Cognition-based Networks: Applying Cognitive Science to Multimedia Wireless Networking

Leonardo Badia*, Daniele Munaretto*, Alberto Testolin†, Andrea Zanella*, Marco Zorzi†, Michele Zorzi*

∗ Department of Information Engineering, University of Padova, Italy

† Department of General Psychology, University of Padova, Italy

E-mail: {*firstname.lastname*}@unipd.it

Abstract—Several techniques for wireless networking, such as opportunistic spectrum access, or self-healing networks, may be seen as using a form of cognition, meaning that they mimic reasoning processes of intelligent beings. We propose to expand this cognition-based process by exploiting the parallel processing power of the infrastructure, so as to go beyond cognition as is meant by these approaches. We leverage novel approaches, taken from cognitive science and artificial intelligence, involving not only supervised but also unsupervised learning, and we envision their application to systems for video over wireless. The transmission of multimedia content, and its adaptation to the condition of the communication infrastructure, i.e., the wireless channel or the content delivery network, are envisioned as particularly critical steps for the development of latest generation mobile networks. For this scenario, we propose and evaluate a video classifier based on a Restricted Boltzmann Machine that tries to extract abstract features of videos from the analysis of the sizes of a few coded frames. These features can then be exploited by the communication network itself to optimize video transmission based on its content.

Index Terms—QoE; video delivery; SSIM; resource management; video admission control; feature extraction; generative models.

I. Introduction

Wireless communication networks are becoming more and more complex. Communication over unused TV bands exploiting the so-called white spaces through cognitive radios [1], self-organizing ad hoc networks for temporary events and emergency applications [2], or the envisioned next generation cellular systems [3], all provide clear evidence of this trend. As the consumers demands and their dependence on wireless services, and correspondingly the complexity of networks, grow, there is a need to efficiently manage the scarce resources without overburdening the network control and the end users. Current policy-based approaches are wasteful and will not be able to deal with the heterogeneous wireless networks of the future. Applying cognition is a way to deal with the challenges of future wireless networks. In different forms, cognition is applied in several networks today, e.g., from cognitive radios to biologically inspired networking solutions [4]–[7].

However, we argue that, in order to efficiently use the wireless resources across a network, it is critical to consider all the players in a wireless ecosystem in a more coherent manner. Thus, we propose the new concept of cognition-based network, which is more than a collection of smart or cognitive radios and instead includes a cognitive infrastructure for learning, management/decision, and dissemination of information.

Our research approach focuses on deconstructing cognition into the following core functionalities, that are part of the so-called cognition cycle proposed in [5]: 1) situational awareness in the form of observation and sensing ability, 2) learning from new data and situations, and 3) recollection and exploitation of past experiences and situations to plan, decide and act. We will consider both supervised and unsupervised learning techniques, and will focus on several system concepts: (i) an ideal, centralized, full-knowledge system as an optimistic benchmark; (ii) a practical system in which each node individually runs model-based optimization algorithms using local information; (iii) a collaborative system in which nodes exchange information with their neighbors, and where network-level optimization might be seen as an emergent property of a community of wireless devices. The application of these approaches enables the derivation of a wireless ecosystem with more advanced capabilities, i.e., able to identify operation paradigms and learn from new data, so as to make autonomous decisions and act similar to what a human being would do.

As a possible application of this approach, we will show how machine learning techniques can be exploited to evaluate efficient video encoding rates to transmit video content over wireless. In particular, we consider a set of several video flows, each encoding with different compression rates. The quality experienced by the end user when receiving each flow under a given compression rate can be estimated quite accurately by means of objective evaluations such as the self-similarity (SSIM) index. In this way, it can be shown that different encoding rates do not always result in the same quality of experience for the end user, mostly depending on the actual content of the video. This poses a critical challenge in that, ideally, the resource manager would like to compress the video down to a rate that significantly reduces the amount of data without decreasing the quality perceived by the end user. However, such a choice varies from flow to flow.

