The 29th Annual Workshop of the Swedish Artificial Intelligence Society (SAIS)

2–3 June 2016, Malmö, Sweden

Editor
Paul Davidsson

Linköping Electronic Conference Proceedings, No. 129
Linköping University Electronic Press
Linköping, Sweden, 2016
ISSN: 1650-3686
eISSN: 1650-3740
ISBN: 978-91-7685-720-5
URL: http://www.ep.liu.se/ecp/contents.asp?issue=129

Conference organization

Program chair
Paul Davidsson

Program committee
Eva Blomqvist
Anton Borg
Paul Davidsson
Christos Dimitrakakis
Göran Falkman
John Folkesson
Johan Hagelbäck
Fredrik Heintz
Anders Holst
Magnus Johnsson
Niklas Lavesson
Jacek Malec
Slawomir Nowaczyk
Panagiotis Papapetrou
Federico Pecora

Local organization
Paul Davidsson
Karolina Rosenqvist
Richard Topgaard

Table of Contents

Non-Linear Hyperspectral Subspace Mapping using Stacked Autoencoder

Niclas Wadströmer and David Gustafsson
Swedish Defence Research Agency (FOI), Sweden
email: {niclas.wadstromer, david.gustafsson}@foi.se

Abstract

Stacked Auto-Encoder (SAE) is a rather new machine learning approach which utilize unlabelled training data to learn a deep hierarchical representation of features. SAE:s can be used to learn a feature representation that preserve key information of the features, but has a lower dimensionality than the original feature space. The learnt representation is a non-linear transformation that maps the original features to a space of lower dimensionality. Hyperspectral data are high dimensional while the information conveyed by the data about the scene can be represented in a space of considerably lower dimensionality. Transformation of the hyperspectral data into a representation in a space of lower dimensionality which preserve the most important information is crucial in many applications. We show how unlabelled hyperspectral signatures can be used to train a SAE. The focus for analysis is what type of spectral information is preserved in the hierarchical SAE representation. Results from hyperspectral images of natural scenes with man-made objects placed in the scene is presented. Example of how SAE:s can be used for anomaly detection, detection of anomalous spectral signatures, is also presented.

1 Introduction

Hyperspectral imaging reveal information about the scene that can not be perceived with a visual camera. The spectrum radiated from each point of the surfaces in the scene is captured in many separate narrow wavelength bands. Usually a visual camera have three wavelength bands capturing red, green and blue colours. A hyperspectral sensor can capture tenths, hundreds and even thousands of wavelength bands. The spectrum for single pixels can be viewed and compared with each other. But an image could have millions of pixels each with hundreds of wavelength bands. Thus an image could be composed of 10^9 values. The information in a hyperspectral image can not easily be comprehended by a human observer. On the other hand is not unlikely that a scene only contains a handful of materials suggesting that maybe the information in the image can be found in a low dimensional space. So how can we find a suitable space which will fit the information without discarding any information? What dimension of the space can we expect to find? What class of mappings between the spaces should be expected to be efficient?

Here we will assume that the hyperspectral data can be found in a subspace of the measurement space. It is also possible that the scene consists of a countable set of different spectrums. In this case can the data be represented by an index number referring to the spectrum found in each pixel.

But for now let us assume that the data is composed of spectrums from a subspace of lower dimensionality.

Many analysis algorithms have a computational complexity which is proportional to the number of dimensions, some have exponential complexity, and are thus costly to work with. If it is possible to map the data to a space of lower dimensionality than less computational resources will be needed.

Deep learning using stacked autoencoders have shown in many examples to give a compressed representation that is useful for many different applications. For example a representation trained for a specific classifier has shown to be useful for classification of classes not at all considered in the original training.

There is also a question of how easy it is to find the important information. It may be possible that the information in the encoded space is more difficult to find than the same information in the uncoded data.

2 Hyperspectral imaging

A hyperspectral image can be viewed as set of spectral measurements $x_i : i = 1, 2, \ldots, N$ where i a spatial index. Each measurement x_i is represented as a M-dimensional vector

$$x_i = [L_i(\lambda_1), \ldots, L_i(\lambda_M)] \tag{1}$$

where $L_i(\lambda_k)$ is the measurement of spectral band λ_k and M is the number of spectral bands. This representation stresses the spectral dimension of the data, while the spatial relations are ignored. As the focus of the paper is spectral dimensionality reduction the spectral vector representation is used.

2.1 Dimensionality Reduction

Hyperspectral images (HSI) contain both spatial and spectral information about the scene at hand. How much and what kind of information the spatial and the spectral dimensions carry depends on the situation. Often it is of interest to determine how much and what information that is carried by the spectral dimensions solely. The strong focus on the spectral dimension in hyperspectral imaging can partly be explained by the strong connection between hyperspectral signatures and material properties in a scene. So, henceforth in this paper we will only consider the spectral dimensions. Thus, a hyperspectral image will be thought of as a large number of pixels, each with a spectrum, and with no spatial relation to any other pixels.

Hyperspectral data represented as vectors, Equation (1), is often of high dimensionality (i.e. the number of sampled spectral bands are many). The spectral bands are typically highly correlated which indicate that the data resides in a space of lower dimensionality than M. The dimensionality of the hyperspectral space is also limited by the number of materials with different spectral signatures. If there is a countable number of materials in the scene and each material has its own unique spectrum then the scene can be represented by a number for each material. In this case the information in the scene is digital. If each material in the scene has its unique spectrum and the radiation varies with the lighting then it is possible to reduce the number of dimensions needed to represent the data, and represent the data in a common low dimensional space.

One way of considering dimensionality reduction which is related to the sensor is to reduced the number of wavelength bands and possible adjust their widths to find a smaller set of spectral bands containing the interesting information about the scene. In dimensionality reduction by sub-band selection the problem is to determine how many and which spectral bands that are required to solve the problem at hand. This kind of dimensionality reduction may influence the construction of the sensor. Fewer spectral bands may mean less complexity in the sensor.

Other kinds of dimensionality reduction requires that all bands are captured and then transformed into a space of lower dimensionality. It may still be of importance to reduce the number of dimensions to reduce complexity of signal analysis which then can be done in a subspace of lower dimensionality. It is not given that analysis of the data in the low dimensional space is less complex. However, many results show that the inner representation actually is meaningful and that the representation is useful and makes for example classifiers are more easily trained on the reduced representation. Linear transformations of the hyperspectral data, such as Principal Component Anlysis (PCA) [4] and Independent Component Analysis (ICA) [8], for dimensionality reduction are frequently used pre-processing methods. Kernel-PCA [5], which is a non-linear extension of the (linear) PCA-transformation, is also a frequently used method.

Dimensionality reduction is an encoding problem meaning that we seek an efficient representation of the data and in this context efficient means few dimensions. Efficient representation could also mean few bits if the information is discrete. Encoding problems compared to classification and regression means that no annotated data is needed for the training. We assume that the data contain the information of interest and nothing else. If there is some kind of noise that can be discarded this can be captured by requiring that the data is reconstructed at least as well as a limit on the given error measure is obtained.

Neural network (NN), and recently deep neural network, has been used for dimensionality reduction of hyperspectral images. Zeng an Trussel [10] used NN to implicitly reduce the dimensionality of hyperspectral images in a classification setting. A data dependent error function - sum of square error (SSE) - was combined with a sparseness criteria on the weight in the NN which penalizes non-zero weights. The NN was trained for classifying hyperspectral signatures and was trained using classified spectral signatures of different materials. The dimensionality reduction was implicit in the training procedure through the sparseness criterion and no explicit low dimensional representation of the spectral signatures was generated.

Chen et al. [3] propose a DL based approach for classification of hyperspectral signatures. Chen et al. uses Stacked Auto Encoder (SAE), as described in Section 3.1, to pre-train the NN using unlabelled training data. The input features to the AE:s are spectral signatures, spatial representation and a joint spectral-spatial representation. The spatial information is represented with a flattened neighbouring region of a PCA transformation in the spectral domain. The SAE is fine tuned using a labelled training set. Chen et al. [2] use a similar approach but instead of using SAE in the unsupervised pre-train step, they use layered Restricted Boltzmann Machines (RBM). The dimensionality reduction, using SAE:s and layered RBM, was implicit and as an pre-processing step for supervised learning of a classifier.

The paper addresses how SAE:s can be used for dimensionality reduction of hyperspectral signatures without explicitly connecting the learnt representation to a specific classification task. What spectral information is preserved?

3 Deep learning and Autoencoder

Deep Learning (DL) is a relatively new approach in pattern recognition that has achieved remarkable results in many applications ([9, 1, 6]). The key concept in DL is to represent the features in a way that improves further analysis. The features is, in DL, represented hierarchically where the different levels in the hierarchy represent different levels of abstraction.

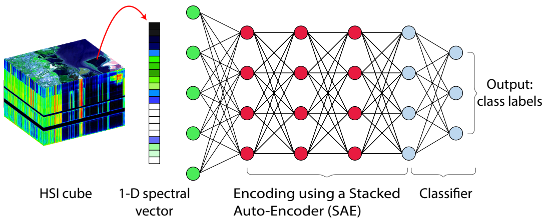


Figure 1: A 1-D spectral feature vector is hierarchically represented - the red nodes - using a Stacked Autoencoder (SAE). The output of the encoding part of the SAE generate a compressed representation of the vector used as features for the classifier. The SAE - red nodes - is optimized using an unlabelled training set (see figure 2), while the classifier require labelled data.The ability represent hyperspectral signatures in a low dimensional space using SAE is investigated.

A hierarchical feature representation can, in principal, be learnt from training data using a multilayer neural network (also called Multi Layer Perceptron (MLP)). The weights in a MLP are adjusted by minimizing an error function, commonly the mean-square error (MSE) - equation 9, over a training set commonly using the Back-Propagation (BP) algorithm. BP calculate the gradient, i.e. the derivative of the error function with respect to the weights, of the MLP and the weights are updated by moving in the negative gradient direction.

It is hard, or even impossible, to train a deep MLP, i.e. a MLP with 2 or more hidden layer, due to the *the vanishing gradient problem*. BP calculate the gradient by propagate the error from the output layer toward the input layer and the derivatives decreases as the error is propagated through the layers in the MLP. The gradient is small in the bottom layers of the MLP and the weights are almost unchanged after updating, which make learning very slow (or impossible).

Instead of training a deep neural network directly it can be trained in two phases:

1. **Pre-training** In this phase unlabelled training data is used to learn a hierarchical rep-

resentation of the data. A layer-wise greedy strategy to learn a latent representation of the unlabelled data is used. Autoencoders (AE) and Restricted Boltzmann Machines (RBM) are two frequently used techniques.

2. **Fine-tuning** The weights obtained from the pre-training phase are used for initialising the weights in the full deep neural network. The hierarchical representation learnt in the pre-training phase is a good starting point for fine tuning the deep neural network using labelled data and the back-propagation algorithm.

3.1 Autoencoder

An AutoEncoder (AE) is neural network trained to reconstruct or reproduce its inputs as its outputs. An AE is composed of two parts; an *encoder* and a *decoder*. The encoder takes an input X and maps it to a hidden (or latent) representation U. The latent representation is often of lower dimensionality which imply that the encoder compress the information in the features. The decoder reconstruct the input X from the latent representation U. An AE is an neural network composed of one or more hidden layers which maps the input features onto itself. An AE can be learnt using the BP algorithm and unlabelled training examples. No labels are required because an AE maps the input onto it self independent of any class labels. In the training an optimal encoder and decoder of the input features though the latent representation is learnt. The latent representation can be viewed as a compression of the features containing the most important information.

Multilayer AEs, called Stacked AutoEncoders (SAEs), are constructed using a greedy layer-wise strategy. An AE is trained using an unlabelled training set and some features X. The trained AE maps the features X of the training set to the learnt latent representation, called, U. U is a representation of lower dimensionality than the original features. An AE is trained using U as features resulting in a latent representation V of even lower dimensionality and so on (See Figure 2).

A deep structure, SAE, is constructed by stacking the greedy layer-wise learnt AEs. The encoding part of the SAE maps the original features through a hierarchical representation to a low dimensional compressed representation.

Let $\{\mathbf{x}_i\}_{i=1}^{N_t}$ be a set of training vectors where each vector is the spectral information from one pixel, $\mathbf{x}_i = (x_1, x_2, ..., x_M)$, each vector describes the spectral intensity of M spectral bands. Let $\{v\}_{i\,i=1}^{N_v}$ be a set of validation vectors. Let $\Phi(x)$ be a non-linear activation function in this case

$$\Phi(x) = 1.7159\tanh(\frac{2}{3}x) \tag{2}$$

A node in the neural network performs the function

$$f_n(\mathbf{x}) = \Phi(w_0 + \sum_{i=1}^{M} w_i x_i) \tag{3}$$

which can also be written

$$f_n(\mathbf{x}) = \Phi(\mathbf{w}\mathbf{x}^\top). \tag{4}$$

A layer in the network is composed of n nodes

$$f(\mathbf{x}) = \Phi(\mathbf{W}\mathbf{x}^\top) \tag{5}$$

where $\mathbf{W}$ is a matrix with one row for each node or output signal and one column for each input signal.

An autoencoder (one layer in a stacked autoencoder) consists of two layers of nodes, a hidden layer and a output layer. The output layer have a linear activation function, thus

$$f(\mathbf{x}) = \mathbf{W_d}\Phi(\mathbf{W}_e\mathbf{x}) \tag{6}$$

where $\mathbf{W}_e$ is the parameters of the encoding layer and $\mathbf{W}_d$ is the parameters of the decoding layer.

Training is done by iteratively adjusting the parameters of the network using backprogagation.

Let e_{SAE} and d_{SAE} be the encoder respective decoder parts of the SAE then reconstruction of a sample x_n is defined as

$$\tilde{x}_n = d_{SAE} \circ e_{SAE}(x_n) \tag{7}$$

where $\circ$ is the function composition operator. The reconstruction residual r_n of a sample x_n is defined as

$$r_n = x_n - \tilde{x}_n \tag{8}$$

and reconstruction error is defined as $e_i = \|r_i\|$. Norms of interest include the $\|x\|_{L_1}$, $\|x\|_{L_2}$ and $\|x\|_{L_\infty} = max(|x_1|, |x_2|, \cdots, |x_m|)$ norms. The reconstruction error e_i of a hyper spectral signature

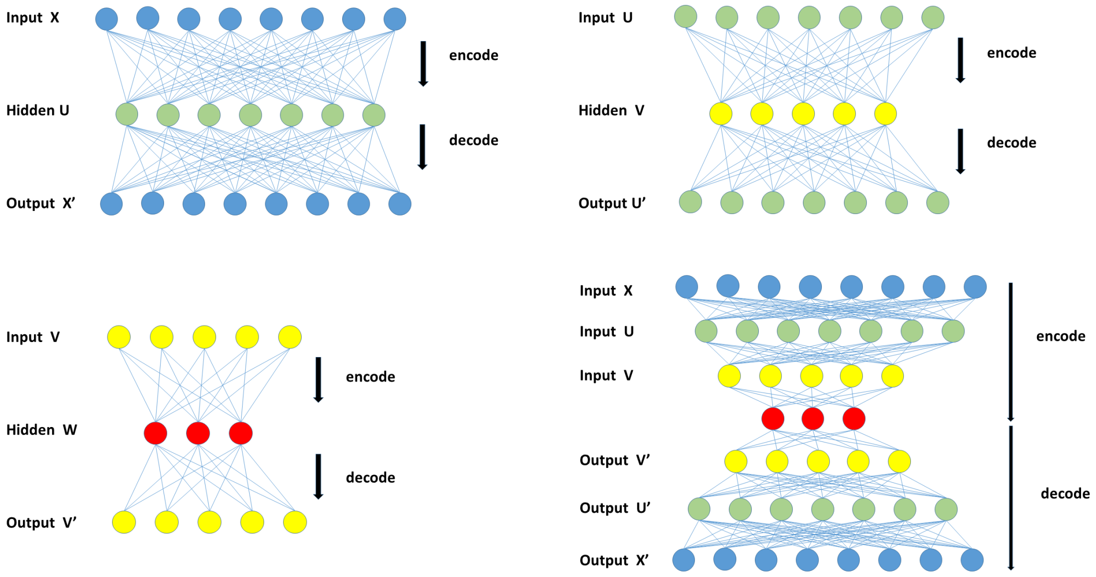


Figure 2: Stacked Autoencoder (SAE) is a greedy strategy to learn a hierarchical feature representation of the data. An AE which maps the features X onto it self though a hidden layer U is learnt. The hidden layer U is a latent representation of lower dimensionality of the features. The latent representation U is used as input to a another AE which learn a latent representation V and so on. A deep structure is composed by stacking the layer-wise trained AEs and is hence called a SAE. A SAE can be used to encode (and decode) original features to any of the dimension the latent representations.

indicate how well it can be approximated in the subspace by the SAE.

The mean-square (reconstruction) error is defined as

$$\mathcal{E} = \frac{1}{N} \sum_{n=1}^{N} \|e_i\|_2^2. \tag{9}$$

where N is the number of samples in training set.

The mean-square reconstruction error of an SAE indicate the performance of the hierarchical representation of the features. The reconstruction error also gives information about the dimensionality of the original data i.e. the number of dimensions which are required for a good reconstruction.

3.2 Dimensionality Reduction using stacked autoencoder

There is a choice of the topology of the neural net, that is the number of hidden layers and the number of neurons in each layer. The topology will determine the number of parameters in the encoder.

There is a relationship between the number of parameters and what the net can be expected to represent, how complex data that can be encoded with the mapping. There is also a relationship between the number of parameters and the amount of training data that is required to be able to determine the parameters. If there are too many parameters then there is a risk that the system will be over trained meaning the the system will learn too much detail of the data. If the system is overtrained the system will not generalise well to coming data from the same source.

There are two important variables, the number of parameters and the reconstruction error.

This is a source coding problem. A small increase in the reconstruction usually means a large improvement in the data rate.

We assume or hope that data can be reduced with respect to the number of dimensions while retaining the information about the scene. Not much is known and thus can be assumed about the space in which the relevant aspects of the scene is represented. We hope that a trained multilevel NN will be able to reconstruct the original data with little distortion.

The output from one layer is the input to the next layer so the output of the first layer should be adapted to be suitable as input to the next layer.

A proper choice of the learning rate is critical for the convergence of the learning algorithm in reasonable time. However, there is no guarantee that the algorithm will converge, converge at all or converge to the global minimum point. The learning rate depends on the number of inputs to the neurons. In a SAE the learning rate should be adapted to the depth of the layer, or to the number of inputs that the neurons have in the particular layer.

