



Smart Services Using Voice and Images

Alexander I. Iliev^{1,2}(✉) and Peter L. Stanchev^{2,3}

¹ UC Berkeley Global, Berkeley, CA, USA
ailiev@berkeley.edu

² Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

³ Kettering University, Flint, USA
pstanche@kettering.edu

Abstract. In this chapter, we will emphasize on some of the most prominent advances in smart technologies that formulate the smart city ecosystem. Furthermore, we will be highlighting the automation of numerous developments based on the extraction and analysis of digital media, using speech and images. At present, there is a multitude of practical systems used for personalization and recommendation of different media. On the other hand, there are assorted types of services in different areas that are directly benefiting from these advancements. Most of them were created with human-machine interaction methodology in mind, where people had to interact with the machines in various ways. In the past this type of interaction has been completed through the use of conventional interfaces such as a mouse and a keyboard, where the user had to type a response manually, which was in turn recorded by the machine for subsequent analysis. Therefore, in order to simplify these types of interactions and lead to improvement of services, new methodologies must be studied, discovered and developed so as to improve services such as recommendation and personalization services.

Keywords: Smart systems · Emotion recognition · Voice recognition · Image recognition

1 Introduction

Smart services provide addition value for both service providers and consumers. The continuous growth of data according to different sources [1, 2] leads to an increasing necessity of novel and adequate technologies that can structure and analyze vast amounts of new data every day. This inevitably leads to the need of creating smart machines that have a profound impact on our daily lives. The increase of computational power as well as minimization of electronics in every field is also a factor in the complex equation of smart technology development. Simply put, there are too many factors that can contribute to the advance of smart devices, smart cars, smart homes and even smart cities in a global Internet of Things system. This is why the topic of smart system development in multiple areas is the focus of this chapter.

Despite all of this, technology is not the main driver behind the need to develop more advanced concepts for smart cities. Technology merely serves the need we have.

Recent studies [3] by the United Nations Population Fund suggest that by the year 2030 approximately 66% of the world's population (or about 5 billion people) will live in urban areas. According to this report, in 2014 this number was 54% (about 3.3 billion people). By the growth of the world's population and the increase of smart technologies and devices worldwide, it is inevitable that more and more smart technologies should emerge in years to come.

ACM Transactions on Internet of Things (ACM TIOT) is a new ACM journal that will publish novel research contributions and experience reports in several research domains whose synergy and interrelations enable the IoT vision. TIOT focuses on system designs, end-to-end architectures, and enabling technologies, and on publishing results and insights corroborated by a strong experimental component.

Voice is easy to capture, retain and process. These characteristics of voice make it an extremely attractive source for gathering intelligence. Thus, the reason why we will spend some time in this chapter to discuss smart technologies based on speech signals. Nowadays it is an absolute necessity to study and use speech and images as a source of automation. It is an advanced way to think about future developing technologies. The amalgamation between Digital signal processing (DSP) and Artificial Intelligence is the expected way of solving sophisticated smart technology problems and is already happening. Through the usage of DSP, we can contrive complex feature vectors as inputs to machine learning algorithms in order to complete specific smart tasks. The informational capability of speech signals remains undisputed.

2 Smart Systems Through Voice

In today's world we are constantly bombarded by an ever-growing information flow. It is all around us in everything we do. The steady increase of options in every aspect of our lives requires an exponential growth of information gathering and processing. Making simple choices when selecting food in our local shops, selecting the right doctor based on specific criteria or choosing an entertainment tonight are some of the many choices we are up against on a daily basis. These, among many other choices must be made by all of us in order to fit the modern paradigm we live in. The selection process must therefore be aided by the implementation of smart algorithms for nearly every aspect of our lives. We see the growth of smart systems implemented in different industries such as the automobile industry, banking, healthcare, insurance, security, shopping, and entertainment to name a few. For many of these algorithms data is collected from human activity that usually is the basis for deducing specific behavior using deep learning methods. In many cases, these behavioral patterns are typical for different people and in these cases personalization of services becomes possible. The data sources used for this kind of activity is usually gathered silently in the background. Amazon is constantly gathering information on the latest products we search for or the ones we already bought. Google and Facebook built their enormous empires around personalizing ads based on search records and habits we have when interacting with their systems. These services are based on textual information we typed when searching for people, services or products. Based on this kind of search there is a way of creating new technology and services to aid us in the sea of choices we are up against.

Depending on the task, the designer of such system can collect information from different layers of speech such as:

1. *Who is talking: that is a trivial speaker recognition task. This can also be specific to gender recognition only (men, woman or a child) or more specific as to what individual is talking from a number of speakers in the room;*
2. *What a speaker is saying: this is known as speech recognition and it aims to collect a specific verbal message from each speaker. In many cases a smart system can be tuned to gather specific words that a speaker is using to deduce some logic from the speech. In this layer of information, the systems can even be tuned to map specific verbal messages and words to specific people, if there are several speakers in the particular environment, thus combining the first two points above;*
3. *How was the voice message conveyed: this captures yet another extremely important layer of information from speakers? By capturing the way each speaker is talking we can gather important information about each person's behavior and mood. This type of information is especially valuable for smart system development in areas such as security, customer service interaction, banking, etc.*

Speech recognition is a very developed and extremely interesting field when it comes to implementing the result of it in today's Smart Systems. It is often that we see researchers using the Mel-frequency cepstral coefficients (MFCCs) to extract specific attributes for speech recognition, and then using methods like Dynamic Time Warping (DTW) in combination to capture some interesting patterns and achieve voice command/speech recognition as a result [4]. In their research, the authors mention some of the different techniques to be incorporated in the field of gender recognition. Pre-emphasis filtering is applied to the speech signal and as a result the magnitude of the higher frequencies was increased in order to improve the overall SNR Ratio. This signal was then passed through a so-called framing technique wherein it is divided into N samples with M distance between them ($M < N$). Then after Hamming window was applied, a Discrete Fourier Transform (DFT) was performed in order to convert the signal from time to frequency domain and access to magnitude frequency spectrum was obtained. The MFCC recognizes the linguistic and relevant content and discard every other signal. Access to the Mel spectrum was given through the use of triangular shaped filters or Mel Filter Banks. The signal was then transformed back into time domain by using Discrete Cosine Transform (DCT) and then at the final stage data energy and spectrum were collected. Feature matching, DTW is used to compare 2-time series to measure their similarity. The input signal is compared to the variables to determine the accurate output.

Another interesting and very practical example of a smart system that determines *what the speaker is saying* is described in [5]. There, a smart system was developed, which could help people with disabilities, senior citizens or be installed in every household to simplify our lives by aiding us control our home appliances with voice commands. The project was implemented at a fairly low cost and was claimed to be energy efficient. It utilizes a Texas Instruments TMS320 DSP processor along with a microphone, and when specific conditions are met, it sends a signal via Universal Asynchronous Receiver-Transmitter (UART) hardware to the XBee Transmitter (Tx), which is an embedded

solution that provides wireless end-point connectivity to different devices. Voice Recognition is done via Zero Crossing with Peak Amplitude (ZCPA) – a method measuring the number of times a given frequency is crossing the zero line. This helps distinguishing between various commands. The XBee Tx transmits the signals via the UART module. Any device that has the UART interface can interact directly with the XBee RF module and will receive the signal through the XBee Receiver (Rx). The XBee Rx has a microcontroller attached to it receives the transmitted control characters and compares them against a set of specific ones. If there is a match, the microcontroller will switch the corresponding relay on or off, thus controlling the state of the appliance.

In another example [6], the gender of a specific speaker is determined based on a speech sample using a one-dimensional Convolutional Neural Network (CNN). CNNs are commonly used for image recognition, but in this case, it was used for gender classification. Voice is firstly passed through a pre-emphasis filter, as it is susceptible to noise interference. Voice samples sometimes can contain a lot of redundant data and thus grow to be too long. Feature extraction is a methodology we use to collect and then reduce the number of samples into a more manageable set of features. When dealing with speech we often used a combination of different MFCCs and even Short-Term Fourier Transform (STFT) techniques in order to extract various attributes. This helps in producing a training set wherein we can use feature selection techniques to extract variables that resemble the target variable. Evolutionary Search algorithms like PSO and Wolf Search are employed as well. Then using Classification Techniques, labeling the samples is achieved by using test samples to identify if the speaker is a male or a female, thus verify the validity of the results.