In this context, we observed that the quality experienced by the users can be significantly well estimated without actually inspecting the video content, but just using the sizes of the video frames, which are easy to derive directly from the encoding process. Also, we noted that temporal and spatial correlation properties of the video flows lead to easily predictable structures in terms of frame size. Thus, a machine learning approach appears to be quite useful to learn, and in certain cases even predict, the end quality experienced by the user under a given compression level of the video flow.

In this paper we present our vision of "cognition-based

networks" as an evolution of the cognitive paradigm, and we discuss its application to the optimization of multimedia transmission over wireless networks. We then present some preliminary results that support our approach, showing that machine learning techniques can indeed be exploited to classify video sequences and infer the end quality of a video flow from the size of just a few frames. This opens new perspectives of context aware wireless networking that can be used to optimize radio resource management yielding a better quality of experience for the wireless users.

The rest of this paper is organized as follows. In Section II, we describe general related work on cognition-based networks. Section III details models for video transmission over wireless networks and discusses the need for cross-layer optimization between the source coding rate selection and the conditions of the communication infrastructure. Section IV presents our proposed methodology, which makes use of a Restricted Boltzmann Machine to infer a forecast of the resulting video quality from a sample of the initial video frames of the flow. Some numerical results are presented and discussed in Section V, and finally Section VI concludes the paper and highlights possible future developments, where cognitive resource management within the communication network is applied to other parts of the protocol stack.

II. COGNITION-BASED NETWORKS

Advanced systems such as heterogeneous networks or wireless platforms for emergency applications are becoming more and more complex and will be hard to manage and optimize without the ability to self-organize and learn. The goal is to develop systems that use data-driven and model-based learning as the basis of self-organization and optimal resource utilization in wireless networks.

Cognition as a way to deal with the challenges of future wireless networks has been suggested in various forms, including cognitive radios [5], as well as self-organizing and biologically-inspired networks [6], [7]. Adaptability at the physical layer has been extended to communication networks in a new paradigm called cognitive radio network [4], where the spectrum owned by primary users is shared by (secondary) cognitive radios, which should communicate without harming the primary users. These works still adopt a rather traditional view of communications, and do not fully address and/or exploit the essence of cognition, which includes intelligent observation, learning, decision-making, and emergent and collaborative behaviors, which has been considered only by very few studies. Among these, [8], [9] are worth mentioning, in that they propose such an approach to be applied to the whole protocol stack, not only to dynamic spectrum access.

Following these steps, we advocate to employ cognition across the protocol stack in a wireless networking ecosystem for achieving autonomous network operation and efficient resource usage. In a *cognition-based* network, each node both exploits local information to achieve its goals, and shares it with its neighbors to promote collaborative intelligence and emergent behaviors. Thus, nodes are able to adapt to the environment conditions, using a collaborative framework based on the concepts of collective or group intelligence [7], used by most living organisms, from bacteria to humans.

Models of learning and behavior have been developed by cognitive scientists over many years. Recent computational approaches have greatly improved our understanding of how cognition and intelligent behavior in humans and animals emerges. However, it may not be possible to directly apply such models to our own context, e.g., as they may require a prohibitive amount of computational resources that are not available in widespread communications devices, such as smartphones or even sensors. Therefore, some deep rethinking may be needed in order to be able to apply current knowledge on cognitive models to our scenario of interest. In addition, cognitive science itself is a very dynamic field, and novel and more powerful models are recently being proposed that may be suitable for application in the area of wireless system resource management. Examples of recent approaches that may be taken as a starting point include learning of generative models, where the "learner" estimates a probabilistic model of the system, and evolutionary computation [10].

We aim for a network architecture in which communication nodes perform sensing and data collection as well as dissemination of appropriate information. Furthermore, we investigate a novel cognition-based model to represent and optimize the protocol stack, using machine learning techniques.

The building blocks for such an approach are as follows.