First experiment. A hyperspectral image from a natural scene was chosen. The scene contains a gravel road and vegetaion close to the road. There are man-made object placed in the scene. In the context where the data were captured the detection of the objects is of interest as could also a classification of the area in terms of different types of terrain. If the data can be represented in a lower dimensional space then the information might be easier to access.

We could also try if the reduced representation could be used for anomaly detection. If data does not fit the trained implicit model then the reconstruction should be worse than for common data from which the model was trained.

The representation obtained with a stacked autoencoder could be compared to for example a PCA, choosing spectral bands and representation with a gaussian mixture model. What should be compared is the number of parameters in the encoder, the number of dimensions in the obtained representation and the reconstruction error. Also the computational complexity of the methods is of interest, both the using the encoder and determining the encoder.

4 Experiments

Let $\{\mathbf{x}_i\}_{i=1}^{N_t}$ be a set of training vectors where each vector is the spectral information from one pixel, $\mathbf{x}_i = (x_1, x_2, ..., x_M)$, each vector describes the spectral intensity of M spectral bands. Let $\{v\}_{i\,i=1}^{\,N_v}$ be a set of validation vectors.

In our experiments we have used Rasmus Berg Palm's [7] Matlab implementation of a stacked autoencoder.

We have used hyperspectral images (e.g. see Fig. 3) from a natural scene with man made objects placed on the ground. The scene is 5-10 x 5-10 meters and the objects are a few decimeters in size. The images are mostly of undergrowth and some tree trunks. There are also parts of a gravel road. There are some calibration boards in the images but those are disregarded in the experiments. Figure 3 shows a visual image of one of the scenes and a mask showing where in the image to find background, objects and calibration boards.

These hyperspectral images have been collected for a project investigating among other things methods for anomaly detection of surface laid objects. There is a number of different objects.

The available vectors were divided into a training set and a validation set.

First we want to find out if it is possible to find an efficient representation of the hyperspectral data using a stacked autoencoder. With efficient we mean a representation in a space with few dimensions from which it is possible to reconstruct the original data with a small reconstruction error using mean square error to measure the error. The mean square error is a general error measure and not very specific which seems reasonable since we do not have a specific application in mind.

In our data (Fig. 4) the distribution of the intensity varies between the wavelength bands which means that the relative error of some bands contribute much more then other bands to the total mean square error. We hope that information in the original hyperspectral signal is retained in the encoded data.

We will use a stacked autoencoder to learn an efficient representation of the background data excluding the man-made objects in the scene. The compact representation could be used for anomaly detection. Here we consider detection of anomalous pixels since each spectrum is encoded separately without any regard to any other pixels. In many cases the objects have spatial properties that could be considered. But in this case we only want to explore the spectral properties.

If the man-made objects are different from the typical natural scene then they should be detectable in the sense that the stacked autoencoder fitted for the background will not be able to reconstruct the spectrum of the objects as well as spectrums from the background.

The data was normalised to improve convergence of the training algorithm by translation of each spectral band to give zero mean. The total set of spectral values was then normlised to unit variance. Figure 5 shows a few normalised spectrums.

In classification problems there is a choice of target values. In our case with a stacked autoencoder the last activation function was chosen to be a linear function. The training example data was shuffled for every epoch of the training. There is a choice of batch or stochastic training. In stochastic training the net is adapted to one data vector at a time. In batch learning the gradient is computed as an average for a set of training vectors. This will give a better approximation of the gradient but the stochastic training can get advantage from the random walk. In the implementation we used there was a significant difference in speed of going through an epoch in the learning algorithm. We choose a solution between the extremes by using small batches. The learning rate was determined by trial and error so that the learning algorithm converged reasonable fast. If it learning rate was chosen to high then the algorithm did not converge but kept jumping around a minimum value.

The stacked autoencoder is trained layer by layer. When the first layer is trained then the data is mapped to the latent representation and then these data are used to train the next layer. When all layers are trained they are put together into a stacked autoencoder with several hidden layers. The full stack is trained to fine tune parameters. In the case of a classifier the inner most hidden layer is attached to a few layers used for classification of the data and then the classifier is trained which include updates to the autoencoder layers apart from training the layers doing the classification.

Figure 7 shows an example of the training. 7(a) shows ten data vectors (spectrums) from the validation set and 7(b) shows the reconstruction of the spectrums given by the autoencoder and 7(c) shows the reconstruction residual (Eq 8). The data need to be normalised to suit the autoencoders learning algorithm, 7(d) is the ten vectors of 7a normalized.7(e) is the reconstructed vectors which are reconstructed in 7(b). 7(f) is the reconstruction residual (Eq 8) for the normalized vectors.

7(g) shows an image of the encoding matrix $\mathbf{W}_e$ of the first layer in the stacked autoencoder. The matrix has 60 rows one for each node and 240 columns, one for each insignal. Each point in the image represents the coefficient of one insignal to one node. 7(h) is the endcoded vectors and 7(i) is the corresponding reconstructed vectors. An image of the reconstruction matrix $\mathbf{W}_d$ is shown in 7(j). This matrix has 60 rows, one for each insignal and 240 rows, one for each output signal.

This SAE is composed of two layers. 7(k) shows an image of the parameters of the encoder in the second layer which has 60 input signals (columns) and 30 output signals (rows). 7(l) shows the encoded signals and finally 7(m) is an image of the reconstruction matrix wich 60 rows (insignals) and 30 columns (outputs signals).

The data in the innermost layer supposedly contain most of the information and this is shown by a small reconstruction error. For a specific classifier it is possible that it would suffice with fewer dimensions. From our experiments it seems that the original data has low dimensionality. Only a few nodes is needed in the hidden layer, e. g. 240 - 8 - 240 nodes. This is consistent with using PCA to reduce the dimensions. Only a few dimension contain most of the energy of the signal.

The representation obtained by the autoencoder may be used to detect anomalies. Figure 3 (bottom) shows the reconstruction error. Most of the man-made objects have considerably higher reconstruction error than most background pixels. It seems that the reconstruction error can be used as a measure of anomaly.

We also made an experiment with constructed data consisting of a handful of spectral signatures or functions. In this case the problem could be seen as a discrete coding problem. The innermost layer need only represent the index of the spectral signal. Thus only one dimension would suffice with the level indicating which spectral signal is current. We did not find a stacked autoencoder with only one node in the innermost layer but using 5 nodes in the innermost layer would give perfect reconstruction. In this case it means up to the noise that was added to the training and validation data.

There is a significant difference with regard to computation time when considering batch versus

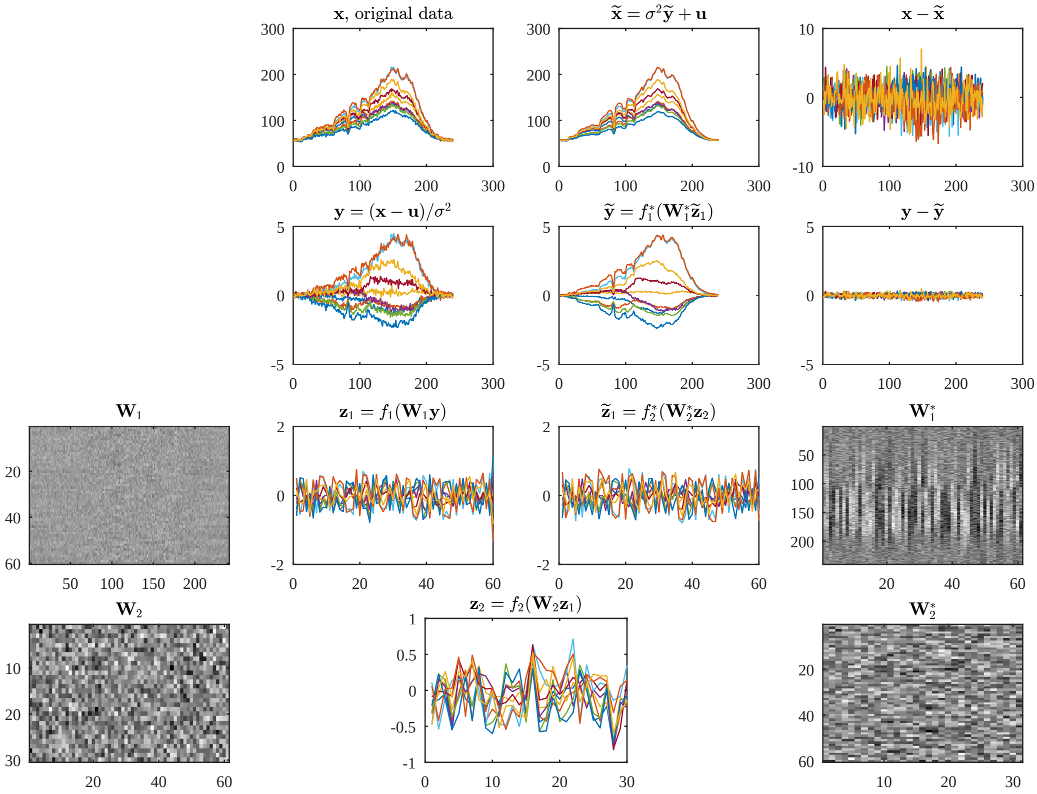


Figure 7: Data in the steps of a stacked autoencoder. The graphs are refeered to row by row as, a b c, d e f, g h i j, k l m. The first row shows (a) ten original spectral data vectors, (b) the ten corresponding reconstructed spectral data vectors from the SAE and (c) the reconstruction residual. The second row shows (d) the corresponding normalized data vectors, (e) the corresponding reconstructed data vectors and (f) the reconstruction residual. The fourth row shows (g) an image of the matrix with the encoder parameters, (h) the latent representation of the data vectors, (i) the reconstructed data vectors and (j) an image of the matrix with the decoder parameters. The fifth row shows (k) an image of the matrix with the encoder parameters of the second layer, (l) the latent representation of the data vectors in the second layer and (m) an image of the matrix with the decoder parameters of the second layer.

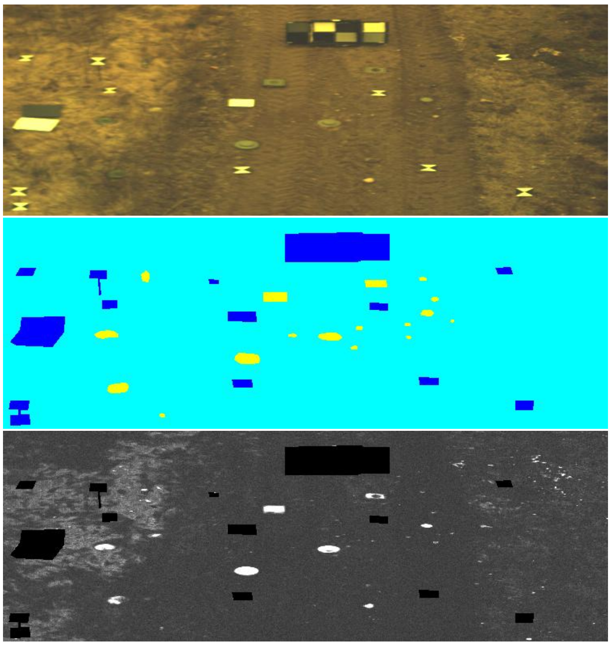


Figure 3: (Top) A visual image of the scene from which the test data is taken. (Middle) The different regions in the image, background (turquoise) objects (yellow) and reference boards (blue). (Bottom) The reconstruction error when spectrums are represented by a stacked autoencoder with 240,60,30,60,240 nodes in the layers.

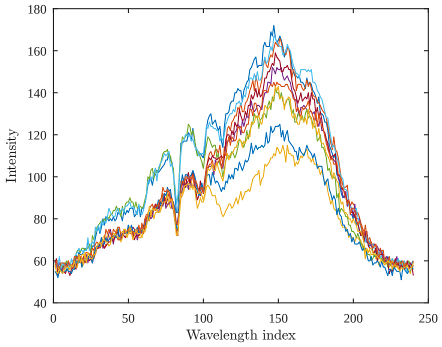


Figure 4: Examples of spectrums from some pixels in the test data.

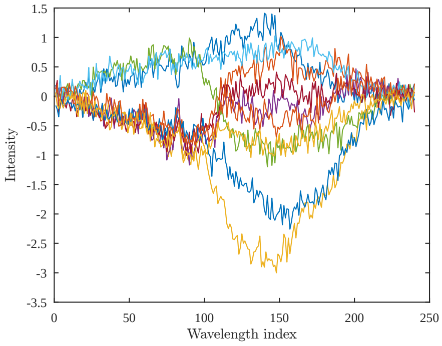


Figure 5: Examples of normalised spectrums used in the training.

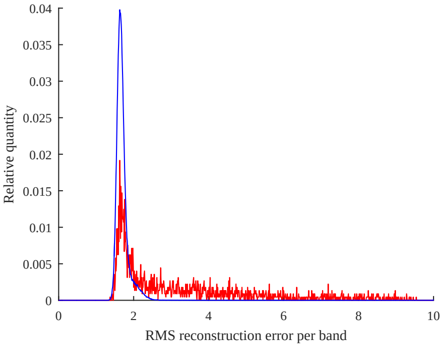


Figure 6: Histogram of the RMS error per spectral band for back- and forground pixels separately.

random choice of a single training vector. There is a significant gain in time to compute all vectors once.

5 Conclusion

In this paper we describe how stacked autoencoders (SAE) can be used to reduce the spectral dimensionality of hyper spectral data. We show that the hierarchical representation learned by a SAE can encode the spectral information with small mean square error. The results are shown using hyperspectral signatures from images of a natural scene containing man-made objects. We also show how the SAE can be used for anomaly detection, with promising results, on the same dataset.

Even if SAEs is a promising tool for dimensionality reduction and learning compact spectral representations it is a rather steep learning curve before one can apply SAEs.

Acknowledgement

This work is funded by the Swedish Defence Materiel Administration, FMV and the Swedish Armed Forces.

References

[1] Yoshua Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

[2] Y. Chen, X. Zhao, and X. Jia. Spectral - spatial classification of hyperspectral data based on deep belief network. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, PP(99):1–12, 2015.

[3] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 7(6):2094–2107, June 2014.

[4] M. D. Farrell and R. M. Mersereau. On the impact of pca dimension reduction for hyperspectral detection of difficult targets. *IEEE Geoscience and Remote Sensing Letters*, 2(2):192–195, April 2005.

[5] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Kernel principal component analysis for feature reduction in hyperspectrale images analysis. In *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, pages 238–241, June 2006.

[6] David Gustafsson, Henrik Petersson, and Mattias Enstedt. Deep learning: Concepts and selected applications. Technical Report FOI-D-0701-SE, Swedish Defence Research Agency (FOI), Sweden, December 2015.

[7] R. B. Palm. Prediction as a candidate for learning deep hierarchical models of data. Master's thesis, Technical University of Denmark, DTU Informatics, 2012.

[8] A. Villa, J. Chanussot, C. Jutten, J. A. Benediktsson, and S. Moussaoui. On the use of ica for hyperspectral image analysis. In *2009 IEEE International Geoscience and Remote Sensing Symposium*, volume 4, pages IV–97–IV–100, July 2009.

[9] Xue wen Chen and Xiaotong Lin. Big data deep learning: Challenges and perspectives. *Access, IEEE*, 2:514–525, 2014.

[10] Huiwen Zeng and H.J. Trussell. Dimensionality reduction in hyperspectral image classification. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 2, pages 913–916 Vol.2, Oct 2004.

Real Time Heart Rate Monitoring From Facial RGB Color Video Using Webcam

H. Rahman, M.U. Ahmed, S. Begum, P. Funk

***Abstract*— Heart Rate (HR) is one of the most important Physiological parameter and a vital indicator of people's physiological state and is therefore important to monitor. Monitoring of HR often involves high costs and complex application of sensors and sensor systems. Research progressing during last decade focuses more on noncontact based systems which are simple, low-cost and comfortable to use. Still most of the noncontact based systems are fit for lab environments in offline situation but needs to progress considerably before they can be applied in real time applications. This paper presents a real time HR monitoring method using a webcam of a laptop computer. The heart rate is obtained through facial skin color variation caused by blood circulation. Three different signal processing methods such as Fast Fourier Transform (FFT), Independent Component Analysis (ICA) and Principal Component Analysis (PCA) have been applied on the color channels in video recordings and the blood volume pulse (BVP) is extracted from the facial regions. HR is subsequently quantified and compared to corresponding reference measurements. The obtained results show that there is a high degrees of agreement between the proposed experiments and reference measurements. This technology has significant potential for advancing personal health care and telemedicine. Further improvements of the proposed algorithm considering environmental illumination and movement can be very useful in many real time applications such as driver monitoring.**

I. INTRODUCTION

The non-contact physiological parameters monitoring idea has come from the cardiovascular system of human body. The cardiovascular system permits blood to circulate in the body due to continuous blood pumping by heart. Our Heart pumps blood through the blood vessels of this system and for each heart beat blood circulation creates color variation in Facial skin. Therefore, it is possible to extract HR from the color variation of the facial skin. In 1995, the first noncontact health monitoring system was investigated by Costa et al. [1]. They used camera images in order to extract physiological parameters using color variation of the skin. But their approaches did not report quantitative results; they reported only a graph of heartbeats and also failed to show any correlation with reference ECG signals. After this first attempt further progress was moderate and in 2005 another novel method was introduced for the measurement of computer user's emotional state using the facial thermal image using a thermal camera [2]. The experiment was conducted by 12 users and the authors found some interesting fact between stress and blood flow. According to their experiment user stress is correlated with increased blood flow in the frontal vessel of the forehead. In 2006, Takano et al. shows that RR (Respiratory Rate), HR and BVP are possible to extract simultaneously using a camera [3]. They captured images of a part of the subject's skin and then the changes in the average image brightness of the region of interest (ROI) are measured for a short time. They used MATLAB custom functions for filtering and spectral analysis. Finally, they could able to extract HR and HRV (Heart Rate Variability). The system can detect HR for a certain period of time but the efficiency of their system is unknown. Later in 2007, Garbey et al. developed a contact-free measurement of cardiac pulse based on the analysis of thermal images using FFT algorithm [4]. Their experiment shows that the temperature of the vessel is modulated by pulsative blood flow is directed at recovering the frequency of the component signal with the highest energy content. The effort is directed at recovering the frequency of the component signal with the highest energy content. After appropriate processing, the thermal image signal can yield quantitative information about blood flow velocity, respiratory function etc. The noncontact methods using camera further improved in the same year by Kenneth et al [5]. They presented a system capable of capturing two PPG (Photoplethysmogram) signals simultaneously at two different wavelengths using non-contact system. Ten test persons participated in their experiment where both camera and PPG sensors were used for data collection. Their proposed system extracted oxygen saturation (SpO2) successfully but the efficiency is not compared. Additionally, they showed that the system was capable of obtaining good quality PPG signals from deep tissue. Another successful attempt was done in 2008 by Verkruysse et al. [6]. Few Simple, inexpensive digital cameras were used to extract HR and RR from facial video recorded in ambient light. Their system is able to extract HR and RR from 30 seconds to a few minutes which is a major flaw to apply in real time applications.

