COMMUNICATIONS

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

VOLUME: 59 **Number: 2** **YEAR: 2017**

Faculy of Sciences, Ankara University
06100 Beşevler, Ankara-Turkey
ISSN 1303-6009

COMMUNICATIONS

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 59 **Number : 2** **Year :2017**

This Journal is published two issues in a year by the Faculty of Sciences, University of Ankara. Articles and any other material published in this journal represent the opinions of the author(s) and should not be construed to reflect the opinions of the Editor(s) and the Publisher(s).

Correspondence Address:
COMMUNICATIONS
EDITORIAL OFFICE
Ankara University, Faculty of Sciences,
06100 Tandoğan, ANKARA – TURKEY
Tel: (90) 312-212 67 20 Fax: (90) 312-223 23 95
e-mail: commun@science.ankara.edu.tr
http://communications.science.ankara.edu.tr/index.php?series=A2A3

Print:
Ankara University Press
İncitaş Sokak No:10 06510 Beşevler
ANKARA – TURKEY
Tel: (90) 312-213 65 65

COMMUNICATIONS

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 59 **Number : 2** **Year :2017**

©Ankara University Press, Ankara 2017

COMMUNICATIONS

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Available online: August 01, 2017

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 59, Number 2, Pages 1–9 (2017)
DOI: 10.1501/commua1-2_0000000099
ISSN 1303-6009
http://communications.science.ankara.edu.tr/index.php?series=A2

ENTROPY SQUEEZING OF A MULTI-PHOTON JAYNES-CUMMINGS ATOM IN THE PRESENCE OF NOISE

HÜNKAR KAYHAN

Abstract. In this work, we study the entropy squeezing of a two-level atom interacting with a single-mode quantum field by a multi-photon Jaynes-Cummings Model in the presence of the two-state random phase telegraph noise. We show that the entropy squeezing is very sensitive to the noise. It disappears in time quickly due to the strongly destructive effect of the noise.

The Jaynes-Cummings Model (JCM) [1, 2, 3] is the basic model for describing the interaction of a two-level atom with a single-mode cavity quantum field under the rotating-wave approximation. This model reveals crucial non-classical properties such as sub-Poissonian statistics, anti-bunching, squeezing and collapse and revival phenomena [4, 5]. Of the several interests to the model, one has been devoted to the squeezing properties of the atom [6, 7, 8, 9, 10]. In these works, the atomic squeezing properties were studied on the base of the Heisenberg uncertainty relation (HUR). But, HUR cannot provide sufficient information about the atomic squeezing in particular when the atomic inversion vanishes. As an alternative to the HUR, Hirschman [11] studied quantum uncertainty by using quantum entropy theory. And the limitations of the HUR have been overcome by using the entropic uncertainty relation (EUR) [12, 13]. Fang *et.al.* [14] found that EUR can be used as a general criterion for the squeezing of an atom. Accordingly, they proposed a measure of the squeezing of an atom the so-called squeezed in entropy in order to obtain sufficient information on atomic squeezing. The entropy squeezing of the atom has been studied extensively [15, 16, 17, 18, 19]. These works reveal that the entropy squeezing based on the EUR is more precise than the variance squeezing based on the HUR, as a measure of the atomic squeezing.

For the realistic situations, the JCM-type atom-field interactions should be considered with a decoherence mechanism. One consideration was formulated by Joshi *et.al.* [20, 21] in which the authors re-describe the JCM with the random telegraph noise. For the realization of this noise, the authors give some situations

Received by the editors: May 29, 2017; Accepted: July 19, 2017.
PACS numbers. 03.67.-a, 42.50.Lc, 42.50.Ct.
Key words and phrases. Entropy squeezing, noise, Jaynes-Cummings Model.

such as the source of the field or the instability in the atomic vapor production. This noise influences the dipole or the transverse relaxation of the interaction. The resulting decoherence mechanism conserves the energy of the system, but destructs the quantum coherence.

In this work, we study the entropy squeezing of a two-level atom interacting with a single-mode quantum field by a multi-photon JCM in the presence of the two-state random phase telegraph noise. We show that the entropy squeezing is very sensitive to the noise. It disappears in time quickly due to the strongly destructive effect of the noise.

The Hamiltonian of a multi-photon JCM with resonance between the atomic transition and the field frequency [22, 23] is given by $(\hbar = 1)$

$$H = \omega \frac{S_z}{2} + \omega a^\dagger a + g(S_+ a^k + S_- a^{\dagger k}) \tag{1}$$

where $S_\pm, S_z$ are the spin-1/2 operators, $a, a^\dagger$ denote the annihilation and the creation operators of the field, ω is the atomic transition frequency and the field frequency. g is the coupling coefficient which gives the interaction strength between the atom and the field and k represents the k-photon process. The experimental realization of the multi-photon process can seen in a trapped ion [24].

