*XXIX Simpozijum o novim tehnologijama u poštanskom i telekomunikacionom saobraćaju – **PosTel 2011**, Beograd, 06. i 07. decembar 2011.*

SOCIAL SIGNAL PROCESSING (SSP) IN TELECOMMUNICATION SERVICES

Andrej Košir
Faculty of Electrical Engineering, University of Ljubljana, Slovenia

Abstract: *The ability of modern telecommunication services to adapt more and more aspects of their service to users' needs is becoming of crucial importance. This is due to the fact that these services are getting increasingly complex and are also entering the user's extremely sensitive social domain. Approaches developed to mitigate these issues rely on user data acquired in a non-intrusive way. Social signal processing is a novel approach to acquiring and understanding user-related data and utilizing it in a targeted domain. This paper presents selected aspects of social signal processing in user-centered communication services. We focus on visual information acquisition via video cameras and video processing as one of the promising SSP data and processing approaches.*

Keywords: *Social Signal Processing, Telecommunication Services, Social Events*

1. Introduction

The development of telecommunication services has brought telecommunication systems and their users to the point when the user-centric approach and service adaptation to users is not an exception but is gradually becoming common practice. The reason for this is as follows: the complexity of services and ways to handle services and content have reached the point where a vast majority of users would not be able or willing to make the effort required to use and control them. User adaptation approaches have developed from simple adaptations [1] to what are now context-aware adaptations to users [2]. The ever-present difficulty of adapting services to users is that of user- and context-related data collection. Despite the fact that there has been considerable progress in the field of sensors and their accessibility (accelerometers, positioning systems in smart phones, etc.), user data collection still remains one of the unsolved issues.

There are several reasons for this. Among others, there is the sensitive nature of user data and the problem of the limited possibilities of implicit user data collection (explicit acquisition can be irritating for users). Besides, the next steps towards more efficient user adaptation require measuring not only the user's location and status, but also the user's emotional state, his relationship to other users, etc. This is where social signal processing (SSP) comes in [3]. The goal of SSP is to measure and process social signals. In the telecommunications field, the goal is also to utilize the results of this analysis in order to improve users' quality of experience (QoS). As is well known, user quality of experience for communication services is difficult to estimate, either from social signal analysis or from low-level service data such as packet delays and jitter. Among other reasons this is due to the fact that a user's opinion on a given communication service or metadata content item may reveal high variations from user to user and for the same user in different situations. These and other reasons indicate that an effective user adaptation service must take users' social signals into account.

As our efforts in the field are mainly concentrated on video analysis-related SSP, this paper mainly deals with video analysis-based systems and development. Video acquisition is regarded as a non-intrusive method, even though it is not free of security and privacy issues. It is believed that main social signals intentionally or non-intentionally sent out by the user can be recognized on the basis of visual signals, since during the evolution of natural selection such skills have improved the chances of survival of individuals and groups. Audio signals are also an important example of social signals in everyday life. In terms of social signal acquisition and the analysis of sound, there are some limitations such as the following: it is impractical to record sound produced by a large number of users. Besides, the surrounding noise heavily impedes the possibility of social content recognition, such as a user's emotional state [4].

The rest of the paper is organized as follows. In Section 2 we list and discuss several aspects of social signal processing. Section 3 contains historical notes on the origins of SSP relevant to understanding the modern and expected future role of SSP in the field of telecommunications. Section 4 describes selected goals of SSP in modern telecommunication services. A broader range of tasks at least partially solved using SSP is given in Section 5. SSP as a scientific and research challenge is presented in Section 6 and its expected future development is given in Section 7. The paper ends with the conclusions in Section 8.

2. What is Social Signal Processing (SSP)?

There are several different descriptions and definitions of SSP. They are focused on different aspects and fields of application, as well as on different approaches and methods used to solve targeted tasks.

In order to introduce SSP, we need to define what a social signal is. The broadest definition of signal is anything that caries information. Therefore a social signal is any signal that carries information on a user's social status, his mood, his relationship to others, etc. SSP is now simply described as the processing of social signals.