Network architecture — Future systems are likely to have positional and temporal awareness as well, so that the sensing data can be spatio-temporally tagged. As an enabler for the collection of real system data, this architecture is critical to integrate cognition across the network [8] and must enable network-wide observation and sensing (beyond spectrum occupancy, and including quantities such as traffic, channel statistics, error events, interference) as well as data collection from various layers of the protocol stack.

Machine learning aided network optimization — An important component is the ability of the cognitive entity to adapt its operation based on past situations that are likely to repeat again, also including collaborative behaviors. Proper statistical principles can be used to develop generative models that discover distributed and non-trivial representations of the input. These models will then be used to design techniques in which nodes are considered as agents in a collaborative framework, where partial coordination and a shared objective will lead to emergent behavior and collective intelligence.

Practical evaluations — Several applications can serve as proofs of concept and also to derive novel cognition-based networking techniques, depending on the problem at hand. For example, in the following we will report our findings on the application of a cognition-based approach to the transmission of video content over a wireless network, involving management of the encoding rate of video and the automatic recognition of the related quality parameters.

III. VIDEO OVER WIRELESS

The problem of video over wireless is very relevant, since in 2012 the mobile video traffic already exceeded 50% of

the total data traffic in the Internet, and forecasts foresee a further increment to over 66% of the total mobile data traffic by the end of 2017 [11]. Therefore, supporting high quality video services by means of femto-cells and WiFi hotspots is challenging and requires to adjust the video encoding rates according to the available transmission resources.

In this paper, as in [12], [13], we consider the transmission of video content encoded through Scalable Video Coding (SVC), an extension of Advanced Video Coding (AVC) mechanisms such as H.264 [14]. The general idea of this scheme is to consider the bitstream transporting the video as divided into layers, one Base Layer (BL) and multiple Enhancement Layers (ELs). The BL consists of independently encoded video frames and by itself is only able to provide a low-quality transmission of the video content. The ELs are incrementally encoded based on the BL, and therefore require much lower transmission capacity, but cannot be decoded by themselves. However, they are able to provide further enhancements to the quality of the end user.

Thus, depending on the state of the communication infrastructure (communication channel, communication network), it may be convenient to transmit just the BL, or the ELs as well. Also, retransmission policies in case of errors can be variable [15]. Generally speaking, although there are analytical frameworks to determine the solution to this optimization problem and its inherent trade-offs, the formulation itself is made challenging by the fact that the required interaction is not simply a generic "cross-layer" exchange, but indeed it encompasses the entire protocol stack, from the application to the physical layer, also traversing transport, network, and medium access layers.

Instead of deriving a theoretical abstract framework, we specifically focus on an H.264-AVC format as reported in the Joint Scalable Video Model software of [16]. This means that we consider $N_R = 18$ possible compression levels; such a number results in a granularity, which is fine enough to enable general considerations. Thus, the approaches derived in the following can also be easily extended to higher values of N_R. The N_R compression levels are denoted with increasing index from 1, corresponding to the flow with best quality (and lightest compression) to N_R, where quality is worst and compression is heaviest.

We adopt a rather simple top-down approach, where we take into account that using high encoding rates will result in intense bandwidth usage but will determine a good quality for the end user; conversely, choosing lower encoding rates will cause to transmit less data (hence saving bandwidth, or enabling the transmission of more users on the same channel), but also result in a lower, sometimes unacceptable, quality. We denote with $R_v(c)$ the data rate of the video when the compression level is c, with $c = 1, \ldots, N_R$. An interesting parameter is the rate scaling factor (RSF) for a compression level c, denoted as $\rho_v(c)$, and defined as the logarithm of the normalized video rate [12]:

$$\rho_v(c) = \log\left(\frac{R_v(c)}{R_v(1)}\right) \qquad (1)$$

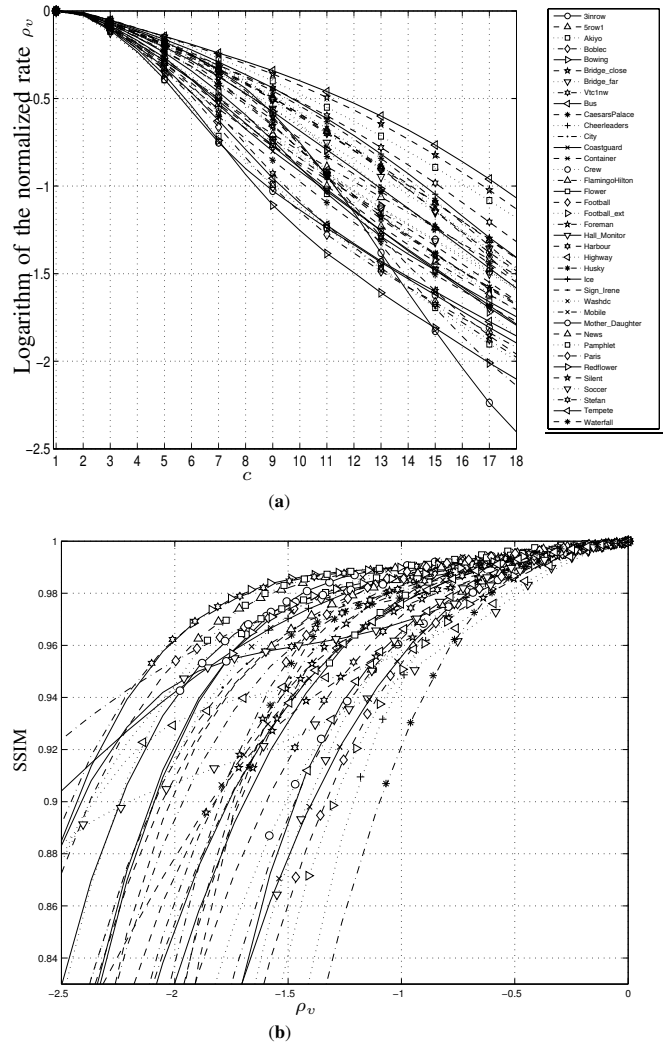where the normalization is made with reference to the highest

Fig. 1. (a) Logarithm of the normalized rate, ρ_v, *versus* compression level c for different video clips. (b) SSIM *versus* ρ_v.

video rate (lowest compression), i.e., level 1.

The reason for taking the logarithm is to empirically evaluate the applicability of a Weber-Fechner-like law [17] to the scaling procedure. Indeed, such a law is relevant in terms of both determining the minimum noticeable difference in quality (Weber law) and also tuning the experienced quality depending on the logarithm of the stimulus (Fechner law), which is the RSF in our case.

To substantiate this argument, we collected a set of $N_V = 38$ test video clips, encoded in CIF format, and for each clip v we generated N_R different H.264 encoded versions with progressively decreasing rates, $R_v(c)$, with $c = 1, \ldots, N_R$. All the flows have a periodic structure of 16 frames, that in the H.264 setup constitute a Group of Pictures (GoP), a video unit carrying both base and enhancement frames that is periodically repeated. This means that the structure of the GoP can be thought of as consisting of an independently encoded frame (type I for H.264 videos), i.e., a BL frame, and 15 incrementally encoded frames (type P for H.264 videos) belonging to ELs. The actual size of a frame, however, depends on both its kind (i.e., on the specific layer of the incremental hierarchy it belongs to) and the specific video content.

The scaling law of the video rate $R_v(c)$ with the compression level c depends on the video content. As visible from Fig. 1(a), ρ_v decreases when the compression is increased, in an approximately linear fashion, albeit with a great variability from one video to another. We observe that the relative rate scaling with the quantization level is generally more pronounced for slowly varying videos, such as *Akiyo* or *Bowing*, while more dynamical videos, as *Football*, are less amenable to compression. However, this intuitive relation between the perceived dynamics of a video content and its potential for compression is not always verified. This is the case, for instance, of the *Bridge_close* video clip that, while appearing rather static at visual inspection, is still quite resistant to compression, because of the continuous variation of the light reflections on the water surface. It is therefore, of great interest to develop automatic methods to estimate the rate scaling law of a video based on its content.