In the case of the interaction with the random phase telegraph noise, the coupling coefficient is modified as [20]

$$g(t) = g_0 e^{-i\phi(t)} \tag{2}$$

where g_0 is the non-noisy coupling coefficient and $\phi(t)$ represents the random telegraph which fluctuates between two states of the noise denoted by (a) and $(-a)$. These random fluctuations obey the Poisson jump process. The fluctuations of $\phi(t)$ are also Markovian which allows one to take the average over the stochastic fluctuations. The average time between these jumps is called the mean dwell time.

The multi-photon JCM in the presence of the random phase telegraph noise becomes

$$H = \omega \frac{S_z}{2} + \omega a^\dagger a + g_0(e^{-i\phi(t)} S_+ a^k + e^{i\phi(t)} S_- a^{\dagger k}) \tag{3}$$

For the initial state of the system, we assume for simplicity that the atom is in the excited state $|e\rangle$ and the field is in the Fock state $|n\rangle$. In this case, the initial state of the system is

$$\rho(0) = |n, e\rangle\langle n, e| \tag{4}$$

In order to find an exact solution to the system under the noise, we use the Burshtein equation [25, 26, 27] by the solution method in Ref. [28] in which we studied the entanglement of atom-field interaction by the JCM with two-state random phase telegraph noise. We also considered some other applications of the Burshtein equation elsewhere for investigating entanglement dynamics in different atom-field systems with this noise [29] . The Burshtein equation is defined as

$$\frac{\partial}{\partial t}V_\alpha(t) = -iM(\alpha)V_\alpha(t) - \frac{1}{T}\sum_\beta[\delta_{\alpha\beta} - \mathrm{f}(\alpha|\beta)]V_\beta(t) \tag{5}$$

where α and β represent the phase of the noise with the values (a) and $(-a)$, the function $\mathrm{f}(\alpha|\beta)$ is the probability of $\phi(t)$ to change its state such that $\mathrm{f}(a,-a) = \mathrm{f}(-a,a) = 1$ and $\mathrm{f}(a,a) = \mathrm{f}(-a,-a) = 0$. The time-dependent element $V_\alpha(t)$ is the α-fixed state component of the vector $\hat{V}(t)$ which is the transpose of the matrix $[\rho_n^{11}(t), \rho_n^{22}(t), \rho_n^{12}(t), \rho_n^{21}(t)]$. $M(\alpha)$ is called the effective Liouville operator with the fixed α-state of the noise obtained from the equation $\hat{\dot{V}}^k = -iM^{kl}\hat{V}^l$. T is the mean dwell time which determines the strength of the dephasing induced by the noise. The smaller T, the stronger noise. In the basis $|n,e\rangle$ and $|n+k,g\rangle$, the following expressions for the stochastic evolution of the elements of the density matrix of the system can be obtained from von Neumann-Lioville equation

$$\begin{aligned}
\frac{d\rho_n^{11}(t)}{dt} &= ig_0\sqrt{\frac{(n+k)!}{n!}}[e^{-i\phi}\rho_n^{12}(t) - e^{i\phi}\rho_n^{21}(t)] \\
\frac{d\rho_n^{22}(t)}{dt} &= ig_0\sqrt{\frac{(n+k)!}{n!}}[e^{i\phi}\rho_n^{21}(t) - e^{-i\phi}\rho_n^{12}(t)] \\
\frac{d\rho_n^{12}(t)}{dt} &= ig_0\sqrt{\frac{(n+k)!}{n!}}e^{i\phi}[\rho_n^{11}(t) - \rho_n^{22}(t)] \\
\frac{d\rho_n^{21}(t)}{dt} &= ig_0\sqrt{\frac{(n+k)!}{n!}}e^{-i\phi}[\rho_n^{22}(t) - \rho_n^{11}(t)]
\end{aligned} \tag{6}$$

where the diagonal elements are

$$\begin{aligned}
\rho_n^{11}(t) &= \langle n,e|\rho(t)|n,e\rangle \\
\rho_n^{22}(t) &= \langle n+k,g|\rho(t)|n+k,g\rangle
\end{aligned} \tag{7}$$

and the off-diagonal elements are

$$\begin{aligned}
\rho_n^{12}(t) &= \langle n,e|\rho(t)|n+k,g\rangle \\
\rho_n^{21}(t) &= \langle n+k,g|\rho(t)|n,e\rangle
\end{aligned} \tag{8}$$