In [7], an attempt to show the importance of emotion recognition in speech is displayed. There are many kinds of emotions, so the challenge is to label and train for each and every emotion. Depending on implementation different set of emotions can be considered and each emotion represents a separate label. These labels comprise of three general types based on the way speech was conducted or more specifically: acted, elicited, and neutral. Any speech has to undergo the usual digital signal processing where it is: converted to digital signal, quant sized, pre-processed, and windowed, then feature extraction followed by selection techniques are used. There are different types of features to be selected for this purpose:

1. *Prosodic Features*: these are perceived by humans such as intonations and rhythms. An example may be when we put an emphasis on certain words while talking;
2. *Spectral Features*: MFCCs converted to frequency domain and then energy calculated using Mel Filter Banks, Linear Prediction Cepstral Coefficients (LPCC) that contain vocal tract feature specifics; Gammatone Frequency Cepstral Coefficients (GFCC) these are obtained by a similar technique as in the MFCC extraction, but using Gammatone filter-bank instead of Mel-filter bank; Log-Frequency Power Coefficients (LFPC) mimic logarithmic filtering characteristics of the human auditory system;
3. *Voice Quality Features*: These are unintentional features that can be used to differentiate emotions in speech. Properties such as jitter, shimmer, Harmonics to Noise Ratio (HNR) can be paramount in the quest of constructing the perfect feature vectors and determine the emotional state of the speaker. A correlation between voice quality and emotion was previously suggested by [8];

4. *Teager Energy Operator (TEO) based features* [9]: these are used to detect stress features in our speech. In essence, speech is produced by non-linear vortex-airflow interference in the human vocal system. While producing a sound in a stressful situation the muscle tension of the speaker is affected, which results in an alteration of the airflow.

3 Smart Services in the Medical Field Through Voice

Smart Systems based on voice have been implemented in a vast number of areas and one of the most important one is the medical field. An interesting product that is monitoring patients by using Digital Signal Processing (DSP) and Deep Learning (DL) is described in [10]. This paper proposed a method that is improving reaction time of medical staff when dealing with disabled people when using technological advancements related to IoT. For that purpose, a Raspberry Pi (RPi) based system has been developed in order to serve people with disabilities. The system is activated via voice and is interpreting them through the usage of a Support Vector Machine (SVM) then generating signals for medical personal as a result. This research was applied in a real-world medical scenario. RPi boards gained popularity for many reasons related to how compact they are, their low-price range, as well as the powerful capabilities that they have. Their size makes them suitable for many mobile and IoT applications. Some of the specifications for the 4th generation RPi in 2020 are: 1, 2 or 4 GB memory, with a Broadcom Quad core CPU - Cortex-A72 (ARM v8) 64-bit SoC@1.5 GHz processing power.

Python scripts have been written to enable the voice recording of the signals. The main communication interface was implemented using an LCD touchscreen [11]. These small devices are built with minimization in mind, so they use micro SD cards as hard drives, from which a mini version of Linux system called Raspbian is running. These features make this portable device a very attractive, nice and easy-to-use plug-and-play device. In order to extract and collect the features, the Discreet Cosine Transform (DCT) was used [12]. SVM was employed for classification [13] while feeding voice waveform parameters into the input of the system. The signal was firstly sent to the DCT section, so that frequency features were extracted and then provided to the second stage for classification into the SVM. The Montreal Affective Voices (MAV) dataset was utilized in this product [14], since it comprised of an assorted set of voice samples covering wide range of emotions for training, validation and testing. The following nine emotions were of interest: happiness, sadness, anger, disgust, surprise, fear, pain, pleasure and neutral. All of the samples were exemplified in 90 non-verbal voice samples.

Naturally, when it comes to selecting the features a good performance should be the first thing in mind. On the other hand, not having the minimum set of features, as well as when they are too many, it may hurt the recognition rate and consequently it may result in poor performance both as a computational load as well as bad results. One way to counter that is to run a number of tests in the training stage in order to test performance with different feature combination as well as varying a number of hyper parameters. Unless the input signal is comprised of pixels and feeds hundreds and even thousands of features, normally an optimized feature vector consists of up to about 100 features. In the case of this particular work, the authors reported running tests with different number

of features containing 15 to 50 different frequency components. It is shown that SVM significantly surpasses the performance of the K-nearest neighbors (KNN) algorithm when it's using the 25 DCT feature set. Furthermore, the true positive and false positive rates are analyzed, as well as the precision and recall for both MAV and Real-time datasets [10, 15].

In an intriguing research titled: "Wearable IoT sensor-based healthcare system for identifying and controlling chikungunya virus" [16], the authors propose a novel technique for identifying and controlling the outbreak of chikungunya virus (CHV). IoT and fog-based healthcare system is discussed from a new angle, in light of the disease. CHV spreads quickly in geographically affected areas as claimed by the authors, and these areas are usually not well prepared for in-time diagnosis and treatment, hence new methods are needed using modern technological advancements. Cloud computing along with mobile technologies and wearable Internet of Things (IoT) sensors are well suited to provide tangible solution. The dangers of the CHV virus that had been running rampant in many parts of the world are discussed and a differentiation between the Zika and the CHV Virus is also provided. This smart system is typically divided in 3 layers, the first being the *Sensor Layer*. This layer collects data from various sensors that have been taken from the user's body, environment sensor, drug sensor, and climatic sensor. It passes on the information to the second layer that is the *FOG Layer*. The classification process is implemented. Firstly, the collected data is being processed and a classification whether the person is infected or not is performed. There is another algorithm that follows to collect the information from the concerned user's cell phone and sends out messages to the medical professionals in case of a medical concern. The third layer is the *Cloud Layer*. It completes all the rest of the functionality that was not presented in the FOG layer. It stores all the data and transmits it to hospitals or government agencies. It follows all necessary security protocols.

In another work titled "BSN-Care: A Secure IoT-Based Modern Healthcare System Using Body Sensor Network" [17], the researchers mention the prediction that by the end of 2050, 22% of the world's population would comprise of older citizens. As such, Health monitoring system takes an important stance in day-to-day activities. The article also talks about different BSN Care systems already in the market. Code blue developed by Harvard University, and Alarm-net developed by University of Virginia are among the few BSN Care systems already on the market. This smart system has biosensors attached to different parts of the users' body as they measure and send various data. A Local Processing Unit (LPU) identifies the data and acts on it according to some pre-set specifications. For example, if the Beats Per Minute (BPM) is less than 120, the LPU doesn't do anything, but if the BPM is above 120, it notifies the family members of this anomaly. If BPM is more than 130, and if the person lives alone, it notifies the emergency services. Obviously, security is a big part of such system and this paper identifies the different security points in the system. To name a few: Data Privacy, Data Integrity, Localization, Data Freshness are some of these points. Authentication is performed in two levels: the first level comprises of BSN Server issuing security credentials to LPU to initiate a secure channel. The second level is an authentication process where the system gives the LPU a random number; a shadow to initiate security measures. This way, if an attacker tries to modify or extract information, the system can detect it.

4 Smart Systems in Robotics Using Voice and Emotion Recognition

The topic of robotics and control systems is becoming more and more popular in recent years. To make the robots be more human-like we have to be able to talk to them in order to facilitate more natural environment for human-machine interaction. Commands are therefore given to robots by voice, then a signal-processing unit interprets them, and action is taken. Convolutional Neural Networks (CNNs) are usually used for training batches of images in order to find objects in images, so it is unusual to use CNN based systems with speech as presented in [18]. There, the authors used voice to detect emotion in companion robots through CNN. The system was implemented in Python via the main two libraries: Keras and Tensorflow used as backend. A brief description of the emotional content extraction from voice was presented in this paper. When it comes to choosing the appropriate number of emotions however, various researches are using a different subsets based on their ability to obtain specific datasets or simply based on the application of their work. Cowie and Cornelius performed one of the fundamental works in that area in 2003 [19], in which they suggested the so-called “big-six” emotions, namely: anger, happiness, sadness, fear, surprise, and disgust. Bhatti M. et al. later confirmed their work in 2004 [20], but very frequently we see that smaller emotional datasets are suggested such as in [21], in which only four emotions were used or: anger, happiness, sadness, and neutral. As reported in [18], the “big-six” emotional set was used as described in [19], and the reported accuracy of their system was 71.33%.