According to SSPNet [3], social intelligence [5] is the facet of our cognitive abilities that guides us through the complex web of our everyday interaction. SSP is the autonomous analysis of these interactions. Within our scope this means that we first need to measure these social signals, analyze them to extract relevant information and utilize it in order to improve the quality of service. A need for the introduction of the term social intelligence also indicates that social signals are related to human activity while using telecommunication services, such as making decisions about content items, about using or not using selected services, etc. in a complex manner.

As shown in Figure 1 depicting a situation where only two humans are interacting, social signals are very difficult to describe accurately, measure and process in order to extract the required meaning they convey. Modern techniques of user data acquisition (video cameras, body sensors, etc.) are not yet able to measure social signals in all their complexity, especially not in the non-intrusive way acceptable in real-world applications.

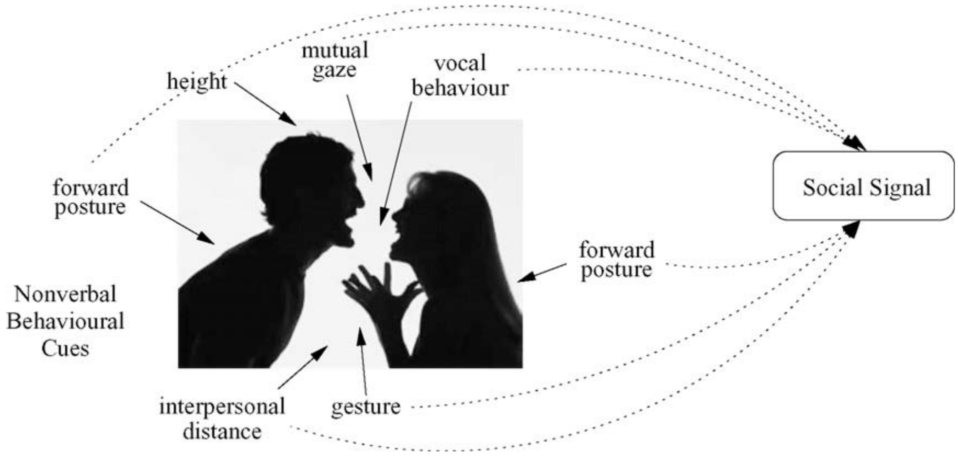

Figure 1. *Social signals from behavioral cues, reproduced from [6].*

3. Historical notes and origins

Social signal processing was introduced by A. Pentland [6], A. Vinciarelly and M. Pantić in 2008 [7]. As is always the case, new content was built on several foundations established before. Among others, these are social intelligence [5] and other areas of the social sciences, sensors, pattern recognition, signal processing, data fusion, data mining and machine learning. In the context of this paper, building blocks are also sound, image and video processing, telecommunication networks and surveillance systems.

Paradigm shifts that allow the introduction of SSP into the field of user-centric telecommunication services [1] are distributed sensor networks [8], pervasive computing [9] and especially the availability of sensors in smart phones with a rapidly growing number of users. These sensors are accelerometers, positioning sensors, illumination sensors and cameras monitoring the user while he is using the device. The number of users that are using smart devices is in itself an important parameter of any user-adapted service. Adaptation approaches such as user interface adaptation and item / service recommender systems are enhanced by the number of users using the same service [4].

4. Social signal processing, video processing and telecommunication services

As previously indicated, the goal of SSP in the context of user-centered personalized telecommunication services [1] is to recognize a user's social signals in real time and in a real environment and to utilize the result to improve the service quality. In this regard, the communication service user interface (the device to support it) is missing in Figure 1. One can imagine that one of two persons is replaced by a communication device such as mobile terminal. The situation would change dramatically in several ways. The device does not provide social signals, at least today's devices do not. but this can change in the future. However, several communication aspects remain present, since the communication is controlled by the person to the extent that the communication device can support it. Therefore, if the device would be able first to measure, understand and utilize human social signals and secondly to use its own social signals in communication, the whole communication process would be better adapted to humans – the users of the telecommunication service.