Noncontact based method for physiological parameters extraction has been further improved in recent years. A novel method was presented by Banitsas et al. in 2009 which is able

---

*Hamidur Rahman is a Ph.D. student at Mälardalen University (MU), Sweden. Shahina Begum is an Associate Professor, Mobyen Uddin Ahmed is a Post-doctoral fellow and Peter Funk is a Professor in MU, Sweden. (E-mail: FirstName.LastName @ mdh.se).

to extract HR information from a user using the camera of a smart phone using user's finger [7]. In different experiments HR and RR were extracted by smart phone camera using user's finger in [8], [9] and [10]. In 2011, Poh et al. proposed an algorithm to extract underlying source signals from R, G, and B color bands using computer webcam [11]. The experiments were conducted using built-in webcam (iSight camera) in indoor environment. BVP was also recorded with spontaneous breathing using a finger BVP sensor and chest belt respiration sensor respectively at a sampling rate of 256 Hz. Finally they extracted HR from 1 minute color facial video using FFT. Similar experiments were taken place using different cameras for physiological parameters extraction which are referred in [12], [13] and [14]. A main drawback of systems for use in personal health care, telemedicine and real time applications such as driver monitoring are that they are not real time and they did not show that how much time the system can extract physiological parameters.

In 2013, Parnandi et al. approached an algorithm to extract HRV using a remote eye tracker [15]. HRV was estimated from the relative distribution of energy in the low frequency (0.04 to 0.15 Hz) and high frequency (0.15 to 0.4 Hz) bands of the power spectrum of the time series of pupillary fluctuations. The system was validated under a range of breathing conditions and under different illumination levels in offline situation. In the same year, several attempts were taken place for non-contact physiological parameters such as [16], [17], [18], [19], [20] and [21]. In 2014, Zhang et al. developed a webcam based noncontact monitoring system for the physiological parameters of drivers [22]. Using iSight camera the facial images are captured for several minutes which are separated into three RGB channels and each channel. FFT is used to measure HR and RR. In the same year, Guo et al. showed similar approach to monitor driver's HRV continuously under real world driving circumstances [23]. Using normal computer webcam one video sequence (15 fps and 640 x 480 resolutions) is taken of the driver face and then HRV is calculated from BVP. Dividing the Face regions into 7 sub regions and average HRV was extracted using ICA method for all the regions. Another approach by Xiaobai Li et al. was proposed for HR measurement framework under realistic human computer interaction (HCI) situations [12]. Some others similar works were done in [8, 24-31].

The most successful noncontact based physiological parameters extraction system has been proposed by Rahman et al. in 2015 [32]. They have developed a simple laptop web camera based method to detect HR, RR and IBI (inter bit interval). The results show about 90% accuracy for physiological parameters extraction using this system. This experiment can extract three physiological parameters in offline for any length of time. From all the related literatures it is seen that most of the noncontact systems to monitor physiological parameters are done in offline and most of them are good for a certain amount of time in lab environment. This paper presents a noncontact HR monitoring system in real time for unlimited amount of time using a web camera which overcomes most of the flaws of the previous works.

The rest of the paper is organized as follows: chapter II describes materials and methods, chapter III highlights feature extraction using image processing tasks for noncontact experiments, chapter IV discusses real time HR monitoring method and the experiments and results have been focused on in chapter V. Finally, chapter VI summarizes the work.

II. MATERIALS AND METHODS

The experiment was taken place in two phases: firstly the real time HR extraction was conducted along with cStress system as a reference. All the facial image frames were saved for offline testing. Secondly HR was extracted again in offline using the saved film image sequences.

A. *Data Collection*

Data acquisition was conducted by 10 participants (all are male) of different ages (25 to 50 years) and skin colors. The experiments were carried out in indoors and with a sufficient amount of ambient sunlight. The participants were informed the aim of the study and they seated at a table in front of a laptop computer at a distance of approximately 0.5 m from the built-in webcam (HP HD webcam). During the experiment, participants were asked to keep still, breathe spontaneously, and face the webcam while their video was recorded for 5 minutes. HR was extracted in real time and saved in an excel file. All facial image frames (24-bit RGB) during real time HR extraction were recorded sequentially at 30 frames per second (fps) with pixel resolution of 640 × 480 and saved in PNG (Portable Network Graphics) format in the laptop. Simultaneously HR was also recorded using ECG sensors and cStress system[1]. After the real time extraction, HR was also extracted again in offline from the saved film image sequences.

B. *Applied Algorithms*

Three algorithms such as FFT, ICA and PCA have been applied at the same time but separately to extract HR in real time using only facial video. The average of the R, G and B signals were calculated for FFT method. For the ICA method [33], the normalized raw traces were decomposed into three independent source signals (R, G and B) based on the joint approximate diagonalization of Eigen matrices (JADE) algorithm [34]. The data collection was supposed to perform in sitting position without any movement but in reality the test persons moved their hands and heads little bit which is the cause of motion artifacts. Therefore, ICA is used which is able to remove motion-artifact by separating the fluctuations caused by small motions or movement. Interestingly, ICA returns the independent components randomly and the component whose power spectrum contained the highest peak is then selected for further analysis. Similarly the normalized raw traces are also decomposed by PCA to find the principal components [35]. This transformation is defined in such a way that the first principal component has the largest possible variance and each succeeding component in turn has the highest variance possible under the constraint that it is

---

[1] http://stressmedicin.se/neuro-psykofysilogiska-matsystem/cstress-matsystem/.

orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. PCA is sensitive to the relative scaling of the original variables. Finally, the Fast Fourier Transform (FFT) is applied on the selected source signal to obtain the power spectrum [36]. The pulse frequency was designated as the frequency that corresponded to the highest power of the spectrum within an operational frequency band.

III. FEATURE EXTRACTION

The main features of the proposed method is 3 independent signals which are called Red signals, Green signals and Blue signals and these signals were produced from the red, blue and green color values of each pixel of all the facial image frames.

A. *Reading Image Frames*

An image frame is the fundamental part of a video or any image source that indicates the start and end point of a video which represents a silent part of that video. Fig. 1(a) shows the real time HR monitoring system to extract a number of image frames one by one at a certain period of time defined by the user. It is also important to notice that the resolution of the video should remain same during each image frame extraction for further calculations. Therefore a novel key frame video extraction algorithm has been used to maintain same resolution that can read image frames automatically one by one [37].

B. *Face Tracking*

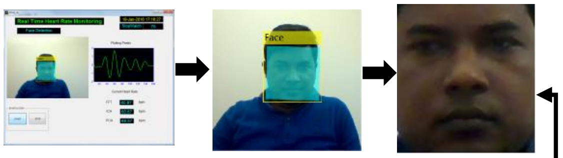
Facial image is the input of the proposed non-contact HR monitoring algorithm and therefore it is very important to track facial part of the user. The real time method needs a powerful face tracking method to perform higher face detection rate. After extracting an image frame in real time, the automatic face detection function '*CascadeObjectDetector'* of Computer Vision Toolbox provided by MATLAB[2] was applied which has been implemented using Viola and Jones method [38]. Later the function was modified to fulfill our own purposes. Fig. 1(b) indicates the detected face.

C. *Region of Interest Selection*

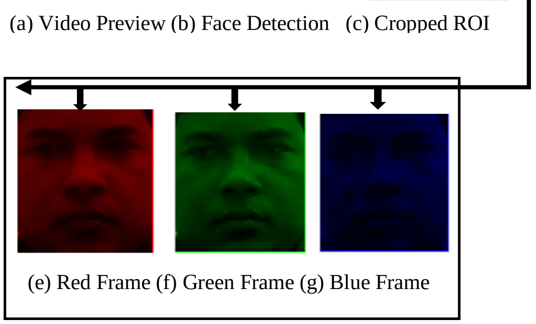
R, G and B color values of each pixel of the facial image frames are the most essential part for this experiment. Hence it was searched a perfect Region of Interest (ROI) over the detected face. The detected face using Viola and Jones method contains some unwanted part which needs to eliminate. To identify the coordinates of the face location in the first frame a boosted cascade classifier was used for the *x* and *y*-coordinates along with the height and width that define a box around the face according to the method in [39]. Therefore the center was selected as 60% width and 80% height of the box as the region of interest which is free from unwanted parts. Only the ROI was then separated from the entire facial image shown in Fig. 1(c) and this ROI is used for further calculations.

D. *RGB Signals Extraction*

R, G, B color values are the fundamental elements of R, G and B signals (together they are called RGB signals) which were extracted from the facial cropped RIO image [40]. Each pixel of the image has 3x1 matrix of color values which consists of Red (R), Green (G) and Blue (B) color of the image. Then the three desired signals Red, Green and Blue signals are produced in two phases. In the first phase the average R, G and B color values are calculated for each image frame shown in Fig. 1(d) and in the second phase the red, green and blue signals are calculated from the summation of all the averaged R, G and B color values indicated in Fig. 1(e-g).



(a) Video Preview (b) Face Detection (c) Cropped ROI



(e) Red Frame (f) Green Frame (g) Blue Frame

(d) RGB Frames

Figure 1. Feature Extraction from each image frame

E. *Signal Detrending*

Detrending is an important signal processing concept which is used to remove unwanted trend from the series. Detrending of signal is useful when it is thought that a feature is distorted from the relationships of interest. In our case, when environmental parameters changes such as temperature or external noise, the collected RGB signals will be drifting and noising. Therefore the signals need to detrend. The RGB signal has been detrended using the method used in [41] based on smoothnes priors approach with the smoothing parameter $\lambda = 10$ and cutoff frequency = 0.059 Hz shown in Fig. 2(h).

E. *Filtering*

Before applying PCA, ICA and FFT the Red, Green and Blue signals in Fig. 2(d-f) formed from all red, green and blue image frames in Fig. 2(a-c) are filtered by Hamming window (128 point, 0.6-2 Hz, for normal HR 36-120) for heart rate [42] shown in Fig. 2(j).

F. *Normalization*

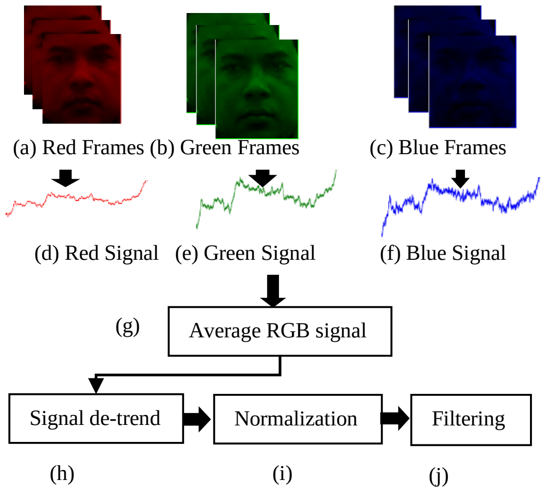
The signal needs to be normalized and the normalization has been performed according to the method mentioned in [43] in Fig. 2(i). Equation (1) shows the normalization formula as below:

---

[2] "MATLAB Computer Vision Toolbox," R2013a ed: The MathWorks Inc., pp. Natick, Massachusetts, United States.

$$X_i(t) = \frac{Y_i(t) - \mu_i(t)}{\delta_i} \tag{1}$$

For each i = R, G and B signals where μ_i is the mean and δ_i is the standard deviation of Y_i.


(a) Red Frames (b) Green Frames (c) Blue Frames

(d) Red Signal (e) Green Signal (f) Blue Signal

Figure 2. RGB Signals Pre-processing

IV. REAL TIME MONITORING SYSTEM

The physiological parameters extraction methods in [32] presents for physiological parameters (HR, RR and IBI) extraction for any length of time in offline. In order to use physiological parameters extraction in real life application like driver monitoring [44] or driver monitoring in semi-autonomous vehicle [45] or in any other health applications [46], a real-time HR extraction method has been implemented in this paper. The proposed method has four main parts which are (i) video display and extract each image frame, (ii) face detection and facial image extraction, (iii) RGB signal extraction and pre-processing and (iv) Extraction of HR using ICA, PCA, FFT methods and displaying the results. For real time HR extraction it is not necessary to save any facial video during experiment. But all the facial images have been saved to extract HR in offline which is also a part of this paper. At first the system previews the user's video in continuous mode and it performs all the works described in fig. 1. When the system reads 50 image frames one by one it stores all the R, G and B color values of the cropped facial image in a temporary database which are sent to the processing unit for further processing. Before applying ICA, PCA and FFT the signals are processed according to the steps in fig. 2 and also described vividly in section III.

To minimize the update time of the results it was investigated using different number of image frames to extract HR and it was found that the system can able to perform best HR monitoring for 50 image frames for the first time which needs 2-3 seconds and after that the update is possible for reading each image frame which is around 500 milliseconds (ms). The update of the result depends on how fast the system can read each image frames. It was investigated that the average update time was about 300 ms to read every image frame and display the results. Hence the system can extract physiological parameters of the user in real time for any length of time. To extract HR in real time at first the number of peaks in frequency domain was calculated for first 50 image frames and also the required time was recorded. Therefore HR is calculated as HR = 60*f_h bpm (beat per minute) where f_h is the extracted frequency of the HR. For example Fig. 3 shows 25 peaks for 600 image frames. The frame rate was 30 fps and therefore the required time to read 600 image frames was 20 second. HR is calculated as below:

$$\begin{aligned} HR &= 60 * f_h \text{ bpm} \\ &= [60 \text{ x (number of Peaks/Time)] bpm} \\ &= 60 \text{ x } (25/20) \text{ bpm} \\ &= 75 \text{ bpm.} \end{aligned}$$

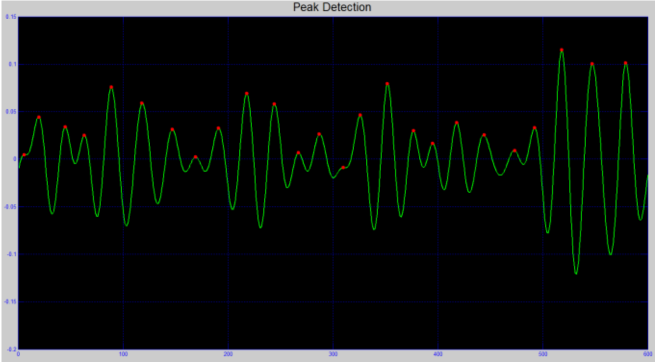


Figure 3. Real Time HR monitoring GUI

When the first 50 image frames are read, the real time HR extraction begins and after that each image frame is added to the data base and the method provides new HR.

A Graphical User Interface (GUI) has been developed using MATLAB to monitor HR in real time which has 3 main sub-sections. Fig. 3 shows the real time HR monitoring Graphical User Interface (GUI) which displays the 3 subsections such as detected face, pulse peak and current HR using the three methods.

19-Jan-2016 17:16:27

StopWatch 75

Current Heart Rate

FFT 60.87 bpm

ICA 60.87 bpm

PCA 69.57 bpm

Figure 4. Real Time HR monitoring GUI

V. EXPERIMENT AND RESULTS

HR was extracted and recorded for 5 minutes for all the 10 test persons in real time using webcam and cStress system using ECG sensors and the extracted HR values were saved in two different excel files. After 5 minutes the real time session was over and HR was extracted again in offline using the saved film sequences using the proposed algorithms of [32]. For each test subject there were 3 separate excel files for the extracted HR; one was for real time method, another one for cStress system for reference and the last one for offline

TABLE I. STATISTICAL ANALYSIS OF HR IN REAL TIME VS OFFLINE CONSIDERING CSTRESS SYSTEM

SUBJECT	PARAMETERS	Real Time			Offline			CSTRESS
		FFT	PCA	ICA	FFT	PCA	ICA	
1	min	46	51	52	63	54	60	64
	max	98	99	90	88	98	88	87
	mean	71	73	70	77	76	78	76
	std	7	9	7	5	7	4	4
	median	70	72	69	77	76	78	76
2	min	57	59	57	58	51	64	53
	max	89	91	89	99	99	99	64
	mean	73	76	73	74	76	74	56
	std	6	5	6	5	6	5	2
	median	73	76	73	73	77	73	56
3	min	57	59	57	61	63	64	58
	max	90	99	90	99	99	99	71
	mean	76	75	76	79	80	80	64
	std	7	5	7	4	5	4	3
	median	78	75	78	80	80	80	64
4	min	54	61	54	54	60	54	64
	max	91	99	91	94	90	94	85
	mean	82	78	82	82	77	82	73
	std	6	6	6	6	4	6	4
	median	83	78	83	84	77	84	72
5	min	51	50	59	55	50	61	41
	max	99	99	99	99	99	99	93
	mean	71	74	72	72	75	75	61
	std	6	6	6	6	7	7	8
	median	71	74	72	71	75	74	61
6	min	63	59	63	55	61	55	71
	max	90	96	90	94	95	94	90
	mean	78	77	78	79	77	79	77
	std	5	7	5	6	6	6	3
	median	77	78	78	79	77	79	77
7	min	46	51	52	64	58	64	61
	max	98	99	90	95	99	95	99
	mean	72	74	70	77	78	77	70
	std	7	10	7	6	5	6	7
	median	70	72	69	76	77	76	68
8	min	51	50	59	54	51	67	71
	max	99	99	99	99	90	99	98
	mean	71	74	72	82	77	83	83
	std	6	6	6	6	5	6	4
	median	71	74	72	83	77	83	84
9	min	65	58	74	66	52	74	66
	max	98	97	98	92	99	92	95
	mean	87	84	87	87	84	87	81
	std	3	4	3	3	5	3	5
	median	70	72	69	77	76	78	76
10	min	47	50	48	61	53	58	66
	max	98	99	88	87	96	90	87
	mean	72	73	68	76	74	76	76
	std	7	8	7	6	7	5	4
	median	71	72	69	75	76	78	75

method. It is necessary to do statistical analysis to find out the efficiency of the proposed method with respect to reference sensor system. Therefore several parameters such as minimum, maximum, average, median and standard deviations were calculated from the extracted HR for the real time extraction method, cStress system and offline method. These statistical parameters for all the 10 test persons are presented in table I. For the evaluation then it is necessary to calculate some statistical analysis. Therefore, the evaluation was made using 2 important statistical parameters such as RSQ (R-squared) and CORREL (Correlation Coefficient) for both real time and offline HR extraction considering cStress system.