However, it is also important to remark that the decrease in the compression rate does not always correspond to an equivalent increase in the quality of experience for the end user. The assessment of a "good," "acceptable," or "bad" quality is heavily varying depending on the video content. Nonetheless, there exist objective indicators of video quality. The Peak Signal to Noise Ratio (PSNR), which has been widely used in the literature until a few years ago as an objective measure of image and video quality, is now generally discredited as video QoS metric because the pixel-by-pixel comparison of two images may fail to capture the visual and structural similarities of the images. Therefore, the research community has been searching for other objective indicators of video quality, such as the widely used structural-similarity index (SSIM) [18], which provides a better indication of the quality of an image (or a video) as *perceived* by a human observer.

The SSIM index can be used to characterize the degradation of the video quality when increasing the compression ratio of a video encoder. For instance, comparing a video clip encoded at a certain rate with the same video with full quality, we can determine the SSIM value for each video frame and, then, consider the average. In this way, we get a real value in the $[0, 1]$ interval: the larger the value, the better the video quality.

If we now plot the mean SSIM versus the normalized log-rate ρ_v, we obtain the graph in Fig. 1(b), where we see that also the scaling of the SSIM with the source rate depends on the characteristics of v. From the application point of view, knowing the SSIM characteristics of videos sharing a common link would make it possible to allot transmission resources in a quality-aware manner, as done in [12]. Unfortunately, the extrapolation of the SSIM characteristic of a video clip by standard frame-comparison methods is time and computational demanding.

However, our investigations when plotting Fig. 1 seem to suggest that there is a strong dependence between the size of the frames in the video and the resulting SSIM. The underlying connection might be that, regardless of how it is actually compressed, the quality experienced by a user for a specific video (and hence, the SSIM) is strongly related to the dynamic evolution of the content. Small-size frames denote either a heavily compressed video or an irrelevant or uninteresting content, hence the perceived quality is low, anyway. Thus, we should be able to infer some of the SSIM characteristics of a video from higher-layer features, such as the size of the H.264 compressed frames when represented as a GoP. Moreover, indicators such as the frame size (at least for a frame in the same position within the GoP structure, which is periodically repeated) are rather easy to predict, since they are slowly varying for the same flow, mostly depending on the characteristics of the video flow (and its represented content) as a whole.

In the following section, we will exploit this correlation to outline an approach based on a generative model that is able to identify the quality characteristics, i.e., the SSIM, from the frame size of the flow. Such a machine learning technique can also be used in a predictive fashion, i.e., to identify the evolution in the user experience for a given flow, also depending on the video compression factor, and can serve to guide the management of network resources that enable video transmission.

IV. RESTRICTED BOLTZMANN MACHINES

A key component of the proposed system is represented by an unsupervised learning module, which we implemented using a Restricted Boltzmann Machine (RBM) [19]. An RBM is a stochastic neural network whose structure, represented in Fig. 2, can be interpreted as an undirected graph including two layers of neurons. This means that a set of observable variables (so-called "visible units") is fully connected to a set of latent variables (so-called "hidden units"). Each edge in the graph has an associated weight, which represents the strength of the connection between the linked variables. The term "restricted" means that the graph is bipartite, i.e., there are no connections among units of the same layer.
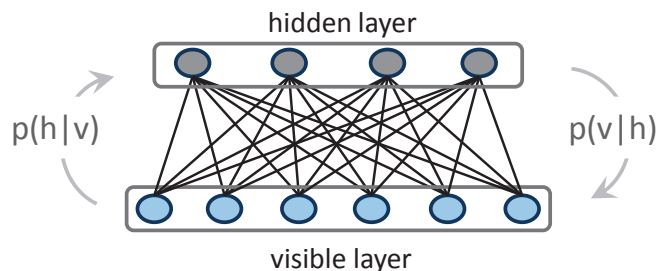

Fig. 2. The structure of a Restricted Boltzmann Machine.