By constructing the elements of the Burshtein equation from these expressions and by using the Laplace transformation techniques [28], one can obtain the following noise-averaged solution

$$\langle\rho_n^{11}(t)\rangle = \frac{1}{2}[1+\sum_{j=1}^{3}\frac{\lambda_j(\lambda_j+\frac{2}{T})}{\prod_{k\neq j}(\lambda_j-\lambda_k)}\exp(\lambda_j t)] \tag{9}$$

$$\langle\rho_n^{22}(t)\rangle = \frac{1}{2}[1-\sum_{j=1}^{3}\frac{\lambda_j(\lambda_j+\frac{2}{T})}{\prod_{k\neq j}(\lambda_j-\lambda_k)}\exp(\lambda_j t)] \tag{10}$$

$$\langle\rho_n^{12}(t)\rangle = ig_0\cos a\sqrt{\frac{(n+k)!}{n!}}\sum_{j=1}^{3}\frac{(\lambda_j+\frac{2}{T})}{\prod_{k\neq j}(\lambda_j-\lambda_k)}\exp(\lambda_j t) \tag{11}$$

and

$$\langle\rho_n^{21}(t)\rangle = \langle\rho_n^{12}(t)\rangle^* \tag{12}$$

λ_js are the roots of the equation

$$\lambda_j^3+\frac{2\lambda_j^2}{T}+4g_0^2\frac{(n+k)!}{n!}\lambda_j+\frac{8g_0^2(n+k)!}{Tn!}\cos^2 a=0 \tag{13}$$

The noise-averaged density matrix of the system $\langle\rho(t)\rangle$ takes the form of

$$\begin{aligned}\langle\rho(t)\rangle = {} & \langle\rho_n^{11}(t)\rangle|n,e\rangle\langle n,e| + \langle\rho_n^{12}(t)\rangle|n,e\rangle\langle n+k,g| \\ & +\langle\rho_n^{21}(t)\rangle|n+k,g\rangle\langle n,e| + \langle\rho_n^{22}(t)\rangle|n+k,g\rangle\langle n+k,g|\end{aligned} \tag{14}$$

The HUR for an atomic system is defined as

$$\Delta S_x\Delta S_y \geq \frac{1}{2}|\langle S_z\rangle| \tag{15}$$

The fluctuations in the components of the Pauli operators are squeezed if

$$V(S_k)=\Delta S_k-\sqrt{\frac{|\langle S_z\rangle|}{2}}<0, \qquad k=x \;\; or \;\; y \tag{16}$$

where $\Delta S_k=\sqrt{\langle S_k^2\rangle-\langle S_k\rangle^2}$. But, this definition of the variance squeezing can not give information when $\langle S_z\rangle=0$. Fang *et.al.*'s definition for the squeezing the so-called the entropy squeezing is

$$E(S_k)=\exp(H(S_k))-2/\sqrt{\exp(H(S_z))}, \qquad k=x \;\; or \;\; y \tag{17}$$

where $H(S_k)$ denotes the information entropy of the component S_k

$$H(S_k)=-\sum_{i=1}^{D}P_i(S_k)\ln(P_i(S_k)), \qquad k=x,y,z \tag{18}$$

where $P_i(S_k)$ represents the probability distribution of D possible measurement outcomes of the S_k component. It is given by $P_i(S_k)=\langle\psi_{ki}|\rho|\psi_{ki}\rangle$ for a quantum system ρ where $|\psi_{ki}\rangle$ is an eigenvector of the component S_k. So, they are the

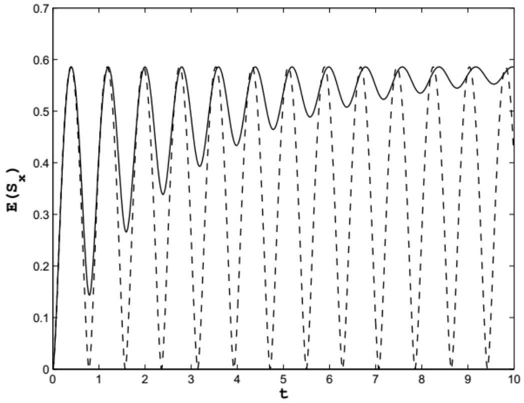


FIGURE 1. Entropy squeezing factor $E(S_x)$ as a function of time t. $n = 3$ and $k = 1$. The non-noisy case $a = 0$ and $T \to \infty$ for dash line and a noisy case $a = 0.4$ and $T = 1$ for solid line.

elements of projective measurements. In this definition, there exists a squeezing in the fluctuations of S_k, if $E(S_k) < 0$.