A big boost in this idea was also driven by Google in their research in the Audio Set project, where an analysis on over 2 million video records taken from YouTube was performed so that a large dataset was created consisting of over 600 audio events. In that research Google was using the Mel-Frequency Cepstral Coefficients (aka MFCCs) as feature detection, extraction and classification parameters from audio. They used GMM based classifier and their work is summarized in [18, 22].

The input layer consisted of 400×12 input neurons. In its core, a conventional CNN system of at least one convolutional layer, after the input layer, followed by pooling, then activation done by a *ReLU* unit, then followed by *Max Pooling*. The research mentioned in [18] used 20 such convolutional layers. The *ReLU* function can be thought of as a way of normalization as it is activation function that turns all negative numbers into zeros. All the rest are positive numbers multiplied by one. This alone will make our network learn non-linear problems.

In their implementation, the researchers used 200 kernels used with size 5×5 . After the feature extraction stage is complete, a typical CNN network ends with the classification stage in which flattening is used, where all the outputs gathered thus far from the network are reorganized into a single flat vector. This layer is immediately followed by a fully connected (or dense) layer consisting of 1000 neurons, which will give us all of the learning we have done so far from all of the neurons. Let's point out that their number had already decreased through applying the convolution. It is at the end of that stage where the previously mentioned six emotions were presented as options at the output. Typically, a *softmax* function is used here that will squash the results to match the classifiers at the end so that we get probabilities for our classes. Then using an *argmax* function the most probable emotion is chosen at the final stage. At the end, the

authors reported that after training the CNN with a set of 200 sound samples the results had a satisfactory performance.

5 Smart Systems in Robotics Using Voice Jitter

Emotion recognition is one of the most fundamental layers when it comes to natural communication. Sensing Emotion from Voice Jitter was proposed in [23]. In their research, the authors provide an interesting concept that involves the usage of voice jitter. This means that instead of using complete voice samples the authors used smaller pieces of speech to make their analysis. As it turns out from this research, short timeframes of speech would suffice with a relatively good confidence to be able to determine emotions from speech. This can be very convenient especially when you have to collect specific speech samples for training and testing. Labeled speech datasets for emotion recognition can be costly so this method provides an alternative to this problem. In turn, the solution not only makes the collection of samples easier, but the calculations in recognizing different emotions are also less costly. All in all, the researchers recorded thirty student volunteers aged between 20 and 25 years each speaking 3 motional states (angry, happy and sad), and there was one sample per recording for each emotion from each speaker (90 recordings in total). Each recording was between 7–15 s. That alone shows how small of a dataset was used in order to perform these experiments and make tangible conclusions.

In more details, here is what this system included:

- **Step 1:** in this step *initialization* of the process was performed, which included recording of all the samples from all the users, one for each emotion as already described, then voice samples were categorized in groups. Framing of each recorded voice sample followed then jitter was calculated [24]. After using a vocoder, the portions that carried voice samples were separated from the noise. The first signal frame of each voice portion was then convolved with a 250 ms Hamming Window. For each of the voice frames, jitter coefficients were calculated.
- **Step 2:** this was called *feature clustering* step in which the features were collected and grouped together. In this stage Vector Quantization (VQ) [25] was used for the K-Means clustering machine learning method. This step was necessary in order to make the task more manageable by introducing a vector codebook for the emotion clustering part. At the end of this stage, for each of the voice features an assignment of each of the three emotional clusters was performed.
- **Step 3:** this step was called *feature matching* in which, a test cluster was ‘matched’ to each of the training clusters. In order to compare the distances between each jitter and the Cepstral Coefficient (CC) in the testing part of the process, from the trained samples in the codebook, the mean Euclidian distance has been adopted as distance measurement.
- **Step 4:** in this so-called *co-efficient computing* stage, the jitter-based coefficient calculations have been revealed. The Mel scale has been utilized in this stage, as applied to each frame, by using thirteen triangular overlapping kernels on each frame in order to recreate the spectrum needed. It is in this stage where the authors mention that

the jitter seemingly reduces the long-term effects of the changes in fundamental frequency F_0 , the pitch (the perceived frequency). The *jitter* is described to be a small fluctuation of the pitch since it is rising or falling during voice production (as oppose to *shimmer*, which captures an amplitude instability). The relative jitter was referred to as *co-efficient*, which was found relevant for usage in this research instead of the CCs.

The recordings were originally completed in two languages and the average accuracy results for English were: 70.04 for happy, 67.10 for angry, and 70.53 for sad. The performance measure used was true acceptance, through Equal Error Rate (EER). Furthermore, the authors compared the Cepstral Constraint system, and the Jitter based system to MFCC and LPCC as described here [25, 26]. In addition, the authors found that, when the number of users is higher, then the EER is lower. In general, this method provides a low-cost framework base, with overall good accuracy for further exploration. In conclusion, emotion recognition based on jitter from voice had a comparable, if not even better performance than in the cases when conventional MFCCs were used.

6 Smart Systems in Entertainment Using Emotions

One of the areas that can greatly benefit from the advancements of smart system development is the entertainment industry. This, not only includes the advancements in personalization and recommendation based on previous experience of the user with the system (concepts that were heavily productized in recent years), but it must also take into account the specific needs of the user at the right time, which can be done either directly or indirectly. This need is especially prominent because of the growth of world media all around the globe. Hundreds and even thousands of new media is created on a daily basis, which makes the problem harder to solve. The solution to this kind of problem is never easy and it can be executed in many different ways. But what makes it smart is the interaction between human and machine. More specifically, the closer the communication between users and autonomous systems to natural person-to-person communication, the better and more natural that system will be. This means that an extra step must be taken in improving that communication. For example, we have seen automated systems with which we speak or show gestures, then some action is usually taken by the system. That is the very first level in human-machine interaction – a direct communication conveying a specific need. However, this may be too intrusive at times and may not have the desired effect by the user, as the communication may not be as natural. The reason for this is that some users get very self-conscious when they are filmed or knowingly recorded. In these kinds of cases, the message might get obscured and may not have the desired effect. But what if a system is introduced where these recordings can be done in the background with a one-time consent of the user (only the first time)? Then in these cases in addition to having something more powerful, we can also have a system that is very natural.

Some research has already been pointing in that direction, as seen in [27]. In this work the authors propose a service-oriented architecture (SOA) for connecting speech to digital information libraries. The scientists use emotion detection from speech as the

main method for meeting individual needs, which can be applied explicitly to content recommendation. Some basic speech criteria and strategies have been previously set out for the task in [28]. Six distinct states of emotion were incorporated there: happy, sad, angry, fear, surprise, and neutral. Discussed were also some of the most important prosodic features extracted and collected in both time and frequency domains. The work was later extended in a more practical case for a recommendation system [29], using gender separation based on voice. One of the most promising features in emotion recognition from speech signals was shown to be the Glottal Symmetry (GS) [30, 31]. This fact was investigated in [28]. Figure 1 [32], depicts a plot of a typical Glottal Symmetry for four emotions. It can be seen that different emotional conditions can be clearly separated. The idea of using GS as feature domain has been extensively tested with an Optimum-Path Forest classifier as shown in [33]. Furthermore, Gaussian Mixture Model (GMM) have also been used with prosodic features, in addition to Tonal and Break Indices (ToBI) [34] and Inter-Sentence time domain statistics [35]. A nice review of all of these algorithms and other various features was provided in [36].