Our proposal is to use an RBM to first extract a useful set of features from the encoded video streams, which enable the automatic assessment of the videos according to some high-level characteristics, such as the dynamics of their content. After each video has been represented using this set of abstract features, a subsequent supervised classification might be easily carried out even by a simple linear classifier, thereby making it possible to effectively assign a label to each GoP in agreement with its SSIM characteristics. The classification process is therefore decomposed into an initial, unsupervised phase and a subsequent supervised phase. The

rationale behind this approach is given by recent advances in learning generative models [20], [21], whose aim is to discover the latent causes that govern the distribution of a set of data by trying to accurately reconstruct the training patterns through maximum likelihood learning. In particular, the introduction of the contrastive divergence algorithm [22] makes it possible to efficiently train very powerful non-linear models, like RBMs, which were usually considered computationally intractable.

Once the generative model extracted meaningful latent features from the data, this high-level, abstract representation can be used to perform different supervised tasks. In the machine learning literature, this process is usually referred to as *transfer learning* [23], because the knowledge extracted from the data can be readily used in many different classification scenarios. The complete weight matrix represents the parameters of the model, which are tuned during the learning phase according to the contrastive divergence algorithm. Intuitively, the visible units are clamped to a particular instance of the data, and the goal of the model is to accurately reconstruct their value by discovering useful features shared among all the patterns in the training set. In our case, the values of visible units correspond to the sizes of the sequence of frames contained in a GoP, and the goal of the hidden units is to discover their latent structure. Once the RBM has learned a set of latent features that characterize the training data, we can use this high-level representation to train a supervised module. The idea is that the non-linear mapping into the feature space captures the main factors of variations in the input data, thereby making it possible to separate the relevant classes even using a linear classifier. This approach is commonly employed in other machine learning techniques, like Support Vector Machines [24], in which the mapping is usually carried out by a kernel function. The use of a linear classifier to assess the degree of explicitness of internal representations in hierarchical generative models has also been shown useful in cognitive modeling approaches [25].

V. NUMERICAL RESULTS AND DISCUSSION

In this section we present a selection of the results we obtained by applying the cognition-based approach to video classification. The basic idea is to design an RBM that is capable of sorting out specific features of a video content from the analysis of application-related data, such as the sequence of frame sizes in a GoP. During the unsupervised training process, the RBM "learns" an internal representation of the input patterns, by extracting some features that may be difficult to capture in the raw data. The RBM maps the input vectors in an abstract structured space, which makes it easier to perform a supervised classification of the input, starting from those abstract representations. Therefore, placing a simple linear classifier on top of the hidden layer of the trained RBM, and performing a supervised classification task, we can get better results than applying the classifier directly to the raw data.

To prove the effectiveness of such an approach, we designed an RBM with 32 input units, corresponding to the (normalized) size of the 16 frames of a GoP, coded at two different quantization levels, namely $c = 1$ (full quality), and $c = 9$
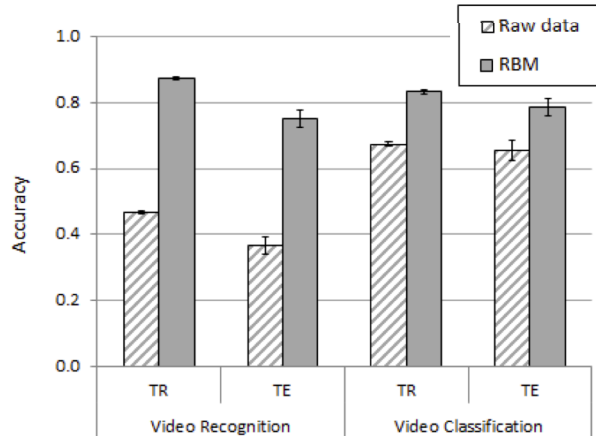

Fig. 3. Video recognizing accuracy for our proposed approach.