5. State of the art – targeted tasks and applications

Selected goals and tasks of SSP addressed by modern data-processing techniques are presented in this section. As already indicated, we focused on real-time video-processing techniques as a source of user-related data from which we extracted users' social signals. We grouped them according to the number of users participating in the communication.

5.1. Single user

The present technologies for understanding the social signals of a single user are not exclusively but mainly focused on the observation of his or her face mimics and sound. The most frequently extracted information is based on the estimation of current users' emotions (mood) by means of video [4] or sound [10] analysis. The results of these methods are given in terms of an emotional space such as VAD [11]. The success of these methods varies according to the circumstances the user is in (accuracy from 60 to 85% for relatively good conditions). The events analysis currently under development covers handshaking, head movements as a sign of approval or disapproval, etc. [3].

The application of these techniques is an improvement of user decision modeling and prediction while using a telecommunication service (recommender systems and others) [12]. It is known that a user's emotional state affects his decisions and is therefore relevant in user decision modeling and prediction. Emotional state is also relevant in estimating a user's attitude toward selected service elements and user interface-related procedures.

5.2. Small number of users

The term small number of users in this context usually means a group of users where each individual can be tracked in order to locate his position and analyze his appearance. At this stage of development, recognizing the social signals of two users is very important, such as handshaking, approval and disapproval of each other, etc. [3]. Video-processing techniques are developed for each specific situation. In terms of pattern recognition, data distribution in the feature space is extremely difficult. It requires sophisticated techniques such as random fields and Monte Carlo methods [13][14]. Besides the significant effort required to develop and implement such algorithms, real-time processing requires unreasonably powerful hardware. This is also due to the fact that the underlying algorithms are not standardized and suitable for parallelization.

Applications of this type of analysis are the analysis of business meetings, analysis of round table debates, political debates, etc. Currently the SSP techniques are not able to analyze and manage ongoing relationships among users in real time but only for offline analysis of listed and other happenings.

5.3. Large number of users – crowd

An example of a large number of users (crowd) are sports event spectators, cinema- and theater-goers, customers of large shopping centers and others [15]. Clearly, individual persons either cannot be tracked due to signal resolution and quality and specific obstacles such as occlusion, or there is no need to do so.

The application of automatic crowd analysis based on video processing is in public space management and design such as shopping centers, multipurpose buildings, public parks, pedestrian areas and others [16]. Special effort has been made to estimate crowd states in terms of the safety analysis of public space.

6. SSP as a scientific and research challenge

In this section, we briefly present a selection of the most difficult challenges for the application of SSP in the field of telecommunication services.

6.1. The nature of processed data

There are two sources of randomness in measured social signals. The first one is the nature of human behavior. It is known that the way in which one decides and acts in a given situation varies from person to person. If exactly the same situation could be repeated twice, even the same person would not decide and act in the same way. His/her actions also depend on his/her history of interaction with the communication device, on the day's events influencing his mood, etc. The outward appearance of one's actions also varies. If a person tries to exactly repeat the movement of his head and/or his hand while, for instance, selecting a telephone contact in a mobile phone directory, there will be high variations among different trials. Therefore, there is always randomness in the production of social signals and consequently in data measured to recognize social signals such as sound and video [6].

The second source of randomness are sensors together with environmental factors such as background noise, illumination, etc. Each sensor of a video camera produces a noise that is added to the recorded image. The same is true for sound sensors, body sensors, etc. This source of randomness is highly dependent on environmental factors and consequently on the situation the user is experiencing. As a result, the quality of measured social signals and the effectiveness of SSP greatly depend on the situation in which users are using their communication devices [13].

6.2. System complexity

The utilization of SSP in telecommunication services can be effective only if a large number of users are using a given set of services [12]. This means that a large number of sensors monitoring large numbers of users in different situations are added to the already complex telecommunication networks, services and different user terminals such as smart phones, etc. The integration of complex subsystems into a user- friendly ubiquity service is of key importance here. This process is progressing slowly, also due to external parameters such as economic factors and local regulations and legislation.