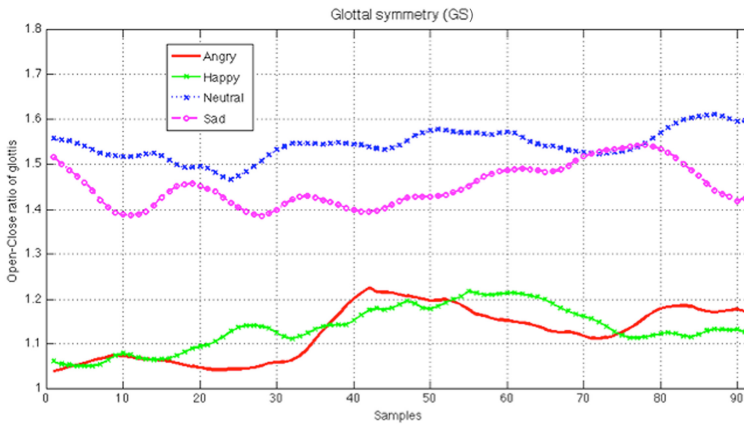


Fig. 1. Glottal Symmetry (GS) using four emotions [32]

In another study [37], an attempt to extract the sentiment from five books using “The Game of Thrones” was made through Natural Language Processing (NLP), implemented via Python’s NLTK package (Natural Language Processing Toolkit). The result was a novel content discovery system, which was exclusively based on human behavior. The idea could easily be applied to any TV or electronic book series as well as movies libraries or music players as described next.

In [38], a music-based approach has been proposed. It is also based on emotion recognition using voice. In their published work the authors use the following five emotional states: happiness, anger, boredom, sadness, and anxiety. Preprocessing is performed for each of the training and testing datasets in order to remove silent sections in the corpus. This was done implementing the Rabiner-Samur algorithm. To segment the speech in order to obtain the voiced sections a trivial end-point detection algorithm was carried out. For silence removing a Short-Time Energy (STE) and a Zero-Crossing

Rate (ZCR) were performed for better accuracy. Then feature extraction was performed. The feature domain solely included Mel Frequency Cepstral Coefficients (MFCCs). They are nonlinearly spread following a logarithmic law in the higher frequencies and are linearly spread in the lower frequencies. They are based on the human auditory hearing and are used quite frequently in the world of perceptual coding. This non-linear scale is also known as Mel-Frequency scale and is implemented using filter banks to obtain the MFCCs. Subsequently training with the newly composed feature vectors was done using two different classifiers. The testing determined the performance of the system and the result was directly applied to selecting a specific musical piece from a preselected musical databank. The authors used the Berlin database, which is a well-transcribed emotional database, comprised of voice sounds. In their speech samples the authors used a gender balanced ten-speaker dataset (five male and five female) of different age. The datasets included 1–5 s long samples. The two classifiers used were GMM and SVM. The results from the confusion matrixes for both showed successful accuracy [38].

In conclusion, SVM model showed better results with average accuracy at about 82%, while the GMM model with Maximum Likelihood (ML) showed a bit over 76% average accuracy when detecting emotions.

7 Smart Services for Finding Similarity in Images

More than 80% of our sensory experiences are visual. Similarity access and their conversion into high-level semantic features using fuzzy production rules, derived with the help of an image mining technique is given in [39]. Image recognition is used in the search of products or objects, by firstly identifying the objects, and then searching the network for similar patterns (IoT) [40].



Fig. 2. The bridge crossing Arno in Arezzo

Similarity among the series of the water lily pond grudes at Giverny painted by Claude Monet is studied in [41]. Many mysteries have been solved regarding the Da Vinci's Mona Lisa painting. The bridge in the painting was found crossing the Arno in Arezzo - near the city. In Fig. 2 is my photograph of the bridge. I was able to find

similarity with the bridge from the painting. In [42] retrieval by contrast and harmony is presented. Examples for search by light-dark, cold-warm and simultaneous contrast are given.

8 Smart Services for Extracting Higher-Level Visual Features

A tool for extracting higher-level visual features for art painting classification based on MPEG-7 descriptors was implemented in the system “Art Painting Image Color Aesthetics and Semantics” [43]. The analysis of the significance of the received characteristics and finding regularities between them can be used as discriminating semantic profile of the art paintings. It can predict several characteristics such as: the artists’ names, movements, themes, techniques, etc. The system allows also creating some datasets, containing extracted attributes or selected part of them labelled with chosen profile such as artist name, movement, scene-type. These datasets are used for further analysis by data mining tools for searching typical combinations of characteristics, which form profiles of artists or movements, or reveal visual specifics, connected to the presented theme in the images. Example for this is given in Fig. 3.

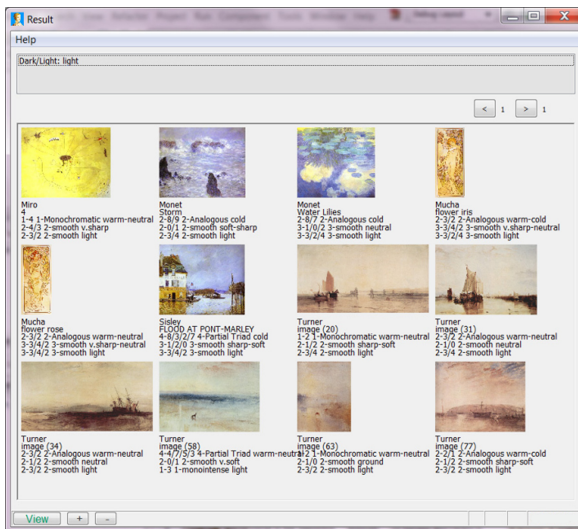


Fig. 3. Result of retrieval from the image base with parameter: “Dark/light contrast = Light” (includes “smooth light” and “monointense light”)

9 Smart Services for Image Data Mining

In the field of image data mining, an approach for extending the learning set of a classification algorithm with additional metadata is developed [44]. It is used as a base for the assignment of appropriate names to find regularities. The analysis of the correspondence

between connections established in the attribute space and existing links between concepts can be used as a test for the creation of an adequate model of the observed world. Meta-PGN classifier is suggested as a possible tool for establishing these connections. This approach is applied on the field of content-based image retrieval of art paintings by designing system architecture for the extraction of specific feature combinations, which represent different sides of artists' styles, periods and movements [45]. The system interacts with the user, displaying those parts of the mental model that are utilized in the name generation process. This interaction is used to further improve and extend the mental model.

10 Smart Services for Artist Identification

Artist identification has been an interesting task for centuries. Machine learning algorithms are used to solve this problem [46]. An artist's profile is obtained through an artist's merged images enhanced with layer and transparency tools. The machine learning J48 algorithm is utilized for classification. Cohen's Kappa, F-measure, and Matthew's correlation coefficient statistics are applied to compare the results obtained.

11 Smart Services for Autonomous Cars

Autonomous cars [47], or cars that run without human control, have been developed over the past several decades, starting in 1977. Currently, we have autonomous cars still in experimental and development stage that have driven autonomously thousands of miles. Some autonomy systems are already being used in the cars such as: Cruise Control, Anti-Lock Brakes. Some systems are just starting to be used: Stability and Traction Control, Pre-Accident Systems, Traffic Jam Assist, Self-Parking Systems. V2V (vehicle to vehicle) communication systems, are designed to prevent crashes in a number of scenarios such as: Intersection assist; Left-turn assist, Do-not-pass warning; Advance warning of a vehicle braking ahead; Forward-collision warning; Blind-spot/lane-change warning.

Computer vision and image processing are one of the main elements used in the autonomous car. Among them is Lidar technology (Light Detection and Ranging). It consists of puck-shaped device that creates a high-resolution map of the environment in real time. Some of the most important algorithms are for Lane Detection, Speed Bump Detection, Traffic Sign Detection, Steer by Wire System and recognition and classifying objects of different road types [48].

Lane Detection: Warning systems have already is available in many new cars. The algorithm includes:

- a) Find the Region of Interest (ROI);
- b) Image noise subtraction;
- c) Edge detection; and
- d) Lane markings.

Traffic Sign Detection: It includes preprocessing using color segmentation, threshold technique, Gaussian filter, canny edge detection and contours detection [49]. The next detection is based on Polynomial Approximation of digital curves technique applied on contours to correctly identify the signboard. Building new infrastructure, this problem can be solved with V2I (vehicle-to-infrastructure) communication through mobility applications. In V2V network, every car, smart traffic signal could send, capture and retransmit signals. Five to ten hops on the network would gather traffic conditions a mile ahead.

Speed Bump Detection: It can be done with training the speed bump detection system using a neural network.