(intermediate quality). The number of units in the hidden layer has been varied from 50 to 200, and finally set to 70 units, which offers a good compromise between the complexity of the model and the accuracy of the results. Therefore, the dataset consists of the vectors of frame sizes for each GoP of all 38 test video clips, coded at $c = 1$ and $c = 9$. The dataset is hence split into Training Set (TR) and Test Set (TE), the first used to train the RBM and the linear classifier, and the second to test the performance of the trained network.

We applied the cognition-based approach to perform three different tasks of increasing complexity: i) assigning a random GoP to the video it belongs to; ii) classifying the GoPs into four SSIM-based classes; iii) estimating the SSIM characteristic of a video. For the second task, we considered the four classes defined in [12], based on the value of ρ for which the SSIM of each video crosses the threshold $F^* = 0.95$ that discriminates between "fair" and "good" perceived quality levels.

In Fig. 3 we report the accuracy obtained in video recognition (left) and video classification (right). In order to assess the advantage offered by the RBM approach, we compare the results obtained by training the linear classifier on the hidden layer of the RBM (shaded columns) with those obtained by applying the classifier directly to the raw data, i.e., the GoP frame size vectors (textured columns). The variation of the classification accuracy in different tests is represented by the error bar on top of each column.

At a first glance, we note that the accuracy in both video recognition and classification achieved by the RBM approach is significantly larger than that obtained by using the classifier on raw data, thus confirming the capability of the RBM to abstract the input data into a better structured representation. The accuracy gap between the two approaches is more marked in video recognition, for inputs taken from both TR and TE. Quite interestingly, we note that the RBM approach achieves comparable accuracy in both video recognition and classification tasks, while the other approach performs better in the second task. The reason is that the input space formed by the frame sizes is not separable in clusters associated to the different videos by means of simple, planar functions, while

the alternative representation provided by the hidden units of the RBM is more amenable to such a linear discrimination. Instead, the raw data seem to be sufficiently structured to permit a (fairly accurate) classification of the videos in the four SSIM-based macro classes, which means that the quality characterization of a video sequence is indeed strictly related with the distribution of the frame size in its GoPs.

Deeper insights on the latter aspect can be gained by comparing the performance of the two approaches in the third task, i.e., the estimation of the SSIM characteristic of a video clip from the size of the frames in a GoP. This analysis, which is presented in detail in [13], confirms that the internal representation learned by the RBM is indeed capable of capturing critical features of the data and can be used to get a fairly good estimate of the SSIM characteristics of a video from the size of the frames of a few GoPs. As shown in [13], such an estimate is sufficiently accurate to be used by video admission control and resource allocation algorithms that aim at maximizing the quality of experience of the end users.

VI. Conclusions and Future Work

In this paper, we presented a cognition-based framework for network optimization and discussed its possible application to the context of multimedia delivery over wireless networks. The proposed approach combines unsupervised and supervised machine learning techniques to supply wireless systems with much deeper awareness about their own operations and the environment, so that they can learn relationships among network parameters, protocols, and context. Consequently, they will plan and make decisions in order to achieve local, end-to-end, and network-wide performance goals through efficient resource management and parameter optimization.

As a proof of concept, we showed that the combination of an unsupervised learning model (an RBM) with a linear classifier is much more effective in extracting content-related features of video sequences than a simpler approach that acts directly upon the raw data, i.e., the size of the encoded video frames. This information can be then exploited in the design of resource management algorithms that keep into account the quality-of-experience of the end users [13].

As a future extension, a promising research direction would be to combine multiple RBMs to obtain a deep learning system [26], which could make it possible to extract higher-order statistical information by building a hierarchical generative model of the training data. Furthermore, we will investigate how these techniques can be applied to other optimization problems (base station selection, protocol parameter optimizations) and we will implement the resulting procedures in a testbed that is presently under development.

Acknowledgment

References

[1] Y. Yuan, P. Bahl, R. Chandra, P.A.. Chou, J.I. Ferrell, T. Moscibroda, and Y. Wu, "KNOWS: Cognitive radio networks over white spaces," *Proc. IEEE DySPAN*, pp. 416–427, 2007.