6.3. Challenges in video processing

Video acquisition and processing as social signal measuring has inherited all (partially) unsolved challenges of classic real-time, real-environment video processing. Selected specific tasks in a controlled environment are solved efficiently [3]. No general-purpose techniques such as general object recognition and tracking are available using classic computing paradigms. Due to the high complexity and variability of human actions producing social signals, the dependence of these solutions on a specific task only is very limiting. Current efforts are focused on simplified tasks called "atom actions", such as drinking water, paying attention, etc. [3]. When more general real situations are tackled, video processing becomes very complex and both time- and space-(memory)consuming for the underlying hardware, see [13]. One can observe that there are several techniques involved, such as 3D transformations [14], Monte Carlo Markov chains, random fields, etc. The need for these approaches indicates the difficult nature of processing data in order to achieve automatic social signal understanding [16][17][18][19][20]. Besides, the results achieved are good in controlled scenes (accuracy up to 85%) but poor in real scenes (accuracy 25%). A difficult challenge is to prepare suitable training sets for the development of algorithms [21] [22].

As briefly described in Section 5, there are promising results on crowd dynamic modeling and analysis. It must be pointed out, though, that these SSP results are not expected to be applicable in telecommunication services in the near future. However, the development of communication services involving a large number of users gathered in the same place and at the same time may benefit from crowd analysis results. Examples of such services are promotional events, fairs, etc.

6.4 Uncontrolled environment

Huge variations in the kind of situations in which users are using their communication devices require not only the measurement of social signals, but also handling all these usage scenarios in terms of machine learning. Note that the dependence of signal processing techniques on situation specifics also contributes to the variations in the types and distributions of data analyzed by machine learning techniques. As is known from pattern recognition algorithms, the higher variability of training data within each classification class relative to the between-class variability lowers the accuracy of results and requires separate training for classification subclasses. This is also true for the majority of applications of social signals in communication systems and leads to non-universal fragmented solutions.

6.5. The impact of SSP on society and the abuse of SSP

It is evident that social signals are delicate and sensitive types of data in terms of user integrity and privacy. Gathering, analyzing and storing these types of data is strictly regulated by legislation. There are immense ways of abusing SSP acquired in one’s private and professional activities.

Apart from misuse, there are several other impacts of modern communication services on modern society such as the problem of the authorship of produced content, the internal dynamics of several traditional society groups from the primary and secondary school community all the way up to political activities and socially marginal groups. Since it is simply not possible to predict and regulate a potentially negative phenomenon in society before it arises and the fact that protective measures are highly dependent on individual awareness regarding regulated issues, makes these issues even more critical.

7. Expected developments in the field in the near future

In this section we describe the expected development of social signal processing in terms of hardware and software. We also provide selected comments and opinions on major challenges to be solved.

7.1. Sensors, data acquisition and hardware development

The availability of sensors has made significant progress in the last few years [8]. MEMS devices (accelerometers), positioning sensors and cameras observing users, when added to modern smart phones and a fast-growing number of devices in use, has made the application of SSP in telecommunication services possible. The company producing Android smart phones and tablets has unofficially announced that they sell one million devices per day worldwide.

Regarding the availability of video cameras, some progress has also been made lately. Classic video surveillance system cameras are in operation in most public places such as shopping centers and cinemas. There are systems integrating over 2000 video cameras. Recent progress has been made in the direction of the distribution of computing power outside centralized processing. Cameras integrating Linux PC platforms capable of relatively complex real-time video processing are available for €600. It is intended that several steps of video-processing should be performed locally in order to save the communication capacity and computing power of central units. There are many hardware platforms combining these two approaches. Typically, a video stream from 2 to 8 cameras is collected and processed in one unit. It consists of two sub-modules, a predefined one for most frequently used processing (noise removal, edge detection, etc.) and a customized one designed according to specific application requirements. The technologies used are conventional microcomputers and FPGA-like circuits for regular and time-critical processes. Several such modules are combined into a single video collection and analysis service.