It should be noted here that the RSQ parameter is used to see how close the obtained signals of HR to the reference signal. RSQ ranges 0 to 1 where 0 indicates that the two signals are not correlated with each other where 1 indicates that the signals are fully correlated which means that the model explains all the variability of the response data around its mean. CORREL (also known as The Pearson product-moment correlation coefficient) is another statistical measurement of the correlation (linear association) between two sets of values. The CORREL value ranges -1 to +1 where +1 indicates a strong positive correlation and -1 indicates a strong negative correlation. These statistical parameters were calculated both for real time and offline with respect to cStress using the parameters from Table I. The statistical analyses of HR in real time for the 10 test subjects are presented in Table II.

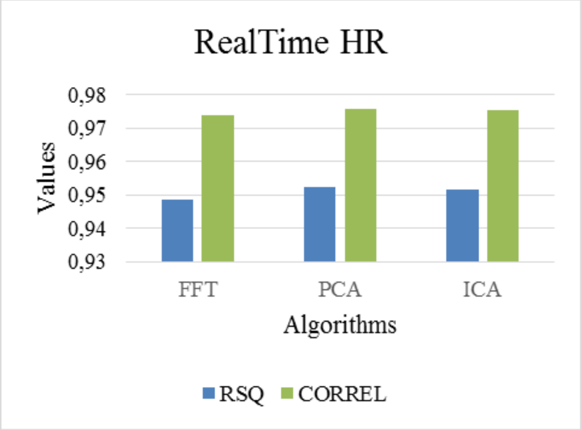


Figure 5a. Comparison among three methods for Real time HR Extraction considering cStress System

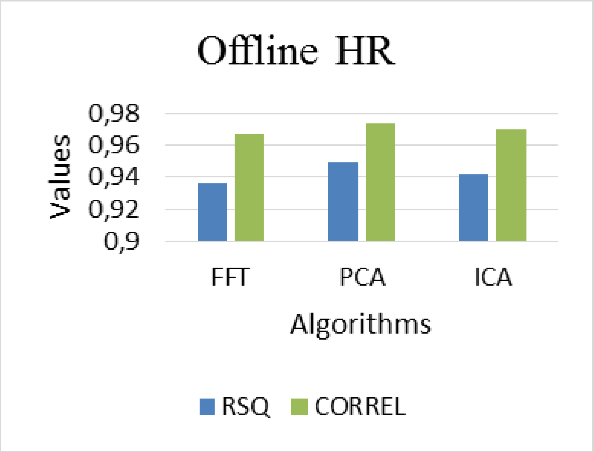


Figure 5b. Comparison among three methods for Offline HR Extraction considering cStress System

TABLE II. STATISTICAL ANALYSIS OF HR IN REAL TIME VS OFFLINE CONSIDERING CSTRESS SYSTEM

SUBJECT	PARAMETER	Real Time			Offline		
		FFT	*ICA*	*PCA*	*FFT*	*ICA*	*PCA*
1	RSQ	0.90	0.89	0.96	0.99	0.95	0.99
	CORREL	0.95	0.94	0.98	0.99	0.97	0.99
2	RSQ	0.95	0.96	0.95	0.91	0.86	0.94
	CORREL	0.97	0.98	0.97	0.96	0.93	0.97
3	RSQ	0.95	0.93	0.95	0.92	0.95	0.94
	CORREL	0.98	0.97	0.97	0.96	0.98	0.97
4	RSQ	0.94	0.97	0.94	0.94	0.99	0.94
	CORREL	0.97	0.98	0.97	0.97	0.99	0.97
5	RSQ	0.99	0.98	0.96	0.97	0.98	0.98
	CORREL	0.99	0.99	0.98	0.98	0.99	0.97
6	RSQ	0.98	0.96	0.98	0.94	0.97	0.94
	CORREL	0.99	0.98	0.99	0.97	0.98	0.97
7	RSQ	0.89	0.94	0.86	0.81	0.90	0.81
	CORREL	0.95	0.97	0.93	0.91	0.95	0.90
8	RSQ	0.93	0.94	0.96	0.94	0.96	0.99
	CORREL	0.966	0.97	0.98	0.97	0.98	0.99
9	RSQ	0.98	0.98	0.97	0.96	0.96	0.95
	CORREL	0.99	0.99	0.98	0.98	0.98	0.97
10	RSQ	0.97	0.97	0.96	0.95	0.95	0.94
	CORREL	0.98	0.98	0.97	0.97	0.97	0.96

As can be seen from Table 1, both the RSQ and CORREL values are close to 90% or more than 90%. CORREL function gives better result than RSQ both in real time and offline. The average RSQ and CORREL values of 10 subjects were also calculated for real time HR methods and offline HR methods by applying three algorithms and presented through bar charts as fig.4a and fig.4b. Average CORREL value is 0.97 for real time and 0.95 for offline which indicates that there is a strong positive correlation between the proposed methods and the reference system. Average RSQ value for real time is 0.93 and offline is 0.91 which also indicates a perfect fitness between the two methods. According to the Figures, real time method shows its best performance compare to the offline methods. It may happen because of the loss of R, G and B color values of the film image sequences during saving in local disk. Lower resolution of the video may be another reason of this performance. It is also seen from the fig.4 that PCA method gives the best results both in real time and offline and ICA methods show better result than FFT.

VI. CONCLUSION

A real time noncontact based HR extraction method is described in this paper using facial video which is easy to implement, low cost and comfortable for real time applications. Here, the main idea is to extract HR from the color variation in the facial skin due to cardiac pulse and the implementation has been done using a simple webcam in indoor environment with constant ambient light. According to the experimental works, both the RSQ and CORREL values shows highest closeness (i.e. >90%) with the reference measurements. From the table and the figures presented in earlier chapter it is noted that the correlation using CORREL parameters (97.5%) is higher than RSQ (96.5%) in real time and among the three methods PCA shows the highest accuracy and ICA works better. Better results (i.e. 99%) can be achieved by taking the average of the three methods and using HD (High Definition) video (1280 x 720 or even 1980 x 1080). This non-contact technology is promising for medical care and others indoor applications due to widespread availability of camera specially webcams. For applications in outdoor environment for example driver monitoring, few things such as variable environmental illumination or head movement should be considered. Also to increase the efficiency, the experiment needs to be done by more test subjects and more verifying systems. Although this paper only addressed the recovery of the cardiac HR, many other important physiological parameters such as, RR, HRV and arterial blood oxygen saturation can potentially be estimated using the proposed technique. Creating a real-time, multi-parameter physiological measurement platform with higher resolution of video based on this technology in driving situation will be the subject of future work.

ACKNOWLEDGMENT

The authors would like to acknowledge the Swedish Knowledge Foundation (KKS), Swedish Governmental Agency for Innovation Systems (VINNOVA), Volvo Car Corporation, The Swedish National Road and Transport Research Institute (VTI), Autoliv AB, Hök instrument AB, and Prevas AB Sweden, for their support of the research projects in this area.

REFERENCES

[1] G. D. Costa, "Optical remote sensing of heartbeats," *Optics Communications*, vol. 117, pp. 395-398, 6/15/ 1995.

[2] C. Puri, J. Levine, L. Olson, I. Pavlidis, and J. Starren, "StressCam: Non-contact Measurement of Users' Emotional States through Thermal Imaging," presented at the Extended Abstracts Proceedings of the 2005 Conference on Human Factors in Computing Systems, CHI 2005, Portland, Oregon, USA, 2005.

[3] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Medical Engineering & Physics*, vol. 29, pp. 853-857, 10// 2007.

[4] M. Garbey, S. Nanfei, A. Merla, and I. Pavlidis, "Contact-Free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery," *Biomedical Engineering, IEEE Transactions on*, vol. 54, pp. 1418-1426, 2007.

[5] K. Humphreys, T. Ward, and C. Markham, "Noncontact simultaneous dual wavelength photoplethysmography: a further step toward

noncontact pulse oximetry," *Rev Sci Instrum*, vol. 78, p. 044304, Apr 2007.

[6] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, pp. 21434-21445, 2008.

[7] S. Suzuki, T. Matsui, S. Gotoh, Y. Mori, B. Takase, and M. Ishihara, "Development of Non-contact Monitoring System of Heart Rate Variability (HRV) - An Approach of Remote Sensing for Ubiquitous Technology," in *Ergonomics and Health Aspects of Work with Computers*. vol. 5624, B.-T. Karsh, Ed., ed: Springer Berlin Heidelberg, 2009, pp. 195-203.

[8] J. Wen Jun, G. Shi Chao, P. Wittek, and Z. Li, "Real-time quantifying heart beat rate from facial video recording on a smart phone using Kalman filters," in *e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on*, 2014, pp. 393-396.

[9] T. I. Papon, I. Ahmad, N. Saquib, and A. Rahman, "Non-invasive heart rate measuring smartphone applications using on-board cameras: A short survey," in *Networking Systems and Security (NSysS), 2015 International Conference on*, 2015, pp. 1-6.

[10] R. C. Peng, X. L. Zhou, W. H. Lin, and Y. T. Zhang, "Extraction of heart rate variability from smartphone photoplethysmograms," *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 1-11, 2015.

[11] P. Ming-Zher, D. J. McDuff, and R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, pp. 7-11, 2011.

[12] L. Xiaobai, C. Jie, Z. Guoying, and M. Pietikainen, "Remote Heart Rate Measurement from Face Videos under Realistic Situations," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, 2014, pp. 4264-4271.

[13] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam- a non-contact method for evaluating cardiac activity," in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, 2011, pp. 405-410.

[14] D. Shao, C. Liu, F. Tsow, Y. Yang, Z. Du, R. Iriya, *et al.*, "Noncontact Monitoring of Blood Oxygen Saturation Using Camera and Dual-Wavelength Imaging System," *Biomedical Engineering, IEEE Transactions on*, vol. PP, pp. 1-1, 2015.

[15] A. Parnandi and R. Gutierrez-Osuna, "Contactless Measurement of Heart Rate Variability from Pupillary Fluctuations," in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, 2013, pp. 191-196.

[16] D. Datcu, M. Cidota, S. Lukosch, and L. Rothkrantz, "Noncontact automatic heart rate analysis in visible spectrum by specific face regions," presented at the Proceedings of the 14th International Conference on Computer Systems and Technologies, Ruse, Bulgaria, 2013.

[17] V. Foteinos, D. Kelaidonis, G. Poulios, P. Vlacheas, V. Stavroulaki, and P. Demestichas, "Cognitive Management for the Internet of Things: A Framework for Enabling Autonomous Applications," *Vehicular Technology Magazine, IEEE*, vol. 8, pp. 90-99, 2013.

[18] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, pp. 1645-1660, 9// 2013.

[19] L. F. Jimenez, A. Parnandi, and R. Gutierrez-Osuna, "Extracting Heart Rate and Respiration rate using cell phone Camera," 2013.

[20] A. M. Mzahm, M. S. Ahmad, and A. Y. C. Tang, "Agents of Things (AoT): An intelligent operational concept of the Internet of Things (IoT)," in *Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on*, 2013, pp. 159-164.

[21] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, "Remote Measurements of Heart and Respiration Rates for Telemedicine," *PLoS ONE*, vol. 8, p. e71384, 2013.

[22] Z. Qi, X. Guo-qing, W. Ming, Z. Yimin, and F. Wei, "Webcam based non-contact real-time monitoring for the physiological parameters of drivers," in *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2014 IEEE 4th Annual International Conference on*, 2014, pp. 648-652.

[23] G. Zhenyu, Z. J. Wang, and S. Zhiqi, "Physiological parameter monitoring of drivers based on video data and independent vector analysis," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 2014, pp. 4374-4378.

[24] F. Bousefsaf, C. Maaoui, and A. Pruski, "Remote assessment of physiological parameters by non-contact technologies to quantify and detect mental stress states," in *Control, Decision and Information Technologies (CoDIT), 2014 International Conference on*, 2014, pp. 719-723.

[25] M. Gerla, L. Eun-Kyu, G. Pau, and L. Uichin, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds," in *Internet of Things (WF-IoT), 2014 IEEE World Forum on*, 2014, pp. 241-246.

[26] W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino, and G. A. Dumont, "Respiratory rate assessment from photoplethysmographic imaging," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 5397-5400.

[27] A. Khelil and D. Soldani, "On the suitability of Device-to-Device communications for road traffic

safety," in *Internet of Things (WF-IoT), 2014 IEEE World Forum on*, 2014, pp. 224-229.

[28] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control,* vol. 13, pp. 102-112, 2014.

[29] P. P. Ray, "Home Health Hub Internet of Things (H³IoT): An architectural framework for monitoring health of elderly people," in *Science Engineering and Management Research (ICSEMR), 2014 International Conference on*, 2014, pp. 1-3.

[30] T. Smilkstein, M. Buenrostro, A. Kenyon, M. Lienemann, and G. Larson, "Heart rate monitoring using Kinect and color amplification," in *Healthcare Innovation Conference (HIC), 2014 IEEE*, 2014, pp. 60-62.

[31] H. Xiaochuan, R. A. Goubran, and X. P. Liu, "Using Eulerian video magnification framework to measure pulse transit time," in *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*, 2014, pp. 1-4.

[32] H. Rahman, M. U. Ahmed, and S. Begum, "Non-contact Physiological Parameters Extraction using Camera," in *The 1st Workshop on Embedded Sensor Systems for Health through Internet of Things (ESS-H IoT), Oct.*, 2015.

[33] P. Comon, "Independent Component Analysis, a new Concept," *Signal Process.,* vol. 36, pp. 287–314, 1994.

[34] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Radar and Signal Processing, IEE Proceedings F,* vol. 140, pp. 362-370, 1993.

[35] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE Journal,* vol. 37, pp. 233-243, 1991.

[36] H. Buijs, A. Pomerleau, M. Fournier, and W. Tam, "Implementation of a fast Fourier transform (FFT) for image processing applications," *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 22, pp. 420-424, 1974.

[37] T. M. Liu, H. J. Zhang, and F. H. Qi, "A novel video key frame extraction algorithm," in *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*, 2002, pp. IV-149-IV-152 vol.4.

[38] P. Viola and M. Jones, "Robust real-time face detection," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001, pp. 747-747.

[39] Z. Deng-feng, Y. Shui-sheng, and H. Shao-hua, "Research on Application of Region-of-Interest in Face Detection," in *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on*, 2010, pp. 192-194.

[40] G. Akti and D. Goularas, "Frequency component extraction from color images for specific sound transformation and analysis," in *Image Processing Theory, Tools and Applications (IPTA), 2012 3rd International Conference on*, 2012, pp. 253-258.

[41] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis," *Biomedical Engineering, IEEE Transactions on,* vol. 49, pp. 172-175, 2002.

[42] K. K. Sundaram and V. Ramachandran, "Analysis of the coefficients of generalized bilinear transformation in the design of 2-D band-pass and band-stop filters and an application in image processing," in *Electrical and Computer Engineering, 2005. Canadian Conference on*, 2005, pp. 1233-1236.

[43] D. Cochran, "A consequence of signal normalization in spectrum analysis," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, 1988, pp. 2388-2391 vol.4.

[44] H. Rahman, S. Barua, and S. Begum, "Intelligent Driver Monitoring Based on Physiological Sensor Signals: Application Using Camera," in *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*, 2015, pp. 2637-2642.

[45] H. Rahman, S. Begum, and M. U. Ahmed, "Driver Monitoring in the Context of Autonomous Vehicle," in *13th Scandinavian Conf. on Artificial Intelligence*, Halmstad, Sweden, 2015.

[46] S. Begum, S. Barua, and M. U. Ahmed, "Physiological Sensor Signals Classification for Healthcare using Sensor Data Fusion and Case-Based Reasoning," *Sensors (Special Issue on Sensors Data Fusion for Healthcare),* vol. 7, Jul. 2014.

Towards Evacuation Planning of Groups with Genetic Algorithms

Bjørnar Hansen Leonard Loland Morten Goodwin
Ole-Christoffer Granmo
Department of ICT, University of Agder,
Grimstad, Norway

Abstract

In crisis situations on board ships, it is of utmost importance to have the passengers safely evacuate to the lifeboats in an efficient manner. Existing methods such as marked escape routes and maps are not optimal as pre-planned escape routes may become heavily congested by passengers. Further, the closest lifeboat is not always feasible as lifeboat capacity can be exceeded. Also considering that some evacuees are strongly affiliated, such as families, and that they prefer to evacuate together as a group, it becomes a difficult problem to solve.

This paper models the area to be evacuated as a time-expanded graph with hazard severities as probabilities of survivability for each node. The presented approach applies a multi-objective genetic algorithm with multiple fitness functions to maximize the over all survivability. Finally, the proposed method picks the best evacuation plan from a pool of potential solutions returned by the genetic algorithm.

The solution generates better routing plans than comparable methods, specially in situations where grouping and congestions are considered. In essence this is an essential step towards automatic planning of evacuations which in turn contributes to smoother evacuations of crises situations and saving lives.

1 Introduction

In an ongoing crisis situation, on ships and elsewhere, many challenges have to be faced during evacuation. In the case of fire, as it spreads over time, it produces an ever-increasing amount of lethal heat and smoke, rendering rooms and corridors hazardous or unusable for evacuation. Further, emergency response teams can be late and may not have the capacity to assist everyone efficiently. A direct consequence is that people are often initially left to themselves in evacuation situations. The closest emergency exits may become heavily congested as masses of people converge on them, while the nearest lifeboats quickly reach their maximum capacity, forcing other evacuees to make detours to search for alternatives. It is also possible that escape routes are rendered too dangerous or unusable, and alternate, perhaps non-obvious, routes have to be used. On top of it all, the information required to make the best course of action, such as the locations of people and hazards, may not be available, or be erroneous, during the crisis. All this may lead to valuable time and resources being wasted.

While traditional static signs are meant to guide evacuees safely towards exits, they have shortcomings. They do not change if the evacuation routes become blocked or hazardous. In addition, if too many evacuees decide to take the same escape route, it leads to congestion and overcrowding.

To mitigate the problems that arise during crisis evacuation, research is being conducted on how personal electronic devices—such as smart phones—equipped with sensors can be applied for management of such situations [1, 12, 5]. Their built-in sensors and communication technologies can both gather information and share it among devices [8], and the aggregate of this information can contribute to clarify the current situation.