[2] P. Pawelczak, R. Venkatesha Prasad, L. Xia, and I.G. Niemegeers, "Cognitive radio emergency networks-requirements and design," *Proc. IEEE DySPAN*, pp. 601–606, 2005.

[3] L. Anchora, L. Canzian, L. Badia, and M. Zorzi, "A characterization of resource allocation in LTE systems aimed at game theoretical approaches," *Proc. IEEE CAMAD*, 2010.

[4] E. Hossain, D. Niyato, and Z. Han, Dynamic spectrum access and management in cognitive radio networks, Cambridge Univ. Press, 2009.

[5] J. Mitola III, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, KTH, Sweden, May 2000.

[6] L. Badia, A. Botta, and L. Lenzini, "A genetic approach to joint routing and link scheduling for wireless mesh networks, *Ad Hoc Networks*, vol. 7, no. 4, pp. 654–664, June 2009.

[7] E. Bonabeau, M. Dorigo, and G. Theraulaz. Swarm intelligence: From natural to artificial intelligence. Oxford University Press, 1999.

[8] B.S. Manoj, R. Rao, and M. Zorzi, "Cognet: a cognitive complete knowledge network system," *IEEE Wireless Communications*, vol. 15, no. 6, pp. 81–88, Dec. 2008.

[9] F.H.P. Fitzek and M. Katz. Cognitive Wireless Networks, Springer, 2007.

[10] T. Hope, I. Stoianov, and M. Zorzi, "Through neural stimulation to behavior manipulation: A novel method for analyzing dynamical cognitive models," *Cognitive Science*, vol. 34, pp. 406-433, 2010.

[11] Cisco White Paper, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017," February 2013.

[12] M. Zanforlin, D. Munaretto, A. Zanella, M. Zorzi, "SSIM-based video admission control and resource allocation algorithms," *Proc. IEEE WiVid'14*, May 12-16, 2014, Hammamet, Tunisia.

[13] A. Testolin, M. Zanforlin, M. De Filippo De Grazia, D. Munaretto, A. Zanella, M. Zorzi, M. Zorzi, "A Machine Learning Approach to QoE-based Video Admission Control and Resource Allocation in Wireless Systems," *Proc. IEEE IFIP Med-Hoc-Net 2014*, June 2-4, 2014, Piran, Slovenia.

[14] F.H.P. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation," *IEEE Network*, vol. 15, no. 6, pp. 40–54, Nov/Dec 2001.

[15] L. Badia, N. Baldo, M. Levorato, and M. Zorzi, "A Markov framework for error control techniques based on selective retransmission in video transmission over wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 488–500, Apr. 2010.

[16] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H. 264/AVC standard," *IEEE Trans. Circ. Syst. for Video Techn.*, vol. 17, no. 9, pp. 1103–1120, 2007.

[17] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, "The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment," *Proc. IEEE ICC*, Cape Town, South Africa, 2010.

[18] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600 – 612, Apr. 2004.

[19] D. Ackley, G.E. Hinton, and T.J. Sejnowski, "A learning algorithm for Boltzmann machines", *Cognitive Science*, vol. 9, pp. 147-169, 1995.

[20] P. Dayan, G.E. Hinton, M. Radford, and R.S. Zemel, "The Helmoltz machine", *Neural Computation*, vol. 7, pp. 889-904, 1985.

[21] G.E. Hinton, T.J. Sejnowski. Unsupervised learning: Foundations of neural computation. MIT Press, 1999.

[22] G.E. Hinton, "Training products of experts by minimizing contrastive divergence", *Neural Computation*, vol. 14, pp. 1771-1800, 2002.

[23] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning", *International Conference on Machine Learning*, vol. 7, pp. 1-20, 2011.

[24] V.N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–99, 1999.

[25] M. Zorzi, A. Testolin, and I. Stoianov, "Modeling language and cognition with deep unsupervised learning: a tutorial overview." *Frontiers in Psychology*, vol. 4, article 515, 2013.

[26] G.E. Hinton, and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks." *Science*, vol. 313, pp. 504-507, 2006.