in Bulgaria. His research interests are based in his work field - digital marketing. So it is only logical for the topic of his doctorate to be in marketing automation, search engines and in specifics - the optimization and advancements of specific website results within the search engine result pages.

SEARCH ENGINE OPTIMIZATION: TOOLS & AUTOMATION

EVGENI YORDANOV

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 Automation" 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 number 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. Some techniques include:

- HTML validation
- Text copy enhancement
- Server speed optimization
- Secured data transfer (SSL)

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.

Boian Katzarsky is PhD student in his first year at the Technical university of Sofia - branch Plovdiv, Faculty of Electronics and Automation (FEA). Boian holds a MSc degree in Computer Science from the same university. He is currently working as System Architect at the Bulgarian company NitroBite Ltd. His PhD research interests and activities are in: distributed systems, state synchronization across nodes in distributed systems, operational transformation (OT), conflict-free replication data types (CRDT) and finite-state machines.

DIVERGING TIMELINES OF STATE IN NODES OF DISTRIBUTED SYSTEMS. OPTIMISTIC CHANGES TO STATE THAT MERGE.

BOIAN KATZARSKY

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

The purpose of the doctoral work is to create a framework of data types and algorithms that facilitates synchronization of state in loosely-coupled (with intermittent connection) nodes in distributed systems. This framework will be built upon to implement distributed applications that use these data-types and have data consistency across the whole system by default. In loosely-coupled distributed system the particular data-states of the nodes will be partially different but will converge to a single state given that they are connected to the network and can propagate the changes from their timelines.

The shared state of a distributed application is all the common data, shared between the nodes. It does not refer to the local variables in functions or program stack.

Each node is not always connected to the network all the time (e.g. a mobile device that temporarily lacks connectivity). During that offline time, it makes changes to the data-state. Hence the state diverges from the shared state and this forms a private timeline. Changes made while offline are “optimistic” because they are made without the consent of the other nodes – i.e. without a “transaction” approved by all the nodes. This leads to speedy offline behavior in the node itself but requires a special policy that ensures the changes are merge-able to the common (shared) state.

Upon re-connection the divergent state must be merged to the common state. This

can be performed by (1) merging the latest snapshots of the local and remote data or (2) merging the local and remote queues of operations performed over the state. This leads to two general approaches for converging the state: (1) data merge and (2) operations merge.

For merging of data (1) special data-types are used known as “conflict free replication data types” or CRDTs. For merging of operations (2) an operational-transformation (OT) is used to remedy the fact that the operation will be performed over changed context.

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.

IMAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS

TEODORA HRISTEVA

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

Convolutional Neural Networks emerged from the study of the brain's visual cortex, and they have been used in image recognition since the 1980s. In the last few years, thanks to the increase in computational power, the amount of available training data, and the algorithms for training deep nets, CNNs have managed to achieve super-human performance on some complex visual tasks. They power image search services, self-driving cars, automatic video classification systems and more. But CNNs are not restricted to visual perception: they are also successful at other tasks, such as voice recognition or Natural Language Processing (NLP).

One of the defining characteristics of a good image recognition algorithm are its ability to detect salient regions, i.e. regions which contain the most information. These are most likely focus corners, textures and centers of unique shapes to recognition objects. This is what image recognition is all about, find the most informative features in a given image to quickly recognize and localize the contents of an image.

The decision making processes in recognition systems are based on a weighted sum of inputs and activations. The weights determine how influential a particular feature is, the activation defines a decision boundary, which involves adding a bias to the weighted-sum and passing the result through a squashing function. A useful activation function is the rectifier, it merely just cuts off the negative region. This is what is what

is known as an artificial neuron.

The basis of deep learning systems is such that the decision making processes by artificial neurons can be made more complex and more abstract by arranging them in layers whereby each layer feeds from a layer below it and sends its output result to the layer above it. One can design a learning algorithm based on gradient descent to adjust the weights in such complex networks, a process that can be sped up using a propagation of errors from the output towards the input called backpropagation, which is just chain-rule from calculus for computing weight derivatives in the neural networks.

A mathematical definition of saliency can be found in the definition of a corner, a salient region is a region that is sharply different compared to its immediate neighborhood. Meaning that it is easy to locate a salient region consistently across varying degrees of transformations. Deep learning features also end-up detecting salient features from training samples, salient regions are important for both object recognition and localization.

Nowadays especially in fields such as deep learning, image recognition is performed by learning algorithms. The trained feature detectors are the convolutional layers that can adapt automatically to training data. In this way is done the collecting of a lot of training examples and the choosing of hyper parameters such as how deep the network must be, the learning rate, the activation functions etc. This results in a very well performing recognition system as the features are adapted to the given problem at hand rather than handcrafted features that maybe good for one or few problems and then break for others.