By leveraging this kind of real-time information, an automatic evacuation planning system can help resolve some of the challenges faced during a crisis situation, namely how to avoid casualties, congestion, and confusion. It can automatically determine escape routes for everyone present, taking care to lead evacuees away from hazardous situations, and avoid congestion by taking into account all passengers and their respective escape routes.

1.1 Problem Formulation

Fast, efficient and safe evacuation is important during crisis situations. Whereas current approaches to evacuation planning include pre-planned routes, a benefit could be had from providing real-time evacuation planning. Pre-evacuation planning is limited in that it cannot take into account the particularities of crisis situations as they happen; consequently, evacuation operation can be inefficient and unnecessarily dangerous.

Furthermore, it is of interest to take more of human behavior into account than has been done in related work. Specifically, group affiliation is an important aspect of human life, and it affects the evacuation process.

In line with common practice [6] we treat the escape as a time expanded directed graph of nodes and edges: $G(N, E)$. For this, any node $n_i \in N$

- is either a room (source), lifeboat (sink), super source or super sink,
- holds zero, one or more evacuees without exceeding its capacity $c(n)$, and
- has a survivability $\sigma(n)$ so that $\sigma(n) \in [0, 1]$ indicating the probability of survival for one time step.

Any potential flow from node n_i to node n_j is represented with an edge $e_{i,j} \in E$. A search space s represents a solution, i.e. paths consisting of edges and nodes, for every evacuee.

Additionally, an edge $e_{i,j}$ has capacity $c(e_{i.j})$ and flow value $f(e_{i,j})$. Further this paper extends the common terminology with node congestion $con(n_i)$. While the capacity $c(e_{i,j})$ reduces the flow of which a quantity moves from node n_i to n_j, $con(n_i)$ limits the number of people that can fit inside a room n_i. This is a realistic extension as there are practical limitations to how many people can be in a room at the same time.

The problem relies upon two main functions, namely the overall survivability of a search space, s:

$$f(s) = \sum_{n_i \in s} (\sigma(n_i)) \tag{1}$$

and the overall grouping of a search space, s:

$$g(s) = \sum_{e_{i,j} \in s} \tau(e_{i,j}, s) \tag{2}$$

where

$$\tau(e_{i,j}, s) = \begin{cases} 1 & \textit{if } e_{i,j} \textit{ occours at least twice in } s \textit{ for the same group} \\ 0 & \textit{otherwise} \end{cases} \tag{3}$$

The latter can be informally written as counting the number of edges overlapping within each group. An overlap occurs whenever at least two people use the same path at the same time.

Consequently, this paper has two objectives in prioritized order:

1. Maximizing the survivability: This is formalised, in line with common practice [5] as finding a search space $s^* \in \mathbf{S}$ so that $f(s^*) \geq f(s) \forall_s \in \mathbf{S}$. I.e. maximizing the probability that persons survive in the path chosen for them.
2. Maximizing the grouping: Finding a search space $s^* \in \mathbf{S}$ so that $g(s^*) \geq g(s) \forall_s \in \mathbf{S}$. I.e. maximizing the probability that groups stay together.

Note that we assume information such as physical layout of the ship, locations of people and lifeboats, affiliations and survivability in each room, is known. The authors realize that this may appear as an unrealistic assumption, but note significant effort is being made to collect and aggregate hazard information in similar scenarios from both stationary and smart phone sensors [12, 14].

1.2 Outline

Section 2 describes the related evacuation and evacuation planning. Section 3 continues with genetic algorithms specifically how it was used to solve the evacuation problem. Empirical results in a simulated ship environment are presented in section 4. Finally, conclusion and further work are presented in section 5.

2 Evacuation

Much work exists in the literature on evacuation modelling [11]. Common for evacuations is a five stage process [15]: (1) An alert is raised. (2) The persons present react to the alert. (3) A decision is made to evacuate. (4) The actual evacuation. (5) Verify that everyone has made it to safety.

Most existing work on evacuation planning focuses on off-line solutions aiding step (3) and (4). One of the main lines of research focus on mathematical modelling and solutions based on finding maximal flow in networks [4]. Other work has been based themselves on shortest path in a graph [9], while some recent research has been carried out for stochastic methods for planning safe escape routes [5].

2.1 Group Behavior in Crisis Situations

Groupwise evacuation is grounded in recent social theory. According to the "social attachment" model of human behavior during crisis situations [10], in threatening situations people tend to seek affiliation with familiar persons or attachment figures. This behavior delays the evacuation process; in fact, it has been shown to cause the loss of human lives because people linger together with their group or search for attachment figures instead of promptly evacuating. Evacuation planning without taking into account the strong force of group affiliation would be nigh on pointless, as it is unlikely that evacuees would follow a plan that required group members to go separate ways.

Furthermore, the social attachment model goes against earlier mass panic theories, which claim that chaotic human behavior is the norm when disaster strikes [3]. In contrast those earlier theories, the social attachment model describes evacuation as orderly in most cases. This certainly indicates a higher probability of evacuees displaying an ability to follow the dynamically planned routes than if they were panicking and behaving irrationally.

2.2 Evacuation Planning With Multi-objective Genetic Algorithms

Genetic algorithms (GA) have been used within the field of evacuation previously. In [13], the evacuation planning process is described as a three-part process which is performed as a preparatory measure for the case where actual evacuation is needed: Selecting safe areas, finding optimal paths from buildings to safe areas, and selecting the best safe area for each building is included in planning. The first step, selecting safe areas, was done manually. Next, the optimal paths from buildings to safe areas were determined according to safety and traffic. The last step, selecting the best evacuation routes for each building, was then solved with a genetic algorithm.

Kongsomsaksakul et al. [7] also consider predisaster evacuation planning. In their model, the problem is formulated as a Stackelberg game, where the leader is the evacuation planning authority designating shelter locations. The follower is the collection of evacuees, who according to the given shelter locations determine which shelter to move to and by which path.

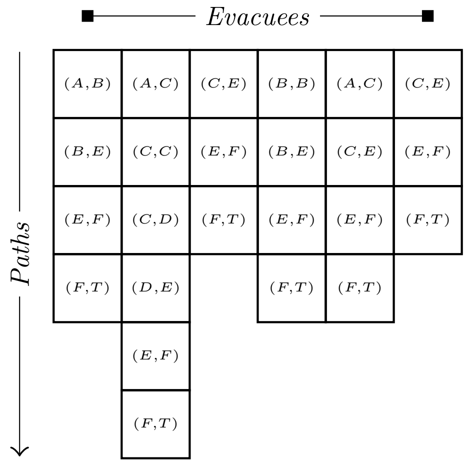
The GA is employed by the planning authority to place shelters. Given a potential solution from the leader, the evacuees decisions are calculated. The result is fed into the GA's fitness function, which is a weighted sum of constraints on egress time, congestion, and shelter capacity.

3 Genetic Algorithm (GA)

The solution bases itself on an initial population of solutions that is further improved by iterations. Solutions within a population are encoded as chromosomes.

The GA implementation uses NSGA-II [2], and was chosen based on promising performance characteristics documented in [16]. It depends on implementation specific components used in every iteration, namely: Selectors selecting pairwise solutions. Crossover combining pairs of solutions. Mutators randomly modifying each solution. Fitness functions used for distinguishing good from bad solutions.

The termination criterion is set to a fixed number



(A,B)	(A,C)	(C,E)	(B,B)	(A,C)	(C,E)
(B,E)	(C,C)	(E,F)	(B,E)	(C,E)	(E,F)
(E,F)	(C,D)	(F,T)	(E,F)	(E,F)	(F,T)
(F,T)	(D,E)		(F,T)	(F,T)	
	(E,F)				
	(F,T)				

Figure 1: A example chromosome containing an evacuation routing plan.

of iterations. After termination, the best solution must be selected from the population. The population contains solutions that are optimized for one or more of the objectives. The final solution is selected by the super selector.

3.1 Encoding of Chromosomes

A chromosome contains path assignments for each evacuee as shown in Figure 1. Note that this is directly related to a selected search path s.

In this example the network includes nodes A-E, sink F and super sink T . The elements of each path, e.g. (A, B), are edges. Each column is a path assignment to either a person or a group which is initially located in the first node in the path. Edges are used in the path to support cases where multiple edges connect two nodes, such as if two oblong, parallel rooms are connected with two or more doors.

3.2 Genetic Operation

The GA starts by selecting a pair of parent chromosomes using binary tournament selection [2] and then performing a crossover using these.

3.2.1 Crossover

A multi-point crossover operator for recombining a pair of two-dimensional chromosomes has been implemented, by using one-point crossover once for each group represented in the chromosome. The crossover point in each parent is randomly selected among potential, valid crossover points. For a crossover point at node n_i to be considered valid, an edge with a target node n_i must exist within the corresponding path in both parent chromosomes. However, the common node does not need to be traversed at the same time.

If the crossover operation creates an invalid child, one of the parents is passed as child instead. Because different children for different lengths are created, the path may extend beyond the time-expansion. By ensuring that the initial population is valid and only passing valid solutions as children, this problem will not occur. It is important to note that this crossover operator can take two identical parents and still produce distinct children. Usually, a chromosome which is recombined with itself will produce children that are perfect copies of itself, and hence with no possible improvement. However, due to the way chromosomes are encoded and crossover is implemented, in this case a parent mating with itself has the possibility of producing offspring which are different and may be better than its parent. This effect arises because edges, which are reusable at different time steps, can occur several times in the same chromosome, which can lead to a single chromosome having multiple time-shifted crossover points with itself. Nevertheless, offspring that are identical to their single parent will still occur if the exact same crossover point is used for both of the parents. Reuse of edges allows for two things. Firstly, it allows waiting in a node by following the holdover edge two or more times in sequence. Secondly, it also occurs when paths are circular. Regardless, such solution paths will be evaluated by fitness functions and handled accordingly.

Note that in a time-expanded network, crossover breaks the sequential timing of the path. To fix the timing of a path, a repair function is applied after the crossover. This adjusts the time of an edge so that the sequential timing is kept.

3.2.2 Mutation

The one-dimensional mutation changes each path of each chromosome in the population by a predefined probability. The mutation generates a new random path starting from the same origin and ending at the super sink.

3.3 Fitness Functions

This section presents two fitness functions motivated by the evacuation criteria (see section 2) :survivability and grouping. As a direct consequence, the fitness functions are used from the problem formulation, namely: Survivability, $f(s)$ (see equation 1) and Grouping, $g(s)$ (see equation 2).

3.4 Super Selector

Unlike traditional GA, NSGA-II does not yield a single solution which can be considered the best one. This is an intended effect of using Pareto ranking. Instead, the solutions present in the first Pareto front are the set of the best solutions which the algorithm could find. Because the solutions with the same rank are mutually non-dominant NSGA-II makes no assumptions as to which, if any, of the objective functions are more important.

Therefore, a single solution must be extracted from the set of solutions yielded by NSGA-II. This can be done manually, which can be suitable in a decision-support system. However, automating the process is often preferable, which can be accomplished by adding a final processing step for the set of solutions NSGA-II yields. This can be realized by using a selection mechanism which is able to rank the solutions, for instance by combining the fitness values in some way. Here we use a prioritized fitness ranking approach.

Prioritizing works as follows. Starting with the highest ranked fitness measure, all solutions' value for it is compared. If one solution has a strictly lower value than all others, then that solution is selected. Otherwise, the set of solutions with the lowest value are compared again, this time on the next-highest ranking fitness measure. This continues until a solution has been found. If all objective functions have been processed in this way and more than one solution are still candidates, the tie is broken arbitrarily.

The objective functions we use are ranked in the following order:

- Endpoint capacity
- Survivability
- Passage congestion
- Room congestion
- Length

The ordering has been determined through informal reasoning. First, we definitely want every evacuee to be assigned to a lifeboat which has room for him or her. This is the highest ranked objective, seeing as failing to accomplish this is considered a hard failure (certain fatality). Second, we want to minimize the time spent in dangerous spaces, measured by survivability. Next comes congestion, which can influence actual survivability and cause evacuees to fail at following their assigned routes. Path length is selected last, because even thought it is desirable to have the shortest paths possible, it is less important than the other objectives.

4 Experimental results

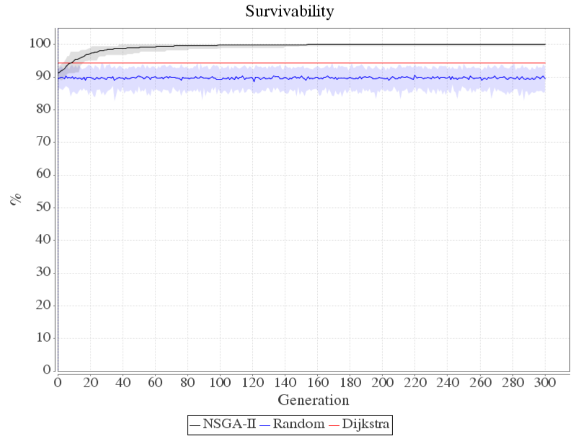
This section provides results from running the algorithms NSGA-II, Djikstra and Random in a network based on the layout of a deck on a real ship, the MS Xpedition owned by Celebrity Cruises. This has in total 40 rooms and 4 lifeboats. For all experiments evacuees are randomly divided into groups and randomly spread out in the network. Survivability is distributed randomly between 0.8 and 1.0.

All results are aggregated averages of 50 iterations with the confidence of 95%.

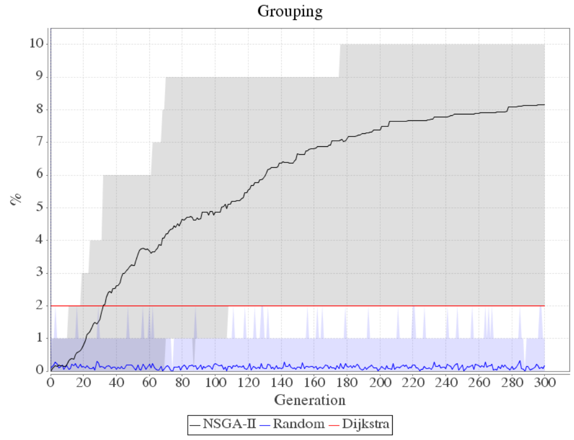
Figure 2(a) and 2(b) show results from survivability for 20 evacuees in a situation with life boat capacity of 5, and random probability of surviving varying from 0.8 to 1.0. The evacuees were divided into random groups of 1-5.

The results from Figure 2(a) indicates that NSGA-II finds the optimal solution (100% survivability), when there is no grouping, after 118 iterations. Djikstra is only able to survive 94% of the evacuees, while random is able to evacuate 90%.

The same experiment is present at Figure 2(b) and shows the grouping of each algorithm. It is



(a) Survivability — $f(s)$



(b) Grouping — $g(s)$

Figure 2: Experiment results in a realistic ship network

noteworthy that the grouping continues to improve throughout the experiment.

A conclusion to be drawn from this is that the approach finds an optimal solution with respect to survivability. Further, without have any impact on the survivability, the approach continues to optimize on the grouping. This is in line with the intent (see section 1.1) that primarily optimises on survivability than continues optimisation on grouping.

4.0.1 Effects of Congestion Heuristic

In Figure 3(a) and , 3(b) the algorithms have been applied with the same parameters. However, in 3(a) NSGA-II was run without the congestion heuristics, while it was present in the experiment presented in 3(b). The effect this has can be seen clearly: When not optimizing for congestion, the criterion is neglected and increases as the genetic algorithm progresses. Congestion even approaches the value of the Random algorithm. Conversely, when congestion is optimized for, the genetic algorithm continually improves the congestion performance, albeit slowly. The results also show that the survivability improves correspondingly. However, for reasons of brevity, this graph is not presented here.

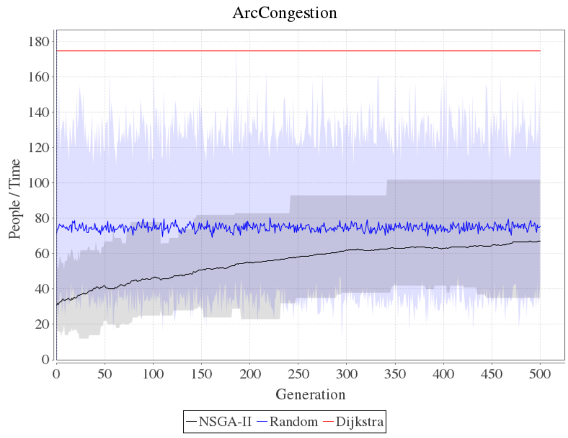
4.0.2 Extensive Experiments

Additional work has been done in both simple test networks and larger randomly generated graphs. The results from these graphs are similar to those presented in this chapter, with mostly negligible differences. The only difference worth mentioning is that randomly generated graphs yields a more difficult optimisation problem which decreases the performance of NSGA-II compared to Djikstra. Hence, NSGA-II result in a lower survivability than Djikstra

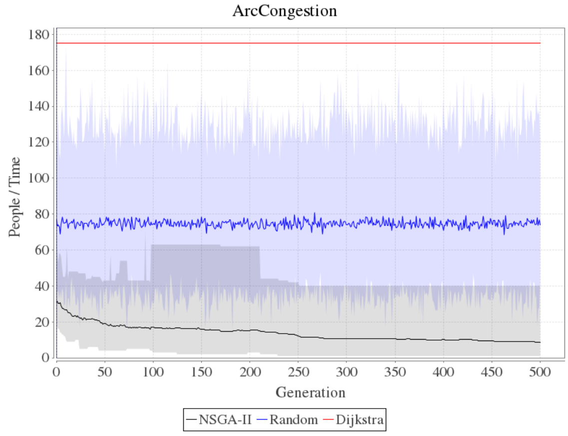
5 Conclusion

The problem of efficient evacuation can be viewed as an optimization problem, for which many techniques have been developed, including genetic algorithms.

NSGA-II is an adaption of the genetic algorithm framework which supports the preservation of diversity among candidate solutions by taking into



(a) Without congestion heuristics.



(b) With congestion heuristics.

Figure 3: Experiment results with and without congestion heuristics.

account Pareto indifference, meaning that no solution is strictly better than others. The technique presented in this paper bases itself on a

The technique we developed to select a solution from the multiple solutions returned by NSGA-II is a prioritized objective approach.

The results clearly show the potential genetic algorithms can have in evacuation planning. In fact, we found that in some simpler scenarios was able to find a solution for most of scenarios that outperformRandom and Djikstra's.

Our results also show that when the fitness functions becomes more complicated, such as considering congestion, the efficiency of the algorithm suffers. Hence, when the complexity increases more generations are needed.