Using local processing (inside the camera or close to the camera) has additional benefits in terms of sensitive data protection. Raw video data is clearly a very sensitive issue and the best solution is that it never leaves the camera and only the processing results are sent out as a higher-level event description.

Another direction of camera development has resulted in high-sensitivity cameras that allow single-person tracking from the distance of one kilometer. Depth cameras where each pixel does not represent illumination or color but the distance to the object are commercially available.

7.2. Human/computer interfaces

As indicated in Section 2, communication between the device and its user should be two-way. At the moment, the research community is focused on user-related data made available to the communication device [1]. Haptic user interfaces are trying to add the reverse direction, when the user has a real-time feedback from the communication device.

Recent progress in bio-inspired robotics promises significant progress in the near future in terms of 5 to 10 years. The higher resolution of neuro-signal measurement directly from human brain activity will enhance communication from the user to the device. Methods for a more direct way of conveying information to the user such as projecting images and even direct stimulation of human brain are also progressing.

Obviously the effective, widely acceptable application of SPP is mainly dependent on human computer interfacing. Once the potential of this research field has been utilized, it is expected that SSP will become a standard part of advanced communication devices [6].

7.3. Communication networks and services

The potential for the application of SSP telecommunication services also depends on communication networks and services, both in terms of a need for SSP and in terms of providing input data for SSP. In other words, an adaptation of services in order to allow the application of SSP will also be required.

The rise of social networks such as Facebook and blogging services such as Google Blog and Twitter is changing the way people communicate. We believe that the future of SSP in telecommunication services mainly depends on the added value it can offer to these new services. The analysis of social signals is not a stand-alone service but can only boost the usage and the quality of user experience of other services.

7.4 Social signals and their utilization

As already indicated, an efficient utilization of SSP in the field of telecommunications depends on user adaptation services that can use this data to add value to service and to improve a user's quality of experience. As a well-adopted technology of user adaptation, recommender systems (RS) [12] are present in many real applications of multimedia content item recommendation such as Amazon books and several music and movie-on-demand services. RS have most of the advantages and drawbacks of user-centric services and are a good example for future development considerations.

RS will benefit from user-related data extracted using SSP. A well-known cold start problem [12] occurs when new users who need help the most do not receive such help as their model (based on their history of service usage) is not available. The answer to this problem is a collaborative recommendation system where user models of similar users are used to help new users. More accurate usage-related data will allow better similar user model selection and directly boost the quality of recommendations. We believe that similar reasoning is valid for most procedures of telecommunication service adaptation to users.

7.5. Integration of communication systems of key importance

The efficient application of SSP in telecommunication services can be achieved only if a large number of users are using and contributing to the same set of communication services. These services should access and share user-related data, including SSP results. Therefore the integration of (tele-)communication systems and services plays a key role here. Users should have one electronic identity and user models should be shared among applications and services despite the risks of personal data abuse involved.

8. Conclusion

We have presented the emerging field of social signal processing (SSP) in the context of telecommunication services. Several aspects of SSP were briefly described and discussed. The state of the art technology mostly focuses on video data acquisition and analysis of social signals of telecommunication service users. We believe that SSP can add value to several existing services and provide grounds to introduce new ones. Besides, it can improve the accessibility of various services to new user groups and allow usage in less intrusive and more comfortable ways for users of communication devices.

More reliable user-related data would also boost the usability of segmentation and modeling of telecommunication service users' techniques such as user retention and churn prediction, targeted advertising and new service creation.

New communication capacities have made completely new services such as virtual life avatars possible. There are many reasons why users prefer a virtual identity over a real one. Among others these are easy ways of experimenting with social scenarios, defining one's past and present social skills and features, living multiple virtual identities, etc. As has often been the case in the field of communication, real applications arose from experiments which resulted from the mere curiosity of their creators at the beginning. It seems obvious that modern communication services involve social components to a higher and growing extent and that automatic understanding of human social signals achieved through SSP will be an essential part of this.