More and more applications with autonomous cars appear such as: Fast-food chains and grocery stores are teaming up with big car companies and tiny startups to test the idea of autonomously shuttling food to customers. During tests in Miami, Michigan and Las Vegas, Domino's Pizza Inc. delivered more than 1,000 pies in Ford sedans plastered with signs that read "self-driving delivery test vehicle" [48].

12 Smart Services for Finding Parking Slot

One of the challenges to build smart cities is the smart parking. Several solutions have been proposed: different types of sensors (magnetometers, light sensors, microphones, etc.), different communication technology (wired, wireless), and different types of cameras. Smart Parking is a system capable of extracting specific information from the captured images and different sensors. Solutions based on computer vision and big data are deployable on top of visual sensor networks. The IoT (Internet of Things) paradigm fits particularly well in urban scenarios as a key technology for the Smart City Concept. In [50], an efficient solution for real-time parking lot occupancy detection based on Convolutional Neural Network classifier is presented. There, a real time image segmentation and analysis, and streaming data are used. It takes into account different light conditions, parts of the day, and seasons. The benchmark collection for parking occupancy detection [51] is used. Problems for solving are:

- a) Significant changes of lighting conditions - sunny, rainy and snowing days;
- b) Different time of the day;
- c) Partially occupant moving cars, people, and additional objects.

OpenCV library (<http://opencv.org/>) and Python can be used to find the frame spaces. The parking classification can be done with mix of the following techniques: background subtraction, defining and analyzing moving cars, applying Gabor filters as feature extractor to train a classifier with empty spaces under different light conditions, using edge detection algorithms. Deep Learning that allow computers to learn complex perception tasks such as Caffe system (<http://caffe.berkeleyvision.org/>) are used to train the neural networks. HAAR CASCADE can be used to detect moving cars [52].

New feature with Tesla called "Enhanced Summon" lets the drivers remotely call their car to drive itself through a parking lot to pick them up, so long as it's within 150 feet.

13 Smart Services for Video Filtering

An efficient video stream filtering solution based on MPEG-7 descriptors is described in [53]. The solution can be adopted for TV-Anytime, On Demand TV, Integrated Digital Television, Set-Top-Boxes, etc. It uses a novel approach called Pivoted Stream. The solution exploits the properties of metric spaces, in order to reduce the computational load of the filtering receiver.

The approach proposed video filtering that makes use of simple additional information (that is, indexes) sent with the video, eliminating the need for users to digest video that wouldn't pass the filter anyway. The approach requires a metric measure of similarity between the filter and the video representative (the feature); this measure is based on the pivot technique [54]. Filtering quality depends on the metric measure adopted. The approach is general and doesn't depend on a specific format to represent the video content, but it assumes the use of MPEG-7 to provide a description of the videos. In particular, it concentrates on the MPEG-7 visual descriptors, which covers basic visual features, such as color, texture, and shapes. Each source sends a video stream associated with an MPEG-7 stream that contains this video stream's description. These streams move through a generic transmission channel to the receiver stations (for example, set-top-boxes, digital multimedia recorders, media center PCs, and so on). Each receiver station has several filters that select (from all the video streams that arrive to the station) and deliver to the user only the video frames deemed relevant. Filtering is based on a comparison between each video frame with the filters, so that only frames similar to one of the filters are delivered to the user.

14 Conclusion

As discussed in this chapter, the need for improving services in areas like mobile services, communications, entertainment, healthcare, robotics, etc. is growing along with the complexity of data itself. Some of these areas are already employing multiple smart systems in various ecosystems to aid the work of personnel at hospitals or public service places, while others are used for entertainment and education. Some additional areas such as banking, security, and various other services related to warfare can also benefit greatly from this kind of technological expansions. In this chapter we showed how textual cues could be used as features to meet the public demand for smart system development. Moreover, we discussed in depth how sophisticated smart systems employ DSP and Machine Learning in order to process voices and images. It was further shown how using audio and visual cues in order to improve our everyday life could use these two areas for the creation of many distinct services. Additional research is necessary to lead to further advancement in smart system technology and improve the global IoT ecosystem.

References

1. <https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/>. Accessed Apr 2020

2. <https://www.networkworld.com/article/3325397/storage/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html>. Accessed Apr 2020
3. Lea, R.: Smart Cities: An Overview of the Technology Trends Driving Smart Cities (2017)
4. Bala, A., et al.: Voice command recognition system based on MFCC and DTW. *Int. J. Eng. Sci. Technol.* **2**(12), 7335–7342 (2010)
5. Parameshachari, B.D., Sawan, K.G., Gooneshwaree, H., Tulsirai, T.G.: A study on smart home control system through speech. *Int. J. Comput. Appl.* **69**(19), 30–39 (2013). 0975 – 8887
6. Alkhawaldeh, R.S.: DGR: gender recognition of human speech using one-dimensional conventional neural network. *Hindawi Sci. Program.* (2019). Article ID 7213717, 12 pages
7. Akçay, M.B., Oğuz, K.: Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication* **116**, 56–76 (2020)
8. Cowie, R., et al.: Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag.* **18**(1), 32–80 (2001). <https://doi.org/10.1109/79.911197>
9. Teager, H., Teager, S.: Evidence for nonlinear sound production mechanisms in the vocal tract. In: Hardcastle, W.J., Marchal, A. (eds.) *Speech Production and Speech Modeling*, pp. 241–261. Springer, Cham (1990). https://doi.org/10.1007/978-94-009-2037-8_10
10. Ghazanfar Latif, A.H., Khan, M., Butt, M., Butt, O.: IoT based real-time voice analysis and smart monitoring system for disabled people. In: *Asia Pacific Journal of Contemporary Education and Communication Technology*, Asia Pacific Institute of Advanced Research (APIAR), vol. 3, no. 2, pp. 227–234 (2017). <https://doi.org/10.25275/apjcectv3i2ict5>. ISBN (eBook) 978 0 9943656 8 2 | ISSN: 2205-6181
11. Smith, B.: Raspberry Pi Assembly Language RASPBIAN Beginners: Hands on Guide. CreateSpace Independent Publishing Platform (2013)
12. Kumar, S.S., RangaBabu, T.: Emotion and gender recognition of speech signals using SVM. *Emotion* **4**(3) (2015)
13. Schölkopf, B., Smola, A.J.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge (2002)
14. Belin, P., Fillion-Bilodeau, S., Gosselin, F.: The montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* **40**(2), 531–539 (2008)
15. Davis, J., Goadrich, M.: The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*, pp. 233–240. ACM, June 2006
16. Sood, S.K., Mahajan, I.: Wearable IoT sensor based healthcare system for identifying and controlling chikungunya virus. *Comput. Ind.* **91**(2017), 33–44 (2017)
17. Gope, P., Hwang, T.: BSN-care: a secure IoT-based modern healthcare system using body sensor network. *IEEE Sensors J.* **16**(5), 1368–1376 (2016)
18. Frant, E., Ispas, I., Dragomir, V., Dascalu, M., Zoltan, E., Stoica, I.C.: Voice based emotion recognition with convolutional neural networks for companion robots. *Romanian J. Inf. Sci. Technol.* **20**(3), 222–240 (2017)
19. Cowie, R., Cornelius, R.: Describing the emotional states that are expressed in speech. *Speech Commun.* **40**, 5–32 (2003)
20. Bhatti, M., Wang, Y., Guan, L.: A neural network approach for human emotion recognition in speech. In: *IEEE International Symposium on Circuits and Systems*, Vancouver, BC, pp. 181–184 (2004)
21. Noda, T., Yano, Y., Doki, S., Okuma, S.: Adaptive emotion recognition in speech by feature selection based on KL-divergence. In: *IEEE International Conference on Systems, Man, and Cybernetics in Taipei, Taiwan*, 8–11 October 2006, pp. 1921–1926 (2006)
22. Murray and Arnott: Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *J. Acoust. Soc. Am.* **93**(i2), 1097–1108 (1993)