5.1 Future Work

Future work includes making more specifically adapted genetic operators such as mutation operators which take into account the grouping aspect. The inherent complexity of the chromosome is likely a hurdle which needs to be overcome. Due to the way the chromosome is defined, very specific constraints are applied to it which limits the effectiveness of the genetic operators, compared to traditional genetic algorithms. The limitations are related to the way each part of the chromosome must be a valid path specification.

References

[1] Maged N Kamel Boulos, Bernd Resch, David N Crowley, John G Breslin, Gunho Sohn, Russ Burtner, William A Pike, Eduardo Jezierski, and Kuo-Yu Slayer Chuang. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, ogc standards and application examples. *International journal of health geographics*, 10(1):67, 2011.

[2] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197, 2002.

[3] Thomas E Drabek and David A McEntire. Emergent phenomena and the sociology of disaster: lessons, trends and opportunities from the research literature. *Disaster Prevention and Management*, 12(2):97–112, 2003.

[4] Lester R Ford and Delbert R Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8(3):399–404, 1956.

[5] Morten Goodwin, Ole-Christoffer Granmo, and Jaziar Radianti. Escape planning in realistic fire scenarios with ant colony optimisation. *Applied Intelligence*, 42(1):24–35, 2015.

[6] H.W. Hamacher and S.A. Tjandra. Mathematical modelling of evacuation problems–a state of the art. *Pedestrian and Evacuation Dynamics*, 2002:227–266, 2002.

[7] Sirisak Kongsomsaksakul, Chao Yang, and Anthony Chen. Shelter location-allocation model for flood evacuation planning. *Journal of the Eastern Asia Society for Transportation Studies*, 6(1):4237–4252, 2005.

[8] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

[9] Gino J Lim, Shabnam Zangeneh, M Reza Baharnemati, and Tiravat Assavapokee. A capacitated network flow optimization approach for short notice evacuation planning. *European Journal of Operational Research*, 223(1):234–245, 2012.

[10] Anthony R Mawson. Understanding mass panic and other collective responses to threat and disaster. *Psychiatry: Interpersonal and biological processes*, 68(2):95–113, 2005.

[11] Adam J Pel, Michiel CJ Bliemer, and Serge P Hoogendoorn. A review on travel behaviour modelling in dynamic traffic simulation models for evacuations. *Transportation*, 39(1):97–123, 2012.

[12] Jaziar Radianti, Ole-Christoffer Granmo, Noureddine Bouhmala, Parvaneh Sarshar, Anis Yazidi, and Jose Gonzalez. Crowd models for emergency evacuation: A review targeting human-centered sensing. In *In Proceeding of 46 th Hawaii International Conference for System Sciences*, 2013.

[13] Mohammad Saadatseresht, Ali Mansourian, and Mohammad Taleai. Evacuation planning using multiobjective evolutionary optimization approach. *European Journal of Operational Research*, 198(1):305–314, 2009.

[14] Parvaneh Sarshar, Jaziar Radianti, and Jose J Gonzalez. On the impacts of utilizing smartphones on organizing rescue teams and evacuation procedures. In *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management, ISCRAM*, pages 24–27, 2015.

[15] James MacGregor Smith. Evacuation networks. In *Encyclopedia of optimization Ed. by Christodoulos A. Floudas and Panos M. Pardalos.*, pages 940–950. Springer, 2009.

[16] Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary computation*, 8(2):173–195, 2000.

Deep Learning for Social Media Analysis in Crises Situations

Mehdi Ben Lazreg   Morten Goodwin   Ole-Christoffer Granmo

University of Agder, Grimstad, Norway

Abstract

Social media has become an important open communication medium during crises. This has motivated much work on social media data analysis for crises situations using machine learning techniques but has mostly been carried out by traditional techniques. Those methods have shown mixed results and are criticised for being unable to generalize beyond the scope of the designed study. Since every crisis is special, such retrospect models have little value. In contrast, deep learning shows very promising results by learning in noisy environments such as image classification and game playing. It has, therefore great potential to play a significant role in the future social media analysis in noisy crises situations. This position paper proposes an approach to improve the social media analysis in crises situations to achieve better understanding and decision support during a crisis. In this approach, we aim to use Deep Learning to extract features and patterns related to the text and concepts available in crisis related social media posts and use them to provide an overview of the crisis.

Keywords: deep learning, social media, crises situations

1 Introduction

A vast variety of natural and human-caused crises occur around the world. The diversity and immediacy of these crises cause severe challenges not only for the people affected and responders, but also to the research community. Some of the unsolved research challenges include:

- How can machine learning be used to detect a crisis as soon as it occurs from external sources including social media?
- How can machine learning use crisis related social media data to acquire information about a crisis' status and progression?
- How can artificial intelligence support the responders in making the correct decisions during a crisis?

Social media plays a pivotal role in most crises today, from getting life signs from people affected to communicating with responders. However, most research using social media in crises situations are one-off solutions with a specialised technique or addressed area [1]. A one-off solution means that the research focuses on finding a technique that yields the best result in a specific crisis. The technique fails to generalize beyond the study. This finding has several implications due to the distinct nature of crisis. Firstly, crises are diverse, and range from natural, technological, financial to political crises, and even previously unforeseen types of crises. Secondly, crises evolve through time. Different aspects of a crisis change as time passes, which means what is learnt earlier on may not be applicable later in the crisis. Finally, crises are unpredictable in nature as unexpected event may occur.

Deep Learning (DL) has the potential to improve social media analysis in crises situations because of its ability to learn patterns from unlabelled data [17]. This property has enabled DL to produce breakthroughs in the domain of image, text and speech recognition. Moreover, DL has the ability to generalize learnt patterns beyond data similar to the training data, which can be advantageous while dealing with social media analysis in crises situations. Despite the breakthroughs brought by DL, improvements are still to be made the further optimise it and improve its performance [14]. This paper proposes to explore the uninvestigated area of how the emerging advantages of DL can be expanded upon to address the pertinent challenges of evolving crisis analytics for social media.

This paper is organised as follows. In section 2 we discuss the use of social media in crises situations with a special focus on the use of machine learning. Section 3 continues with Deep Learning, and section 4 proposes an approach for applying Deep Learning to for social media analysis. Finally, section 5 concludes.

2 Machine Learning in Social Media Analysis in Crises Situations

Social media has become an open crises communication medium. As an example, during the tsunami in the Philippines in 2012, 558126 tweets were produced in 8 different languages in the course of the seven first hours following the crises [2]. Similarly, 20000 tweets/day were registered midst the 2012 Sandy storm in New York, and 5000 tweets/second were reported during the 2011 earthquake in Virginia (US) [2].

There is no doubt that valuable, high throughput data is produced on social media only seconds after a crisis occurs. However, processing and inferring valuable knowledge from such data are difficult for several reasons. The messages are typically brief, informal, and heterogeneous (mix of languages, acronyms, and misspellings) with varying quality, and it is often required to know the context of the message to understand its meaning. Moreover, posts on other mundane events are also part of the data, which introduces additional noise for training. To address the challenges of detecting and classifying a crisis in heterogeneous data, supervised and unsupervised machine learning techniques were used.

2.1 Supervised learning

In order to classify a social media message as part of one particular crisis event, several features related to the message need to be used, including the nature of the message (factual, emotional or subjective), the information provided, the information source, credibility, time and location. Note that some of the features can be automatically extracted, but others need human labelling.

From the message examples, the supervised learning algorithm learns a predictive function (representing the relation between the features and particular crises) so that it can classify any new unknown message as part of one of the categories of crises. Several approaches have been applied for this: Naïve Bayes and Support Vector Machine (SVM) [5][6], Random Forests [7], and Logistic Regression [8]. Further, to mitigate the complexity of social media messages, some research focuses on only analysing tweets with certain tags [4]. As example, the tag “breaking news” and “news” can be used to identify breaking news tweets [9]. In the same way, the occurrence of the words “earthquake” and “landslide” were used as features in a SVM classifier to classify earthquakes [8] and landslides [10] respectively.

However useful in reducing the complexity of the data, this approach neglects the potential valuable information contained in the text such as the status of the crisis, potential victims, needed resources, and so on. In a supervised approach, (human made) labels are necessary for training the classifiers, but they might be highly difficult to obtain, especially in case of multi-language messages or context knowledge [4]. Furthermore, such labels are not always reliable and may not be available at the time of the crisis. Moreover, reusing a classifier trained on data from previous disasters may not perform well in practise and intuitively returns a loss of accuracy even if the crises have a lot in common. To solve this issue research has been carried out to use unsupervised learning techniques.

2.2 Unsupervised learning

Unsupervised methods are used to identify patterns in unlabelled data. They are most useful when the information seekers do not know specifically what information to look for in the data – which is the case in many crises situations. An example is grouping tweets into stories (clusters of tweets) after a keyword filter [11]. This method reduces the number of social media messages to be handled by humans since it groups equivalent messages together. Another application using unsupervised learning identifies events related to public and safety with a spatio-temporal clustering approach

[12]. In addition to strictly clustering elements into groups, soft clusters have been used to allow items to simultaneously belong to several clusters with variant degrees. In this methods, the tweets similarity is based on words they contain and the length of the tweets [13]. The approach was applied on data from the Indonesia earthquake (2009) and it detected aspects related to the crisis (relief, deaths, missing persons, and so on).

3 Deep Learning

Imitating the efficiency of the human brain has been a huge challenge for the artificial intelligence field. The emergence of DL has fuelled a paradigm shift and made it a more achievable goal. DL is a machine learning technique with its roots in Neural Networks that allows learnt models composed of multiple processing layers so that the knowledge state has several layers of abstraction. DL is particularly valuable because it is shown to find complex structures in large data using an algorithm to update its internal representation of each layer in a way that other state-of-the-art algorithms are not [14]. Conventional supervised machine learning techniques require careful engineering to transform raw data into suitable features for classification of inputs, whereas, DL fed with raw data, discovers the representation and features needed for detection and classification. DL techniques with back propagation and deep convolutional nets have brought breakthroughs in image processing while recurrent nets have brought amazing advancements in sequential data analysis such as text and speech recognition [15]. DL is used to analyse X-ray images to detect potential diseases [16], and to recognise handwriting [17]. DL is applied to online tasks perhaps most notably, it was the first AI machine to beat a human expert in the game of Go – by most AI scholars considered one of the most complex games for artificial intelligence [18]. Further, DL is successfully applied to text mining to organise text documents in databases by topic [19], to analyse costumers review on a given product and deduce what they think about it [20]. Moreover, it has been explored in chemical text mining to recognise drugs and chemical compounds [21], and sentiment analysis [22]. However, DL has to a very little degree been explored for crises management [4]. Despite the breakthroughs brought by DL, using DL for unsupervised learning has not been much explored and was for a long time overshadowed by the success of supervised learning [14]. Unsupervised learning is important to explore since, in that paradigm, the AI machines discover structures by observing the data without being told what each feature in the data represents.

4 Social Media Analysis

This paper propose an approach to improve the social media analysis in crises situations to achieve better understanding and decision support during a crisis. The approach is summarised in Figure 1 and consists of moving from low to high level of abstraction. We plan to proceed from: Using DL to transform non-standard words into their canonical form (1). Then, understand the semantic of the text (2). Finally, provide an overview of a crisis development (3).

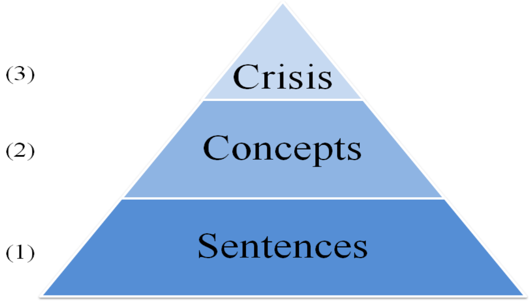


Figure 1: Overview of the proposed approach

4.1 Analysis on words and sentences

The machine needs to recognise words in the sentence to understand it. In social media text this task is challenging since the messages typically contain misspellings, abbreviations, deletions, and phonetic spellings. Traditional supervised learning approaches are extremely dependent on the correctness of the training set. To learn diverse functionality, training sets that represent each category of the data are required. The number of ways a word can be misspelled is huge, which means that traditional approaches fall short [23]. Unsupervised approaches can find similarities between spelling variations of words using clusters containing the correct word as well as its different misspellings. Nevertheless, unsupervised learning also falls short because they depend on rigid metrics for similarity that influence the clusters [23].

To address this challenge, we plan to use DL to deduct high-level abstractions (e.g. meaning) from lower levels (e.g. letters or sub-string). A new configuration at low abstraction level may then lead to a better representation of words and similarities between different spellings. The assumption is that even with limited training sets, a new example could be meaningfully represented using the low-level abstractions. We will look at ways to modify the basic DL algorithm so it can detect features that relate non-standard versions of words or phrases, which will be used to deduct the correct text. For this first phase, we will train and test the developed algorithm on collected texts from crisis events on Twitter. The text contains ungrammatical sentences, non-standard and misspelled words. The algorithm will be evaluated based on correctness metrics and error rates of non-standard words that the algorithm fails to recognise. Due to DL ability to be applied in noisy environments, we predict that this approach will outperform state-of-the-art in the field of text mining.

4.2 Identify and understand concepts

After word and sentence analysis is carried out, the proposed approach moves to a higher level of abstraction: Classifying concepts from multiple social media messages. Explicitly, this means identifying what a writer tries to express (e.g. informing about a situation, crying out for help, express explicit needs, and so on) and understanding concepts from the message. The state-of-the-art in this area mostly centres on supervised learning techniques by training the algorithm on a set of text on each topic to learn a predictive function, which in turn is used to classify a new topic into a previously learnt topic [24]. A limitation of this approach is the scope of predefined topics: If a text about an unforeseen topic is presented to the algorithm, such as a new crisis, it will wrongly classify it as one of the existing topics. A challenge is that crises are diverse, and the number of topics discussed in social media during a single crisis is big, dynamic, and changing. The complexity of social media data makes getting human labelled data for each topic very expensive and time-consuming. Adversely, unsupervised techniques try to look for co-occurrences of terms in the text as a metric of similarity [26] and inferring the word distribution in the set of word the text contains and using their frequencies for document clustering [27].

Using DL to understand a sentence or a document is an ongoing research topic in which progress is still to be made. DL has been used to predict the next word on a sequence of semantically related words [14], and this ability suggests that DL have learnt a semantic representation of the words. DL also has some success in predicting the next character in a sequence of characters which is used to generate text, and in machine translation [14]. To address this challenge we aim to use DL on the normalised text to automatically learn distinct features for each discussed concept. Further, we intend to find a strategy to decompose a document into concepts, segmenting the text into semantically meaningful atomic units. By identifying the underlying concepts of a document or a sentence, a deeper understanding is established. This lies the foundation for building crisis understanding. We will investigate ways to improve unsupervised DL for concept discovery. A semi-supervised method can be used to improve the performance of the data representation. When the model is actually able to represent the unlabelled data, labels can be added to transform the problem from an unsupervised to a supervised learning problem. For this model, we will gather data from social media platforms on crises events, including Twitter, transform the noisy data into a workable form using the approach described in Section 4.1, and use this data to train and test our model. We will empirically verify the result of each test, which will help us understand how the model performed. We will base our evaluation on established correctness metrics and error rates of topic or concepts that the algorithm fails to recognise.

4.3 Crisis understanding

Even though the crisis data is valuable with high throughput, it is small compared to the 7Gb/min of data produced by Twitter alone [3]. Hence, this topic integrates two areas: DL and big data (large and complex datasets that cause traditional data processing application to be inadequate). The aim of this phase is to detect crisis patterns in social media text that can be used to retrieve crisis related messages from big social media data in a way that it gives an overview of its status.

The state-of-the-art in the use of machine learning on social media in a crisis (see Section 2) are one-off solutions with specialised techniques or addressed areas, including the labels needed for training the classifiers which are not always reliable or available. Also, reusing a classifier trained on data from previous disasters may not perform well in practise and intuitively returns a loss of accuracy even if the crises have a lot in common.

One of the key features of DL is the analyses of a big amount of unsupervised data [25], which makes it valuable for big data analytics with unlabelled and uncharacterized data. DL can be used to address the important problems including extracting patterns from massive data, and information retrieval. It can provide a generic solution that infers similarity and dissimilarity patterns between different crises. Nonetheless, DL algorithms can become computationally expensive when dealing with high dimensional data due to its deep layered hierarchy and number of parameters to learn. The computing expensiveness becomes more of a problem in social media where the data is streaming rapidly and changing fast. Methods for incremental learning have been developed to deal with this challenge that includes the use of DL [28]. To address this challenge question, we will use and expand incremental DL to infer information about rare events (crises data) in a mix of a massive and diverse data which, in this stage, include the concepts learnt previously and metadata (including time, location and writer of a tweet). The result will be presented in a spatiotemporal overview of the crisis. A spatiotemporal overview presents the statue of crisis in a location at different points in time, an approach which to very little degree has been investigated [4]. The overview will be used as a decision support system to help take the most appropriate actions to resolve the crisis. We will present the model with a set of crisis practitioners that will test it and provide inputs (in the form of a survey) on how helpful this model would be in the case of crisis. Correctness metrics and error rates of the algorithm will also inform of its abilities.

5 Conclusion

This position paper presents an overview of machine learning techniques used for social media analysis in crises situations today. The current approaches, based on traditional machine learning techniques, are heavily criticised for being one-off studies which cannot be generalized. Since every crisis is special, such retrospect models have little value.

Deep Learning (DL) has the potential to mitigate this problem since it has been shown to very good at generalizing. The paper presents a possible approach for applying DL to crises analysis. The model starts with normalizing social media data. This includes mapping noisy words to the original word. Further, the normalized text can be used to deduce the concepts and the topic of a cluster of texts. Finally, the texts related to a crisis situation are retrieved and a spatiotemporal representation of the crises of the crises is produced based on those texts. In this way, DL can be used to offer a decision support system to crises responders to help them better understand a crisis situation and produce more efficient decisions than traditional machine learning techniques.

References

[1] Maresh-Fuehrer, M.M. and R. Smith, Social media mapping innovations for crisis prevention, response, and evaluation. Computers in Human Behavior, 2016. 54: p. 620-629.

[2] Zielinski, A., et al. Social media text mining and network analysis for decision support in natural crisis management. in ISCRAM 2013 Conference Proceedings - 10th International Conference on Information Systems for Crisis Response and Management. 2013.

[3] Valkanas, G., et al. Mining Twitter Data with Resource Constraints. in Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on. 2014.

[4] Imran, M., et al., Processing Social Media Messages in Mass Emergency: A Survey. CoRR, 2014. abs/1407.7071.

[5] Yin, J., et al., Using social media to enhance

emergency situation awareness. IEEE Intelligent Systems, 2012. 27(6): p. 52-59.