Literature

[1] S. van der Meer, S, Steglich, S. Arbanovski, User-Centric Communications, 2001.

[2] A. Karapantelakis, G. Maguire, Utilizing Social Context for Providing Personalized Services to Mobile Users. In "5th European Conference on Smart Sensing and Context (EuroSSC 2010", pp 28–41, 2010.

[3] SSPNet. Web page Social Signal Processing. http://sspnet.eu/, 16.7. 2011.

[4] M. Tkalčič, U. Burnik, A. Košir, Using affective parameters in a content-based recommender system for images. User model. User-adapt. interact., 2010, vol. 20, no. 4, pp. 279-311.

[5] M. Buchanan, The science of subtle signals, Strategy+Business 482007, pp 68–77.

[6] A. Pentland. Social Signal Processing. IEEE Signal Processing Magazine, 24(4):108-111, 2007.

[7] A. Vinciarelli, M. Pantic, in H. Bourlard, A. Pentland. Social Signal Processing: State-of-the-Art and Future Perspectives of an Emerging Domain. Communications, pp 1061-1070, 2008.
[8] N.M. Freris, H. Kowshik, and P.R. Kumar. Fundamentals of large sensor networks: Connectivity, capacity, clocks, and computation. Proceedings of the IEEE, 98(11): 1828–1846, 2010.
[9] K. Henricksen, J. Indulska, A. Rakotonirainy, Modeling Context Information in Pervasive Computing Systems, Lecture Notes In Computer Science, pp 167-180, 2002.
[10] F. Eyben, M. Wöllmer, B. Schuller, openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor, Proc. ACM Multimedia (MM), ACM, Florence, Italy, ISBN 978-1-60558-933-6, pp. 1459-1462, 25.-29.10.2010
[11] W. Dongrui, T.D. Parsons, E. Mower, S.Narayanan, Speech emotion estimation in 3D space, 2010 IEEE International Conference on Multimedia and Expo (ICME), p. 737-742, 2010.
[12] M. Pogačnik, J. Tasič, M. Meža, A. Košir, Personal content recommender based on a hierarchical user model for the selection of TV programmes. User model. User-adapt. interact., vol. 15, no. 5, pp 425-457, 2005.
[13] T. H. Thi, J. Zhang, in L. Cheng. Human Action Recognition and Localization in Video Using Structured Learning of Local Space-Time Features. 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp 204-211, 2010.
[14] I. Laptev. On Space-Time Interest Points. International Journal of Computer Vision, 64(2/3): 107-123, 2005.
[15] J. C. S. Jacques, S. R. Musse, and C. R. Jung, Crowd Analysis Using Computer Vision Techniques, IEEE Signal Processing Magazine, Vol. 66, Sep. 2010.
[16] V. Saligrama, J. Konrad, in P.-M. Jodoin. Video Anomaly Identification. IEEE Signal Processing Magazine, 27(5):18-33, 2010.
[17] D. Kosmopoulos, S. in P. Chatzis. Robust Visual Behavior Recognition. IEEE Signal Processing Magazine, 27(5):34-45, 2010.
[18] H. J. Seo in P. Milanfar. Action Recognition from One Example. IEEE transactions on pattern analysis and machine intelligence, 33(5):867-882, 2011.
[19] I. Saleemi, K. Shafique, in M. Shah. Probabilistic modeling of scene dynamics for applications in visual surveillance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(8):1472-1485, 2009.
[20] O. Javed, K. Shafique, in M. Shah. Automated Visual Survellance in Realistic Scenarios. IEEE Multimedia, pp 30-39, Jan.-Mar. 2007.
[21] HOHA. Dataset web page HOHA. http://www.irisa.fr/vista/Equipe/People/Laptev/download.html, 15.7. 2011.
[22] KTH. Dataset web page KTH. http://www.nada.kth.se/cvap/actions/, 15.7. 2011.

PROCESIRANJE SOCIJALNIH SIGNALA
U TELEKOMUNIKACIONIM USLUGAMA

Andrej Košir