23. Nazia, H., Mahmuda, N.: Sensing emotion from voice jitter. In: SenSys 2018, Shenzhen, China, November 4–7 2018, pp. 359–360 (2018). ISBN 978-1-4503-5952-8
24. Ganapathy, H.H.S., Mallidi, S.H.: Robust feature extraction using modulation filtering of autoregressive models. *IEEE Trans. Audio, Speech, Lang. Process.* **22**(8), 1285–1295 (2014)
25. Kheder, M.A.D., Bausquet, P.: Dealing with additive noise in speaker recognition systems based on i-vector approach. In: IEEE ICASSP, Canada (2013)
26. Atal, B., Hanauer, S.: Speech analysis and synthesis by linear prediction of the speech wave. *J. Acoust. Soc. Am.* **50**(2), 637–655 (1971)
27. Iliev, A.I., Stanchev, P.L.: Smart multifunctional digital content ecosystem using emotion analysis of voice. In: 18th International Conference on Computer Systems and Technologies CompSysTech 2017, Ruse, Bulgaria, 22–24 June 2017, volume 1369, pp. 58–64. ACM (2017). ISBN 978-1-4503-5234-5
28. Iliev, A.: Monograph: Emotion Recognition From Speech. Lambert Academic Publishing (2012)
29. Iliev, A.I., Stanchev, L.P.: Information retrieval and recommendation using emotion from speech signal. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval, Miami, FL, USA, 10–12 April 2018, pp. 222–225 (2018). <https://doi.org/10.1109/MIPR.2018.00054>
30. Iliev, A.I., Scordilis, M.S.: Spoken emotion recognition using glottal symmetry. *EURASIP J. Adv. Signal Process.* (2011). Article ID 624575, ISSN 1687-6180
31. Iliev, A.I., Scordilis, M.S.: Emotion recognition in speech using inter-sentence glottal statistics. In: Proceedings of the 15th International Conference on systems, Signals and Image Processing, IEEE-IWSSIP 2008, Bratislava, Slovakia, 25–28 June 2008, pp. 465–468 (2008)
32. Iliev, A.I., Stanchev, P.L.: Glottal attributes extracted from speech with application to emotion driven smart systems. In: Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018), KDIR, vol. 1, pp. 297–302, Thomson Reuters, Seville, Spain, 18–20 September 2018. ISBN 978-989-758-330-8
33. Iliev, A.I., Scordilis, M.S., Papa, J.P., Falcão, A.X.: Spoken emotion recognition through optimum-path forest classification using glottal features. *J. Comput. Speech Lang.* **24**(3), 445–460 (2010). ISSN 0885-2308
34. Iliev, A.I., Zhang, Y., Scordilis, M.S.: Spoken emotion classification using ToBI features and GMM. In: Proceedings of the 14th International Workshop on Signals and Image Processing 2007 and the 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. IEEE-IWSSIP 2007, Maribor, Slovenia, 27–30 June 2007, pp. 495–498 (2007). ISSN 16874722, 16874714
35. Iliev, A.I.: Emotion recognition in speech using inter-sentence time-domain statistics. *IJRSET Int. J. Innov. Res. Sci. Eng. Technol.* **5**(3), 3245–3254 (2016)
36. Iliev, A.I.: Feature vectors for emotion recognition in speech. In: National Informatics Conference, Sofia, Bulgaria, pp. 225–238 (2016)
37. Iliev, A.I.: Content discovery using perceptual automation. In: Proceedings of the 10th International Conference on Management of Digital EcoSystems (MEDES 2018), Tokyo, Japan, 25–28 September 2018, pp. 233–238. ACM (2018). <https://doi.org/10.1145/3281375.3281399>. ISBN 978-1-4503-5622-0
38. Lukose, S., Upadhy, S.: Music player based on emotion recognition of voice signals. In: 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT), IEEE 2017, pp. 1751–1754 (2017). ISBN 978-1-5090-6106-8
39. Stanchev, P.: Using image mining for image retrieval. In: IASTED International Conference Computer Science and Technology, Cancun, Mexico, 19–21 May 2003, pp. 214–218 (2003)
40. Viana, W.: Using images to extend smart object discovery in an Internet of Things scenario. file:///C:/Users/pstan/Desktop/4057-829-4030-1-10-20181009.pdf

41. Stanchev, P., Green Jr., D., Dimitrov, B.: Some issues in the art image database systems. *J. Digit. Inf. Manag.* **4**(4), 227–232 (2006)
42. Stanchev, P., Green Jr., D., Dimitrov, B.: High level color similarity retrieval. *Int. J. Inf. Theor. Appl.* **10**(3), 283–287 (2003)
43. Ivanova, K., et al.: Local features in APICAS (analyzing of added value of the descriptors based on MPEG-7 vector quantization). *Int. J. Comput. Sci. Artif. Intell.* **2**(4), 23–32 (2012). ISSN: 2226-4450 (online) 2226-4469 (print)
44. Radenski, A., et al.: Big data techniques, systems, applications, and platforms: case studies from academia. In: *Proceedings of the Federated Conference on Computer Science and Information Systems*, pp. 893–898 (2016)
45. Ivanova, K., Mitov, I., Stanchev, P., Ein-Dor, P., Vanhoof, K.: Establishing correspondences between attribute spaces and complex concept spaces using meta-PGN classifier. In: *Proc. of the 2nd International Conference on Digital Preservation and Presentation of Cultural Heritage*, V. Tarnovo, Bulgaria, IMI-BAS, Sofia, pp. 71–77 (2012). ISSN 1314-4006
46. Stanchev, P., Kolinski, M.: Novel artist identification approach through digital image analysis using machine learning and merged images. In: Rocha, Á., Ferrás, C., Paredes, M. (eds.) *ICITS 2019. AISC*, vol. 918, pp. 465–471. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11890-7_45
47. Stanchev, P., Geske, J.: *Autonomous cars. History. State of art. Research problems*. Springer Communications in Computer and Information Science, vol. 601, pp 1–10 (2016)
48. Viswanathan, V., Hussein, R.: Applications of image processing and real-time embedded systems in autonomous cars: a short review. *Int. J. Image Process. (IJIP)* **11**(2), 36–49 (2017)
49. Salhi, A., Minaoui, B., Fakir, M.: Robust automatic traffic signs detection using fast polygonal approximation of digital curves. In: *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, Marrakech, pp. 433–437 (2014)
50. Amato, G., Carrara, F., Falchi, F., Gennaro, C., Meghini, C., Vairo, C.: Deep learning for decentralized parking lot occupancy detection. *Expert Syst. Appl.* **72**, 327–334 (2017)
51. de Almeida, P.R.L., Oliveira, L.S., Britto Jr., A.S., Silva Jr., E.J., Koerich, A.L.: PKLot – a robust dataset for parking lot classification. *Expert Syst. Appl.* **42**, 4937–4949 (2015)
52. Stanchev, P., Geske, J.: Smart Parking. *Geoinformatics Research Papers*, vol. 5, BS1002 (2017). <https://doi.org/10.2205/codata2017>
53. Falchi, F., Gennaro, C., Savino, P., Stanchev, P.: Efficient video stream filtering. *IEEE Multimed.* 52–61 (2008)
54. Shapiro, M.: ‘The choice of reference points in best-match file searching’. *Comm. ACM* **20**(5), 339–343 (1977)

Journal Subline

LNCS 12630

Richard Chbeir
Guest Editor

Transactions on **Large-Scale Data- and Knowledge- Centered Systems XLVII**

Abdelkader Hameurlain • A Min Tjoa
Editors-in-Chief

**Special Issue on Digital Ecosystems
and Social Networks**

 Springer

Founding Editors

Gerhard Goos

Karlsruhe Institute of Technology, Karlsruhe, Germany

Juris Hartmanis

Cornell University, Ithaca, NY, USA

Editorial Board Members

Elisa Bertino

Purdue University, West Lafayette, IN, USA

Wen Gao

Peking University, Beijing, China

Bernhard Steffen 

TU Dortmund University, Dortmund, Germany

Gerhard Woeginger 

RWTH Aachen, Aachen, Germany

Moti Yung

Columbia University, New York, NY, USA

More information about this subseries at <http://www.springer.com/series/8637>

Abdelkader Hameurlain ·
A Min Tjoa · Richard Chbeir (Eds.)