[6] Power, R., et al., Emergency situation awareness: Twitter case studies, in Information Systems for Crisis Response and Management in Mediterranean Countries. 2014, Springer. p. 218-231.

[7] Imran, M., et al. Aidr: Artificial intelligence for disaster response. in Proceedings of the companion publication of the 23rd international conference on World wide web companion. 2014. International World Wide Web Conferences Steering Committee.

[8] Ashktorab, Z., et al., Tweedr: Mining twitter to inform disaster response. Proc. of ISCRAM, 2014.

[9] Phuvipadawat, S. and T. Murata. Breaking news detection and tracking in Twitter in Web Intelligence and Intelligent Agent Technology (WI-IAT). IEEE/WIC/ACM International Conference on. 2010.

[10] Musaev, A., D. Wang, and C. Pu. LITMUS: Landslide detection by integrating multiple sources. in 11th International Conference Information Systems for Crisis Response and Management (ISCRAM). 2014.

[11] Rogstadius, J., et al., CrisisTracker: Crowdsourced social media curation for disaster awareness. IBM Journal of Research and Development, 2013. 57(5): p. 4: 1-4: 13.

[12] Berlingerio, M., et al. SaferCity: a system for detecting and analyzing incidents from social media. in Data Mining Workshops (ICDMW), 2013.

[13] Kireyev, K., L. Palen, and K. Anderson. Applications of topics models to analysis of disaster-related twitter data. in NIPS Workshop on Applications for Topic Models: Text and Beyond. 2009. Canada: Whistler.

[14] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. Nature, 2015. 521(7553): p. 436-444.

[15] Deng, L., et al. Recent advances in deep learning for speech research at Microsoft. in Acoustics, Speech and Signal Processing (ICASSP), 2013.

[16] Anavi, Y., et al. A comparative study for chest radiograph image retrieval using binary texture and deep learning classification. 37th Annual International Conference of the IEEE in Engineering in Medicine and Biology Society (EMBC), 2015.

[17] Arel, I., D.C. Rose, and T.P. Karnowski, Deep Machine Learning - A New Frontier in Artificial Intelligence Research. IEEE Computational Intelligence Magazine, 2010. 5(4): p. 13-18.

[18] Silver, D., et al., Mastering the game of Go with deep neural networks and tree search. Nature, 2016. 529(7587): p. 484-489.

[19] Kalaimani, E., E. Kirubakaran, and P. Anbalagan, Classification and topic discovery of web documents using semantic deep learner based on semantic smoothing model. International Journal of Applied Engineering Research, 2015. 10(2): p. 2575-2598.

[20] Lu, Y., et al., CHEMDNER system with mixed conditional random fields and multi-scale word clustering. Journal of Cheminformatics, 2015. 7.

[21] Wang, Y., et al., Word vector modeling for sentiment analysis of product reviews, in Communications in Computer and Information Science. 2014. p. 168-180.

[22] Hailong, Z., G. Wenyan, and J. Bo. Machine learning and lexicon based methods for sentiment classification: A survey. in Proceedings - 11th Web Information System and Application Conference, WISA. 2014.

[23] Acharyya, S., et al., Language independent unsupervised learning of short message service dialect. International Journal on Document Analysis and Recognition (IJDAR), 2009. 12(3): p. 175-184.

[24] Aggarwal, C.C. and H. Wang, Text Mining in Social Networks, in Social Network Data Analytics, C.C. Aggarwal, Editor. 2011, Springer US: Boston, MA. p. 353-378.

[25] Najafabadi, M.M., et al., Deep learning applications and challenges in big data analytics. Journal of Big Data, 2015. 2(1): p. 1-21.

[26] Dumais, S.T., Latent semantic analysis. Annual review of information science and technology, 2004. 38(1): p. 188-230.

[27] Blei, D.M., A.Y. Ng, and M.I. Jordan, Latent dirichlet allocation. the Journal of machine Learning research, 2003. 3: p. 993-1022.

[28] Zhou, G., K. Sohn, and H. Lee, Online incremental feature learning with denoising autoencoders. in International Conference on Artificial Intelligence and Statistics, 2012

Center for Applied Intelligent Systems Research (CAISR)

Thorsteinn Rögnvaldsson Antanas Verikas Josef Bigun
Sławomir Nowaczyk Anita Sant'Anna Björn Åstrand
Jens Lundström Stefan Byttner Roland Thörner
Fernando Alonso-Fernandez Martin Cooney Rafael Valencia *

Abstract

Awareness is a broad concept, just like "intelligence", and has many connotations. This paper presents the vision of researchers from Center for Applied Intelligent Systems Research (CAISR) at Halmstad University.

1 Aware systems research definition

Awareness is a broad concept (just like "intelligence") and has many connotations. Relating to research on computing systems, there are two directions: one regarding the logical definition of aware and how to determine if a system is aware or not, another regarding what is required to be aware, i.e. what capabilities are required to be aware. In CAISR we focus on the latter.

Examples along the first direction can be found in, e.g., the papers by Hintikka (1975), Fagin and Halpern (1988) or Modica and Rustichini (1994). They boil down to statements like "Awareness of ϕ if they explicitly know ϕ or they explicitly know they don't explicitly know ϕ" over enumerations of possible worlds, where ϕ is a logical statement that can be true or false. As argued by Devanur and Fortnow (2009), such definitions are of little practical use since there is never endless time to search through all possible objects. It is also relevant to ask if awareness is a purely binary concept. Human awareness works quite differently: we are more aware of recent facts than old facts, even though we know them all. Devanur and Fortnow (2009) suggest a more practical and human-like definition: that awareness of an object is inversely proportional to the time needed to enumerate that object in a certain environment and a context. If you cannot do this within a certain time then you are effectively unaware of the object.

The work by Endsley (1995) is central regarding the second direction: what is required to be aware. Endsley (1995) describes awareness, from a human psychological perspective, as knowledge created through interaction between an agent and its environment, and "knowing what is going on" (see Gutwin and Greenberg, 2002, and references mentioned therein). Similarly, but more recently, Zhao et al. (2012) state, from a computing perspective, that awareness is the ability to perceive, to feel, or to be conscious of events, objects, or sensory patterns, but it may not lead directly to full comprehension. Zhao (2013) refines this into awareness being "a mechanism for obtaining information or materials which are useful for human users, for other systems, or for other parts of the same system, to make decisions". Zhao (2013) further comments that computationally aware systems have been studied for a long time but are often classified based on the event to be aware of. Examples include, context aware, situation aware, intention aware, preference aware, location aware, energy aware, risk aware, chance aware, and so on. Such classifications are not helpful to explore the key properties of aware systems since it divides aware systems research into many subsystems with unclear boundaries between them.

In CAISR we do not pursue research on what it logically means to be aware. We follow the direction of Endsley (1995) and Zhao (2013) and define aware systems research as: *Research on the design of systems that, as autonomously as possible,*

---

[*] Authors are at Center for Applied Intelligent Systems Research, Halmstad University, Sweden. Email addresses follow *firstname.lastname@hh.se* pattern

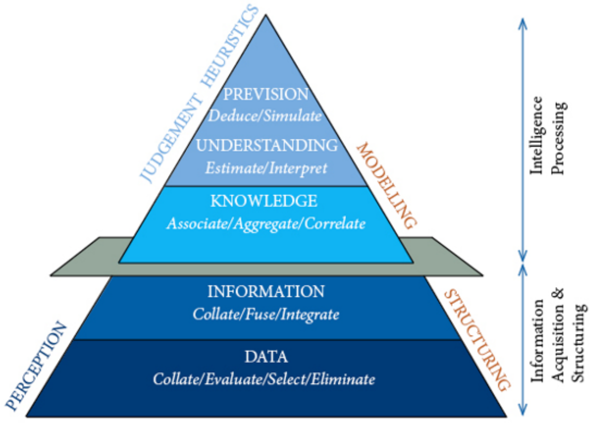


Figure 1: The knowledge pyramid.

can construct knowledge from real life data created through the interaction between a system and its environment. This data necessarily includes streaming data. Such systems should be able to handle events that are unknown at the time of design.

The goal with artificial intelligence (AI) research and development is to construct systems that behave intelligently. However, it is standard to assume that human experts define the task that the system should perform and that the collected data used for building systems reflect the "reality". This means that these systems are "designed" or "programmed", which leads to systems that break when the context changes. Our aim is to approach the construction of AI systems that can do "life-long learning"; systems that require less supervision and handle surprising situations. In order to do so, the systems must become more "aware" and able to learn on their own.

The construction of knowledge (going from perception to knowledge) is often represented by the knowledge pyramid (Ackoff 1989), see Figure 1; the higher a system reaches on the pyramid, the more knowledge it has and the more aware it can be. A version of this structure and knowledge pyramid is also how NASA looks at intelligent monitoring of missions (Degani et al., 2009). A fully aware system will have interaction both upwards and downwards in the pyramid, e.g. events higher up in the pyramid will affect choices on what data to collect.

Much of the work on machine learning (ML) and AI has not considered the knowledge creation aspects of intelligent systems. The common approach is to have humans define the problem in significant detail, for example the data characteristics, the representations used, the model used, etc. and the task is to build a machine that replicates the human decision. There are therefore many open research challenges for each of the stages in the knowledge pyramid:

Data: This deals with the collection of data and the representation of it, answering the question "with what" (Zhao 2013). An open question is how an autonomous system should select what data to collect? With streaming data from all sorts of sources, and with data bases of varying quality, how can the system tell what data are (or will be) relevant? The "with what" decision is to a large part done by humans today, which simplifies the learning problem immensely, but it is clearly one of the most relevant questions for autonomous learning. A related, much more researched, question is how to create general features; features that will apply to many problems (e.g. invariant features in images). Furthermore, with endless streams of data (i.e. in the "internet of things" era) it is impossible, even uninteresting, to save all data. It should be possible to save snapshots, compressed, or aggregated representations of the data. These representations should be learned and be general so that they apply to many different tasks. The fact that the working environment of a system may change frequently, requires considering the plasticity-stability problem seriously (Zhao 2013); features that look unimportant today may end up being important tomorrow. A system that is aware should therefore be curious and never stop exploring.

Information: This relates to questions that begin with "who, what, when and how many" (Rowley 2007), creating "events" from the data in the layer below. Examples of operations that are required for this are classification, rearranging/sorting, aggregating, performing calculations, and selection (see e.g. Curtis and Cobham 2005). Much ML research (including that on deep learning models) has been devoted to this stage, and also AI research for e.g. text and language parsing. Important open research questions here regard autonomous clustering and categorization of events. How can events be grouped into categories, e.g. common or uncommon, normal or odd, for later use? A challenge is to do this autonomously, or with only limited interaction with a human (that can provide hints), and in non-stationary environments. There is a significant body of relevant research on deviation detection, change detection and autonomous clustering.

Knowledge: This level is about creating "rules" from the information (rules can be in the form of models and not necessarily in e.g. predicate knowledge form). This always requires combining information from different sources. For example, is an observed "event" from one set of data sources associated with some other event, and can such associations be formulated into rules (and are these rules correct)? One obvious example is the supervised learning setting, where information "events" (input) are matched to correct responses (target) provided by a human expert and encoded into a rule (model). A very relevant research question here is how human generated knowledge, e.g. in the form of text comments in human curated data bases or models of the environment, can be combined with the machine generated information to create rules. Horeis and Sick (2007) have presented one example for incorporating human experts in this process. Another question deals with knowledge representations (knowledge structures); how can knowledge be represented so that it can be used for reasoning and prediction? A set of well-defined, highly-organized yet dynamic knowledge structures is one prerequisite for achieving awareness. A knowledge structure should evolve over time from experience, thus allowing for learning from data and human experts and be capable of taking into account different kinds of initial domain knowledge. Learning from human experts requires automatic methods to transform textual data into conceptual structures (automated ontology learning). Semantic knowledge self-organization is a very important and desirable property of knowledge structures.

Understanding and prevision: (Sometimes this layer is referred to as the "wisdom" layer.) This layer deals with the question "why" or "what will happen"? It is about the ability to project into the future and reason back into the past. An aware system should be capable of extrapolating information into the future, and be able to estimate and evaluate the consequences of certain actions based on previous observations. In robotics, this can be predicting paths. In other fields it tends to mean reasoning, e.g. ontology-based reasoning. An active research field here is the autonomous creation (learning) of ontologies that can be used for reasoning (Zhou, 2007; Barforush & Rahnama 2012).

In all levels is uncertainty a key aspect. The uncertainty in the data should be propagated to the information level, where it is transformed into an uncertainty in the information, and then on to the knowledge level, etc..

An important point that perhaps is not obvious in our perspective on aware systems is the life-long learning although it is implicit in the "unknown at the time of design". We are not approaching problems where all data is available at once; we are studying systems where learning takes place over time, typically with streaming data. There are already excellent efforts towards automated data mining or model building, e.g. the recent "Feature Lab" by Kanter & Veeramachaneni (2015) or the KXEN system that is now part of SAP (Fogelman-Soulie & Marcade, 2008). These build on the idea that all data is available and the question is how the relationships in this data should be best modeled.

In all levels in the knowledge triangle is the human role and interaction with the systems an important research question. Human can play part in all steps of the knowledge creation, leading to a semi-unsupervised knowledge creation, e.g. providing clues on interesting data representations, clustering events, providing external data, giving feedback on suggested structures, etcetera. What is important is how machine and human create knowledge together, not like in the traditional AI or ML form where humans provide expertize that the machine is expected to replicate. We refer to this as joint human-machine learning.

2 Meeting societal challenges

We list some examples below that are particularly relevant for CAISR, using headings from the EU Horizon 2020 framework program, that tie to activities within CAISR.

2.1 Health, demographic change, wellbeing

Improving individuals' health patterns: The development in wearables has inspired a vision of using self-tracking for personalized (and improved) health. This means wearable devices that log our activities and interact with us, in order to help us improve our lifestyle (eat better food, exercise more, etc.) and, with time, decrease the load on the health care system. Numerous apps in this field are being introduced daily on the market, there are lots of user data being sent to servers all the time, and there are national projects aiming at storing long-term data for individuals. Khosla claims that "In fifteen years, data will transform diagnostics, to the point where automated systems may displace up to 80-percent of physicians' standard work" (Khosla, 2014). Even though the statement is about medical diagnosis and not proactive health patterns per se, it certainly applies also to promoting healthy behaviors. Combining life-logging data with health record data will be, to say the least, challenging (missing data, erroneous data, etc., humans are entering quite a lot of it) but the large quantity of data means that with time we should be able to make good analysis and provide good advice.

Healthy and active ageing: Wearables also apply to healthy and active ageing. However, what is equally important is in-home monitoring (ambient assisted living). Elderly living homes can be equipped with sensors and the sensor data (streams) analyzed and used for providing security solutions, emergency solutions, and assistive services. This offers a decreased cost for elderly living services, while maintaining a high quality of service. Aware systems research is about developing methods for autonomously analyzing such streams of data and construct knowledge about the individuals' living patterns. Without such knowledge it is difficult (impossible) to design a working (and simple) intelligent service for elderly living. In a critical text on ambient intelligence, two Philips researchers (Reddering & Scholten, 2003) express what they consider the most important challenges for ambient intelligence if it should ever become a useful technology: to construct knowledge systems that are simple, that can cope with the diversity and unpredictability of human needs, and that learn to ask and value the unknown.

2.2 Secure, Clean and Efficient Energy

Better models of energy use and demands: Energy production is to an increasing level produced by small-scale renewable energy sources (solar, wind, biofuel). The volatility of wind and solar (and in the future: wave) generation creates problems in balancing the demand with the generation of energy and operating conventional power plants in part load. This requires better prediction of energy demands. On a consumer scale this may be possible to achieve by combining ambient intelligence with smart power meters, but modelling this will require autonomous knowledge creation since the data and variation is so large. Ambient intelligent systems technology can also be used to learn the inhabitants' living patterns and provide feedback in order to improve (lower) the energy consumption. It is difficult to imagine how this can be done without autonomous knowledge creation.

2.3 Smart, green and integrated transport

Improving vehicle uptime: Transport is intimately linked to economic growth. Transport cost and efficiency are intimately linked to vehicle uptime, i.e. the reliability of the vehicle fleet. Transportation vehicles are becoming more and more advanced, with huge amounts of data streaming on the controller area network on-board the vehicles. More and more data are also logged in databases about maintenance operations and vehicle setups. Aware systems will be important for better maintenance solutions that build upon this data. Maintenance of vehicles is not optimal today; there are erroneous repairs done, there are on-road breakdowns that could be prevented, and there many commercial transport vehicles that don't pass the national annual inspections. All this can be improved with systems that allow logging on-board (streaming) data, fleet wide comparisons, and connecting on-board signatures with repair histories. Making sense from these data, learning over the life-time of the fleets, requires autonomous knowledge creation.

3 CAISR research questions

The CAISR focus is on research questions that are general across application areas, across research groups and relevant for external partners. We describe them below in relation to the levels in the knowledge pyramid. They are further refined, with specific contributions we have made and intend to make in CAISR, at the end of this section and in the following section.

Data: The research questions we explore here are how to select what data to collect and how to find general and robust representations of data. This can be by learning representations, by designing representations, or by searching through sets of representations and estimating how good (interesting) they are. This means research on how to autonomously engineer features, or ways to learn representations. Work on automatic feature engineering has been presented by Kanter & Veeramachaneni (2015), Cheng et al (2011), and Paulheim & Furnkranz (2012). Bengio et al. (2013) present a review on learning representations. It also means research on the general applicability of representations, at least for certain types of signals like images, see e.g. Bigun (2006) for a discussion. It means research on measures to determine how interesting a particular representation is, which is related to (but not equal to) measuring interestingness of rules, see e.g. Zhang et al. (2009) for a review of the latter. The plasticity-stability problem mentioned above is very important and something that is usually not handled. There are many practical issues here related to e.g. dealing with missing, flawed or erroneous data. This corresponds to using feedback from higher levels in the knowledge pyramid; what data is expected based on the type of event? Another important issue is how human expertize can be combined with machine work, i.e. how the machine data exploration can be done in interaction with humans. One more important issue is curiosity; an aware system must continuously search for interesting things.

Information: The research questions we explore here are how to do (semi-)autonomous deviation detection and autonomous clustering of events, as well as the maintenance of such categorizations, e.g. dealing with concept drift, seasonal variations, application changes, and so on. Clustering is still something of an art, see von Luxburg et al. (2012), and certainly a challenge to do well in an unsupervised manner and for general types of problems. Iverson (2008) has suggested, and patented, one general data driven solution intended for system maintenance. Angelov (2013) suggests fuzzy clustering as a method to design general purpose cluster structures. But there is still lots of room for improvement, or as von Luxburg et al (2012) put it: "Depending on the use to which a clustering is to be put, the same clustering can either be helpful or useless". It is also important to incorporate humans in the loop; can humans provide initial suggestions for categories, can humans give feedback on suggested categorizations, etc.?