Transactions on Large-Scale Data- and Knowledge- Centered Systems XLVII

Special Issue on Digital Ecosystems
and Social Networks

Editors-in-Chief
Abdelkader Hameurlain
IRIT, Paul Sabatier University
Toulouse, France

A Min Tjoa
IFS, Technical University of Vienna
Vienna, Austria

Guest Editor
Richard Chbeir 
University of Pau and the Adour Region
Anglet, France

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISSN 1869-1994 ISSN 2510-4942 (electronic)
Transactions on Large-Scale Data- and Knowledge-Centered Systems
ISBN 978-3-662-62918-5 ISBN 978-3-662-62919-2 (eBook)
<https://doi.org/10.1007/978-3-662-62919-2>

© Springer-Verlag GmbH Germany, part of Springer Nature 2021

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer-Verlag GmbH, DE
part of Springer Nature
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface

With the rapid advancements of Internet technologies, there is a shift in the focus of online applications towards more interaction and collaboration. Nowadays, there is a move from the Internet as a place of producers and consumers of content to a place of communities where everyone can publish information, interconnect, communicate, collaborate, and share. Moving from services provided by a single entity to more complex or integrated multi-stakeholder services requires new approaches for effective consideration of collaboration.

In this context, Digital Ecosystems have emerged to allow many digital entities/subsystems to interact by exchanging information in a wide variety of ways (Web, Cloud, etc.). They support in-between cooperation and promote collective knowledge sharing in order to provide mutual benefits, as a new way to handle collaboration in a distributed and heterogeneous environment. Interdependencies between entities can be based on automated, semiautomated, or manual relationships. The individual entities have their own purposes to achieve and could be managed locally. However, the efficiency of the whole Digital Ecosystem depends on the consistency of its use and a global, intelligent, balanced coordination of the various resources.

To model, develop, simulate, and validate such complex systems, several challenges remain. The aim of this special issue is to show several current studies addressing these challenges and evincing interesting research directions. The special issue is organized in self-contained papers to provide the greatest reading flexibility. It includes nine papers that have been selected after a very tight peer review, in which each paper has been reviewed by three reviewers. Several topics are addressed in the special issue, but mainly: **Social Big Data, Data Analysis, Cloud-based Feedback, Experience ecosystems, Pervasive Environments, and Smart Systems**. It is organized as follows.

The first paper is titled “Mapping Experience Ecosystems as Emergent Actor-Created Spaces” and is authored by *Andrea Resmini* and *Bertil Lindenfalk*. It introduces a conceptualization of experience ecosystems as semantic blended spaces instantiated by the activities carried out by independent actors moving freely and at will between different products, services, devices, people, and locations in pursuit of individual goals. This conceptualization is anchored to three distinct cultural and socio-technical shifts that characterize the current postdigital condition: the displacement of postmodernism as the cultural dominant; the embodiment of digitality and the emergence of a blended space of action; the occurrence of a postdigital society. It contributes to ongoing conversations on ecosystem-level and systemic design from the point of view of information architecture and user experience in five distinct ways: by centering the discourse on the actor-driven individual experience made possible by the postdigital condition; by framing the problem space from an embodied, spatial, and architectural perspective; by considering the environment systemically as a blend of digital and physical non-contiguous spaces; by recasting the object of design to be the semantic and spatial relationships that exist or could exist between the elements of the actor-centered

eco-system; by introducing a mapping methodology that can be used to capture and spatially describe the relational complexity of said ecosystems for further intervention.

In the second paper of this special issue, *Kurosh Madani, Antonio M. Rinaldi, and Cristiano Russo* propose “A Semantic-Based Strategy to Model Multimedia Social Networks”. Here, the authors explore the decisive role of the social facet of information in our quotidian life. An abstract representation and a proper management of Online Social Networks (OSNs) constitute a new challenge for communities of researchers. In addition, the need to extend OSNs to Multimedia Social Networks (MSNs) comes from the fact that the vast majority of data is unstructured and heterogeneous, making the reuse and integration of information effortful. In this paper, the authors propose a general high-level model to represent and manage MSNs. The proposed approach is based on property graphs represented by a hypergraph structure due to the intrinsically multidimensional nature of social networks and semantic relations to better represent the network’s contents. Using the proposed graph structure singles out several levels of knowledge and helps in analysis of the relationships defined between nodes of the same type or different types. Moreover, the introduction of low-level multimodal features and a formalization of their semantic meanings give a more comprehensive view of the social network structure and content. The proposed data model could be useful for several applications. A case study is proposed in the cultural heritage domain.

In the third paper, titled “Social Big Data: Concepts and Theory”, *Hiroshi Ishikawa and Yukio Yamamoto* explain the basic concepts of social big data and its integrated analysis. First, they explain the outline and examples of the real-world data, open data, and social data that compose social big data. They then describe interactions among the real-world data, open data, and social data. They also introduce basic concepts of an integrated analysis based on the “Ishikawa concept.” Furthermore, after explaining the flow of integrated analysis in line with the basic concept, a data model approach for integrated analysis is introduced.

Hiroshi Ishikawa and Yasushi Miyata propose “Social Big Data: Case Studies” in the fourth paper. Based on the concepts and theory introduced in the previous paper “Social Big Data: Concepts and Theory”, the authors concretely explain hypothesis generation and integrated analysis through several use cases.

In the fifth paper, “Data Analysis in Social Network: A Case Study” is proposed by *Mou De, Anirban Kundu, and Nivedita Ray De Sarkar*. Here, the authors propose a structural design of social networks to study the architecture of social networking sites and its working principles. Typical social networking sites have a three-tier architecture which induces higher searching time for user queries. The proposal presents a load-balancing module for protecting user enquiries before spreading them to the data server. In this paper, query optimization of user queries for faster results is discussed. Experimentation results exhibit possibilities of data (user queries) failure reduction due to external disturbances. The authors have analyzed large-scale data of a social network through a graph to reduce data loss and minimize network failure to maintain scale-free growth in the social network. Properties of the interface module and growth coefficient are analyzed to exhibit benefits of the proposed system architecture for balancing the load from the web server to the data server through the Hash table cache, Log table and index control module with scale-free query optimization.

“Smart Services Using Voice and Images” is proposed as the sixth paper by *Alexander I. Iliev* and *Peter L. Stanchev*. Here, the authors emphasize some of the most prominent advances in smart technologies that formulate the smart city ecosystem. They highlight the automation of numerous developments based on the extraction and analysis of digital media, using speech and images. At present, a multitude of practical systems is used for personalization and recommendation of different media. On the other hand, assorted types of services in different areas directly benefit from these advancements. Most of them were created with human-machine interaction methodology in mind, where people have to interact with the machines in various ways. In the past, this type of interaction has been completed through the use of conventional interfaces such as a mouse and a keyboard, where the user had to type a response manually, which was in turn recorded by the machine for subsequent analysis. Therefore, in order to simplify these types of interactions and lead to improvement of services, new methodologies must be studied, discovered, and developed so as to improve services such as recommendation and personalization services.

The seventh paper is dedicated to “Big Spatial and Spatio-Temporal Data Analytics Systems”, authored by *Polychronis Velentzas*, *Antonio Corral*, and *Michael Vassilakopoulos*. It is true that we are living in the era of Big Data, and Spatial and Spatio-temporal Data are not an exception. Mobile apps, cars, GPS devices, ships, airplanes, medical devices, IoT devices, etc. are generating explosive amounts of data with spatial and temporal characteristics. Social networking systems also generate and store vast amounts of geo-located information, like geo-located tweets, or mobile users’ captured locations. To manage this huge volume of spatial and spatio-temporal data, we need parallel and distributed frameworks. For this reason, modeling, storing, querying, and analyzing big spatial and spatio-temporal data in distributed environments is an active area for research with many interesting challenges. In recent years, a lot of spatial and spatio-temporal analytics systems have emerged. This paper provides a comparative overview of such systems based on a set of characteristics (data types, indexing, partitioning techniques, distributed processing, query language, visualization, and case-studies of applications). The authors present selected systems (the most promising and/or most popular ones), considering their acceptance in the research and advanced-applications communities. More specifically, they present two systems handling spatial data only (SpatialHadoop and GeoSpark) and two systems able to handle spatio-temporal data, too (ST-Hadoop and STARK) and compare their characteristics and capabilities. Moreover, they present in brief other recent/emerging spatial and spatio-temporal analytics systems with interesting characteristics. The paper closes with conclusions arising from investigation of the rather new, though quite large world of ecosystems supporting management of big spatial and spatio-temporal data.

The eighth paper addresses “Cloud Based e-Feedback Services Using Performance Analysis: A Linear Approach”, authored by *Ayan Banerjee* and *Anirban Kundu*. Here, the authors propose an online feedback system having distinct layers to access frameworks through multiple entry points such as student module, administration module, and teacher module which can be operated from any geographically distributed locations. There is no need to install software-based applications and no need for extra hardware expenses to access the proposed cloud-based system, due to the usage of software-as-a-service and platform-as-a-service. Students provide specific

information to the server-side for authenticity regarding entry to feedback questionnaires. Administrative authorities analyze teacher performance based on students' feedback. A teacher observes individual performance from the server. Human effort and human activities have been reduced due to usage of paperless feedback. Teacher performance is measured using preparedness, class-performance, responsiveness, effectiveness, and overall grade. Different nodes are required in the proposed system for distributing and replicating server-side data storage. Time consumption and load distribution of servers are analyzed based on the number of users and servers. Different nodes are accessed by multiple users working with different or the same modules of the system. An energy-efficient framework is incorporated into the proposed system to enhance system performance. The authors have incorporated different weighting factors in the energy efficient framework using distinct layers of the proposed system. Time complexity and space complexity are measured using proposed algorithms. A Web-based approach is required in the proposed system to reduce manpower consumption and workload. A comparative study between existing feedback systems and the proposed feedback system is established based on different characteristics.

The last paper of this special issue is titled "Semantic-Based Automatic Generation of Reconfigurable Distributed Mobile Applications in Pervasive Environments" and is written by *Abderrahim Lakehal*, *Adel Alti*, and *Philippe Roose*. It addresses a practical problem related to mobile applications available through interconnected smart connected objects. Several existing research works suffer from a lack of a distributed semantic-based agile strategy to improve accuracy and increase the system's efficiency. To address this problem, this paper comes up with a new flexible, modular, and hierarchical loosely coupled framework to efficiently generate context-aware applications based on a user's location and his situation. The classification of the user's situations reveals new insight on identifying efficiently hierarchical composite situations in order to meet the quality of the user's constraints. It ensures minimum execution time for context-aware distributed mobile applications using a parallel and distributed strategy. Firstly, the framework filters the contextual user's constraints of different smart-domains into domain-specific user's context. Then, it detects parallel incoming events captured by sensors that are able to identify factorized composite situations. Based on these identifications, the authors automatically generate the application reconfiguration as a collection of adapted services that are deployed on available distributed devices. They compare the situation identification performance of the proposed reasoning approach to efficient map-reduce implementations for healthcare systems. Experimental results show the effectiveness, reusability, and scalability of the proposed approach.

We hope this special issue motivates researchers to take the next step beyond building models to implement, evaluate, compare, and extend proposed approaches. Many people worked long and hard to help this edition become a reality. We gratefully acknowledge and sincerely thank all the editorial board members and reviewers for their timely and insightful valuable comments and evaluations of the manuscripts that greatly improved the quality of the final versions. Of course, we offer thanks to all the authors for their contribution and cooperation. Finally, we express our thanks to the editors of TLDKS for their support and trust in us. Special thanks go to Gabriela

Wagner for her high availability and her valuable work in the realization of this TLDKS volume.

November 2020

Richard Chbeir

Organization

Guest Editor of this Special Issue SI on *Digital Ecosystems and Social Networks*

Richard Chbeir

University of Pau and the Adour Region, France
Richard.chbeir@univ-pau.fr
University Pau & Pays Adour, LIUPPA,
Parc de Montaury, 64600 Anglet, France

Editors-in-Chief

Abdelkader Hameurlain
A Min Tjoa

IRIT, Paul Sabatier University, France
IFS, Technical University of Vienna, Austria

Editorial Board

Reza Akbarinia

Inria, France

Dagmar Auer

FAW, Austria

Djamal Benslimane

Claude Bernard University Lyon 1, France

Stéphane Bressan

National University of Singapore, Singapore

Mirel Cosulschi

University of Craiova, Romania

Dirk Draheim

University of Innsbruck, Austria

Johann Eder

Alpen Adria University Klagenfurt, Austria

Anastasios Gounaris

Aristotle University of Thessaloniki, Greece

Theo Härder

Technical University of Kaiserslautern, Germany

Sergio Ilarri

University of Zaragoza, Spain

Petar Jovanovic

Univ. Politècnica de Catalunya, BarcelonaTech, Spain

Aida Kamišalić Latifić

University of Maribor, Slovenia

Dieter Kranzlmüller

Ludwig-Maximilians-Universität München, Germany

Philippe Lamarre

INSA Lyon, France

Lenka Lhotská

Czech Technical University in Prague, Czech Republic

Vladimir Marik

Czech Technical University in Prague, Czech Republic

Jorge Martinez Gil

Software Competence Center Hagenberg, Austria

Franck Morvan

Paul Sabatier University, IRIT, France

Torben Bach Pedersen

Aalborg University, Denmark

Günther Pernul

University of Regensburg, Germany

Soror Sahri

LIPADE, Paul Descartes University, France

Shaoyi Yin

Paul Sabatier University, France

Feng “George” Yu

Youngstown State University, Youngstown, USA

SI Editorial Board

Abdelkader Hameurlain	Paul Sabatier University, France
Adel Alti	Const. Univ., Algeria
Anastasios Gounaris	Aristotle University of Thessaloniki, Greece
Anna Formica	Istituto di Analisi dei Sistemi ed Informatica-CNR Italy, Italy
Anne Laurent	Univ. Montpellier, France
Apostolos Papadopoulos	Aristotle University of Thessaloniki, Greece
Aristidis Likas	UOI, Greece
Demetris Trihinas	University of Nicosia, Cyprus
Elio Mansour	Université de Pau & des Pays de l'Adour, France
Georgia Koloniari	UOM, Greece
Helen Karatza	Aristotle University of Thessaloniki, Greece, Greece
Hiroshi Ishikawa	Tokyo Metropolitan University, Japan
Joe Tekli	Lebanese American University, Lebanon
Jose R. R. Viqueira	USC, Spain
Joyce El Haddad	UDO, France
Khouloud Salameh	American University of Ras Al Khaimah, UAE
Maria Luisa Sapino	Università degli Studi di Torino, Italy
Masaharu Hirota	Okayama University of Science, Japan
Michael Granitzer	University of Passau, Germany
Ramzi Haraty	Lebanese American University, Lebanon
Richard Chbeir	University Pau & Pays Adour, France
Sabri Allani	University Pau & Pays Adour, France
Salma Sassi	University of Jendouba, Tunisia
Sergio Ilarri	University of Zaragoza, Spain
Yannis Manolopoulos	Open University of Cyprus, Cyprus

Contents

Mapping Experience Ecosystems as Emergent Actor-Created Spaces	1
<i>Andrea Resmini and Bertil Lindenfalk</i>	
A Semantic-Based Strategy to Model Multimedia Social Networks	29
<i>Kurosh Madani, Antonio M. Rinaldi, and Cristiano Russo</i>	
Social Big Data: Concepts and Theory.	51
<i>Hiroshi Ishikawa and Yukio Yamamoto</i>	
Social Big Data: Case Studies.	80
<i>Hiroshi Ishikawa and Yasushi Miyata</i>	
Data Analysis in Social Network: A Case Study	112
<i>Mou De, Anirban Kundu, and Nivedita Ray De Sarkar</i>	
Smart Services Using Voice and Images	137
<i>Alexander I. Iliev and Peter L. Stanchev</i>	
Big Spatial and Spatio-Temporal Data Analytics Systems.	155
<i>Polychronis Velentzas, Antonio Corral, and Michael Vassilakopoulos</i>	
Cloud Based e-Feedback Services Using Performance Analysis: A Linear Approach	181
<i>Ayan Banerjee and Anirban Kundu</i>	
Semantic-Based Automatic Generation of Reconfigurable Distributed Mobile Applications in Pervasive Environments	213
<i>Abderrahim Lakehal, Adel Alti, and Philippe Roose</i>	
Author Index	235