Knowledge: The research questions we explore here are how to associate events from different data sources, including human generated data. An important part is also how this knowledge should be represented. With reallife data, the information provided will (inevitably) be connected with uncertainty, and a question is how to handle the combination of two information sources that both may be uncertain. In this context it is important to also consider how a human can be incorporated to build this knowledge, in a semi-supervised way.

Understanding and prevision: We aim to be able to predict the progress of observed events, and explain why certain things have happened. However, we will be using hand-made ontologies (at least initially), and not do research on the generation of ontologies.

Aware systems research is a systems science, i.e. there are many parts to the system and the results need to address several parts in the knowledge triangle and tie them together. To enable this, we aim to build demonstrators to showcase what this means, with sets of tools for all levels (at least for three levels). These tools will be parts in toolboxes for aware systems. One demonstrator will be the intelligent home environment (the aware system for ambient assisted living). Another one, funded mainly by projects outside of the Knowledge Foundation CAISR funds, will be the self-aware vehicle fleet for increased uptime. A very likely demonstrator is the mobile based decision support system for persons diagnosed with a chronic condition. A possible one is the aware fork-lift truck in a warehouse.

4 CAISR research projects

In this section we summarize the aware systems contributions made in CAISR during the first four years and the proposed contributions during the second half. Contributions have been made mostly on the data level and the information level, and to a very small part on the knowledge level. The data level research has been on constructing robust and general features (HMC2) and/or semi-supervised feature construction (AIMS and SA3L). The information level research has been on using unsupervised clustering to group observations (AIMS and SA3L). The knowledge level research has throughout been on using human supplied labels for the clusters, or observations. The planned contributions during the second half are on the data, the information, and the knowledge levels.

4.1 AIMS (Automatic inventory and mapping of goods)

A robot acquires semantics by linking its world model with human knowledge. A challenge is to find the appropriate level of abstraction before linking the human knowledge and the robot world model. The scientific results include a semi-supervised approach for semantic mapping, introducing human knowledge after unsupervised place categorization, in combination with adaptive cell decomposition of an occupancy grid map. The system autonomously builds a high level spatial model of the world by adopting generic features on the data (occupancy map) and instantiates it, without prior knowledge of the environment. Semantic inference is done on the derived instantiates and semantics are provided as labels accompanied with their functionality and inter relations between them. The proposed adaptive cell decomposition method interprets occupancy maps to bring out underlying spatial characteristics in the data (environment). The information is stored in two corresponding data structures in a format readily usable for humans and machines. Additional knowledge is created by subsequent semantic labeling by using human constructed templates. In addition, we have presented a canonical geometric-semantic model (adjustable according to different scenarios), along with a method for generating and matching these models into the latent structure of the map. The result is a geometric-semantic map there semantics (corridors, pallet rack cells) are encoded into the model through the choice of landmarks (pallet rack pillars).

4.2 HMC2 (Human Motion Classification and Characterization)

The methods and algorithms developed for prediction of physiological parameters of an athlete from EMG data cover have on the data and information level concerned segmentation/structuring of raw signals so that we get robustness of segmentation results in case of changing signal variability. The system assesses signal variability on the information level and makes necessary adjustments in the data level to obtain adequate results of the structuring. Accurate predictions of physiological parameters obtained from trainable models using the extracted features allow assessing state/condition of the athlete, providing short-term advice and gaining knowledge for long-term evaluations and future planning. The reference conditions are obtained by separate measurements.

The methods and algorithms developed for estimation of the fundamental parameters in human locomotion make uses of existing knowledge about human walk, known from research in physiology. This makes the features robust and possible to transfer from the laboratory settings into the real-world. Robust estimation of the fundamental gait events, assessed longitudinal over long time spans, allow generation of new knowledge on e.g. how variability in movement patterns is influenced with successful medication, enabling feedback systems for medication level control and evaluation of long term effects in treatments.

4.3 SA3L (Situation Aware Ambient Assisted Living)

The SA3L project is concerned with developing methods and tools for detection and interpretation of potentially dangerous situations in the home of elderly people. The situations are inherently difficult to specify and generalize due to the diversity of homes, behaviors and the numerous ways of deviating. Thus, a system for detecting deviations based on manually specified rules is difficult (impossible) to build and maintain. We approach this problem by learning the activity patterns in the home.

In the data layer, a new method for representing time-dependent patterns of binary sensor deployed in-homes was proposed and used for modelling human activity patterns. To address the plasticity-stability problem were no prior assumptions made regarding the relevant sensors or the spatio-temporal relations between sensors.

The project has contributed to the modelling of human in-home activity patterns with an unsupervised approach to compare, cluster and relate (in space and time) similar behaviors. Deviations from such normal models are distinguished by indirect if-then rules and thereby contributing to both the information and knowledge layers of the knowledge pyramid. The method has been shown to work both for simulated and real data (in a demonstrator environment).A focus has also been on building up a realistic smart home simulator, in cooperation with Ulster University.

4.4 MoveApp

The goal of the MoveApp project was to develop mobile and wearable systems to support self-management of chronic conditions characterized by motor symptoms. Research activities have focused on the data level of the information pyramid. In particular to find appropriate representations for accelerometer data; representations that facilitate power-efficient, on-line processing of the data onboard the smart watch for long periods of time. Data have been collected on subjects but the study is not analyzed yet.

4.5 Situation aware safety systems (SAS2)

The goal is to develop a system that goes beyond the requirements defined in the current safety standard for automated guided vehicles, and state-of-the-art, by introducing additional functionalities for safe detection of other objects and situations listed, identifying objects and estimating the trajectory of objects in the environment. Such a system would be more situation-aware since actions of moving agents could be foreseen and concerns can be made based on objects' identities (e.g. human or other truck/AGV).

The main research question is how to detect, estimate the trajectory of, and identify objects (and agents) in a warehouse environment, such that actions based on this information lead to safer and more efficient (in terms of productivity) AGV operation. The approach is to use a multi-layer map where each layer is more adapted to the specific purpose. We foresee at least three levels: semantic map (used for reasoning), geometric map (the layout of the environment used for e.g. planning) and spatial-temporal map (used for reaction and obstacle avoidance). A challenge is how interactions between different layers is done and how feedback from higher levels can, for example, be used to improve accuracy, consistency of different layers in the map; how to detect the difference between foreground (static and dynamic obstacles) or background model (static objects).

4.6 Long term multi-layer mapping

Mapping is a classical problem in robotics and is currently well understood how to solve this problem in static environments. However, in the long-term mapping problem it is not clear enough how to deal with changes in the environment and how to manage the scalability of the lifelong mapping process. Arrange maps in different layers depending on data, information and knowledge content (and/or application, e.g. localization, monitor events) helps in adding scalability. However, all these maps must be maintained, i.e. updated to account for changes in the environment. We can refer to this as lifelong mapping, to enable lifelong situation awareness.

The scientific contribution are: a long-term multiple-layer mapping system that accounts for changes in the environment and scalability issues to maintain global maps, acquired along a typical industrial vehicle operation, i.e. approximately 20 km of total path length; a multi-layer map architecture with a useful connotation in the industrial world; a mapping technique to model the dynamics in the environment; a map maintenance strategy to maintain compact but informative maps efficiently.

4.7 Intelligent environments supporting ageing at home

This project is a continuation of the SA3L project, and it aims at building models of human behavior patterns that can be generalized over different environments and individuals. Future capabilities of AAL involve reaction to an otherwise normal sensor reading if it happens concurrently with something else, preventing or alerting of unwanted events or abnormal patterns without the need of constant attention of an operator. Additionally, the user should be able to interact or to "speak" to the system, telling that s/he is aware of the situation (perhaps thanks to the alert). Early detection of diseases is another example of a promising application by monitoring activity patterns of elder people who wish to live in their own homes as long as possible, e.g. detecting a reduction in physical/sleeping activity or a fall. Research questions here include how to autonomously learn the habits of inhabitants, and thus be aware of their activities/status, and how to interact with the inhabitants.

In more detail in regard to the knowledge pyramid, data representations will be required, e.g., to keep track of individuals who interact with the robot. For the information layer, event recognition will occur in a simplified manner after event detection: e.g., an interacting person will be identified or marked as new if a face is detected, a touch such as a hug will be recognized or marked as new if a touch is detected. For the knowledge level, we will manually find features and parameters which could be useful from the literature, our own ideas, and watching interactions (e.g., a person might talk more if the interaction is proceeding well) as a starting point; parameters will be adapted by the robot during interactions based on evaluating success.

4.8 Knowledge discovery and data mining life logging data for fertility

The main contributions of this study are related to the data and information layer of the pyramid. One important challenge in this application stems from the fact that each individual is unique, presenting individual cycles and individual hormonal levels. Therefore, models based on an average representation of the fertility cycle are not accurate for the majority of individuals. The research focus is on developing models that capture domain knowledge, such as information about ovary fertility cycles, but automatically adapt to each individual, providing more accurate information. This requires designing robust or adaptable features.

4.9 Knowledge creation from data streams and service operations

We extend our collaboration with Toyota Material Handling Europe, from autonomous vehicles into the uptime and predictive maintenance of forklift trucks. The planned research contributions relate to all the stages in the knowledge pyramid. In the data level we will continue our research concerning automatically evaluating which signals that are most interesting to monitor, and how to find appropriate representations for various types of data. In the information level of the pyramid we need methods for grouping systems, automatically, into clusters based on various aspects such as configuration, usage or condition. The fleet of vehicles will be heterogeneous, and we will need to investigate new methods for describing operation (both the commonalities and differences). It is a challenge to find methods that can automatically, or with minimal support from human experts, decide on the most interesting configurations to focus on. This is additionally complicated by issues such as concept drift, seasonal variations, application changes, and so on, which all need to be automatically detected and taken into account. In the knowledge level, we will contribute with algorithms related to how machines and humans can jointly create knowledge from information.

The awareness requires analyzing different clues when looking for faults, and in this project we will develop an integrated framework for using several approaches: identification of "normal" or "expected" behavior and detecting deviations from that; finding out common patterns between issues that have proven problematic in the past and looking for new situations that are similar; characterizing events and time points when the condition changes for some reason. By building at all stages of the knowledge pyramid we will also develop online and incremental algorithms for decision support based on very little data for early detection of issues. Those algorithms will be looking at streams of data and processing information as soon as it arrives, giving initial warnings as soon as possible and then refining those decisions as more information becomes available.

4.10 Image analysis

There is a substantial research activity in CAISR directed at image analysis, mostly with a focus on the data and information levels. The most prominent research contribution regards the design of robust, general, translation, rotation and scale invariant features for images.

Knowing where humans are (presence) or who they are (identity) can be fundamental in aware intelligent systems not only for individual events (e.g. "person A is here"), but for the construction and understanding of bigger pictures in critical situations, for example "who is where", "is the right person in the right place (and not wounded)?", or "is a (maybe harmful) person where should not be?" Analysis of human activity over a certain period (something which involves continuous detection) is also necessary to provide information about what a person is doing, either over a short period (activity event) or a longer period (activity pattern).

Image analysis-based environmental monitoring, e.g. studies of long-term changes in aquatic ecosystems, assessment of water quality parameters, is another application area of aware intelligent systems addressed in our research. Semi-supervised training applied to build a model for machine-learned ranking allows assessing automatically the quality of intermediate results produced by the system itself and greatly reduces the risk of propagating intermediate errors into next processing steps.

4.11 Healthcare technology

Laryngology is a healthcare area with great potential for application of aware intelligent systems.Our research here touches on all four levels of the knowledge pyramid. A number of representations of varying granularity and complexity are created from raw voice data and then used for adaptive modelling, which enables having a data point specific model for each data point to be processed. Knowledge extracted from voice data is enriched with knowledge gathered in a set of association rules elicited from subjective self-evaluations/self-assessments using affinity analysis. The set of very simple rules together with self-organized 2D maps extracted for voice data-based models help clinicians gaining comprehensible insights (understanding) concerning specific cases as well as trends and important associations.

4.12 Self monitoring systems

Together with Volvo Technology we have been pursuing ideas on self-monitoring for several years now, focusing on the area of uptime and predictive maintenance. Due to the distributed nature of the vehicle fleet and communication limitations, we have developed methods for modelling the data and automatically finding the most interesting signals to focus on. A city bus is a very complex system and automatically building information about its operation, from the data, is beyond current state-of-the-art. We have published new algorithms for analyzing behavior of the fleet of vehicles and detecting anomalous individuals. Based on outliers detected this way, in combination with historical repair information, we have created knowledge about the state of health of a given vehicle, as well as whether and when it requires a workshop visit.

We have also looked at combining more sources of data. This brings its own issues related to, among other things, awareness of the reliability of each source and ways to resolve inconsistencies between them.

In addition, we have started the BIDAF project, in collaboration with SICS (project coordinator) and University of Skovde, about realizing the promise of advanced, near real-time analytics on uncertain data with high volume and velocity through machine learning techniques. Key challenges include development of a computational platform; machine learning algorithms suitable for handling massive data; analytics methodology for automatic creation of information, knowledge and understanding.

4.13 Robotics

The AIR project (Action and Intention Recognition in Human Interaction with Autonomous Systems) is a project in cooperation with Skovde University (project leader), Orebro University, and the Viktoria Institute. The CAISR contributions in the AIR project explore the use of novel forms of data (e.g. using breath sensors to allow a robot to detect a person's location without using possibly intrusive sensors capable of identifying individuals), as well as deriving rules to generate typical classes of behavior (such as playful motions) and predicting how a person will perceive such behavior over time to achieve good interactions. For the latter, knowledge creation will involve deriving an initial model from expert knowledge refined by data from real interactions (hand-coding by humans) and adapting the model autonomously to accommodate individual preferences by the robot.

References

[1] Hintikka, J., "Impossible Possible Worlds Vindicated", Journal of Philosophical Logic, 4, pp. 475-484 (1975).

[2] Fagin, R. and Halpern, J.Y., “Belief, Awareness, and Limited Reasoning”, Artificial Intelligence, 34, pp. 39-76 (1988).

[3] Modica, S. and Rustichini, A., “Awareness and Partitional Information Structures”, Theory and Decision, 37, pp. 107-124 (1994).

[4] Devanur, N., and Fortnow, L, “A computational theory of awareness and decision making”. In Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge, pp 99-107 (2009)

[5] Endsley, M.R., “Toward a Theory of Situation Awareness in Dynamic Systems”, Human Factors, 37, pp. 32-64 (1995).

[6] Gutwin, C., and Greenberg, S., “A Descriptive Framework of Workspace Awareness for Real-Time Groupware”, Computer Supported Cooperative Work, 11, pp. 411-446 (2002).

[7] Zhao, Q, Hsieh, C-H, Naruse, K, and She, Z., “Awareness Science and Engineering” (editorial), Applied Computational Intelligence and Soft Computing, 2012 (2012).

[8] Zhao, Q., “Computational Awareness: Another Way towards Intelligence”, Computational Intelligence (Eds. Madani et al.), Book Series Studies in Computational Intelligence, 465, pp. 3-14 (2013).

[9] Ackoff, R.L., “From data to wisdom”, Journal of Applied Systems Analysis, 16, pp. 3-9 (1989).

[10] Degani, A., Jorgensen, C, Iverson, D., Shafto, M. and Olson, L., “On Organization of Information: Approach and Early Work”, Technical Report NASA/TM2009215368, NASA AMES (2009)

[11] Rowley, J., “The wisdom hierarchy: representations of the DIKW hierarchy”, 33, pp. 163-180 (2007).

[12] Curtis, G., and Cobham, D., Business Information Systems: Analysis, Design and Practice, FT Prentice Hall, Harlow (2005).

[13] Horeis, T., and Sick, B., “Collaborative Knowledge Discovery & Data Mining: From Knowledge to Experience”, IEEE Symposium on Computational Intelligence and Data Mining 2007 (CIDM 2007), pp. 421-428 (2007)

[14] Zhou, L., “Ontology learning: state of the art and open issues”, Information Technology Management, 8, pp. 241-252 (2007).

[15] Barforush, A.A., and Rahnama, A., “Ontology learning: revisited”, Journal of Web Engineering, 11, pp. 269-289 (2012).

[16] Kanter, M. J., and Veeramachaneni, K., “Deep Feature Synthesis: Towards Automating Data Science Endeavors”, International Conference on Data Science and Advanced Analytics (DSAA), pp, 1-10 (2015)

[17] Fogelman-Souli, F., and Marcad, E., “Industrial Mining of Massive Data Sets”, Mining Massive Data Sets for Security, (NATO ASI Workshop), IOS Press, pp. 44-61 (2008).

[18] Khosla, V., “20-percent doctor included: Speculations and musings of a technology optimist” (2014)

[19] Reddering, K., and Scholten, L., “Understanding Human Cultures”, Chapter 1.2 in The New Everyday: Views on Ambient Intelligence (Eds. Aarts, E., and Marzano, S.), 010 Publishers, Rotterdam (2003)

[20] Cheng, W., Kasneci, G., Graepel, T., Stern, D., and Herbrich, R., “Automated feature generation from structured knowledge”, Proceedings of the 20th ACM Conference on Information and Knowledge Management, ACM, pp. 1395-1404 (2011).

[21] Paulheim, H., and Frnkranz, J., “Unsupervised generation of data mining features from linked open data”, Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS12), ACM, pp. 31:131:12 (2012).

[22] Bengio, Y., Courville, A., and Vincent, P., “Representation Learning: A Review and New Perspectives”, IEEE Transactions on Pattern Analysis and Machine Intelligence (special issue Learning Deep Architectures), 35, pp. 1798-1828 (2013).

[23] Bigun, J., Vision with direction, Springer Berlin (2006).

[24] Zhang, Y., Zhang, L., Nie, G. and Shi, Y., “A Survey of Interestingness Measures for Association Rules”, 2009 International Conference on Business Intelligence and Financial Engineering, IEEE (2009).

[25] von Luxburg, U., Williamson, R. C. and Guyon, I., “Clustering: Science or Art?”, Journal of Machine Learning Research (JMLR): Workshop and Conference Proceedings, 27, pp. 65-79 (2012)

[26] Iverson, D. L., “Inductive monitoring system constructed from nominal system data and its use in realtime system monitoring”, Patent US 7,383,238 B1 (2008).

[27] Angelov, P., Autonomous Learning Systems, John Wiley & Sons (2013).