The probabilities are given as

$$\begin{array}{rcl} P_1(S_x) & = & 1/2(1 + 2Re\langle\rho_n^{12}(t)\rangle) \\ P_2(S_x) & = & 1/2(1 - 2Re\langle\rho_n^{12}(t)\rangle) \\ P_1(S_y) & = & 1/2(1 - 2Im\langle\rho_n^{12}(t)\rangle) \\ P_2(S_y) & = & 1/2(1 + 2Im\langle\rho_n^{12}(t)\rangle) \\ P_1(S_z) & = & \langle\rho_n^{22}(t)\rangle \\ P_2(S_z) & = & \langle\rho_n^{11}(t)\rangle \end{array} \qquad (19)$$

Since the entropy squeezing factor $E(S_k)$ is more reliable in providing information for the squeezing of the atom than the variance squeezing $V(S_k)$, we will only deal with the analysis of the entropy squeezing factor.

We investigate the influence of the noise on the entropy squeezing factor of the atom by the following figures. (In these, we assume that the non-noisy coupling coefficient is unity $g_0 = 1$.) Figures (1)-(2) show that $E(S_x)$ oscillates periodically and has no negative values during the time-evolution of the system. This situation remains unchanged when taking into account the noise. So, there is no entropy squeezing in the S_x component at any time in the absence or in the presence of the noise. For the S_y component as shown in figures (3)-(4), there is an entropy squeezing. $E(S_y)$ oscillates periodically and achieves some negative values during the time-evolution of the system. But, the situation changes when the noise is

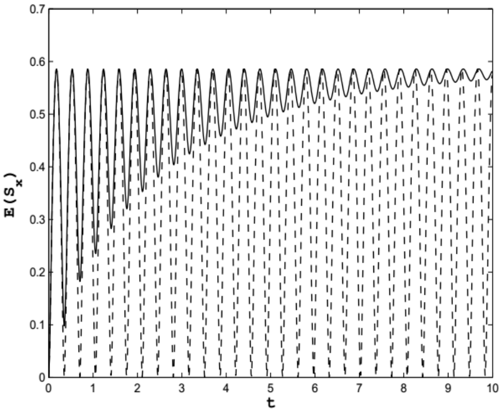


FIGURE 2. Entropy squeezing factor $E(S_x)$ as a function of time t. $n = 3$ and $k = 2$. The non-noisy case $a = 0$ and $T \to \infty$ for dash line and a noisy case $a = 0.4$ and $T = 1$ for solid line.

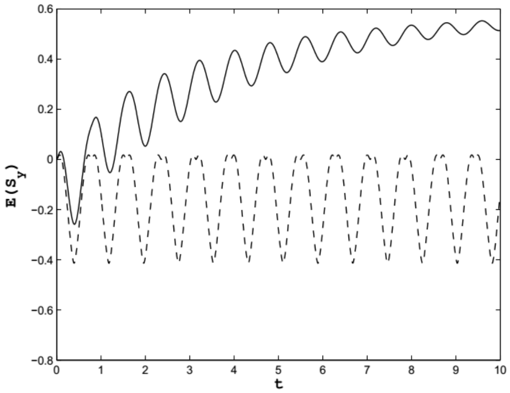


FIGURE 3. Entropy squeezing factor $E(S_x)$ as a function of time t. $n = 3$ and $k = 1$. The non-noisy case $a = 0$ and $T \to \infty$ for dash line and a noisy case $a = 0.4$ and $T = 1$ for solid line.

involved. The negative values of $E(S_y)$ disappear, as time passes. So, the noise obviously destructs gradually the existing squeezing in the S_y component during the time-evolution of the system. In the both components S_x and S_y, as the value of k increases, the decay of the entropy squeezing in these components occurs with a smaller period. Thus, the entropy squeezing is very sensitive to the noise. It disappears in time quickly due to the strongly destructive effect of the noise.

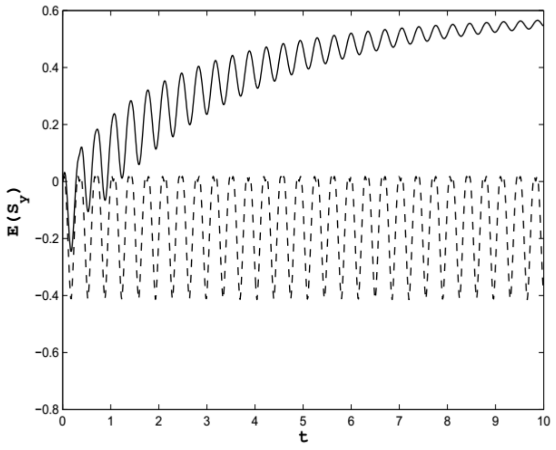


FIGURE 4. Entropy squeezing factor $E(S_y)$ as a function of time t. $n = 3$ and $k = 2$. The non-noisy case $a = 0$ and $T \to \infty$ for dash line and a noisy case $a = 0.4$ and $T = 1$ for solid line.

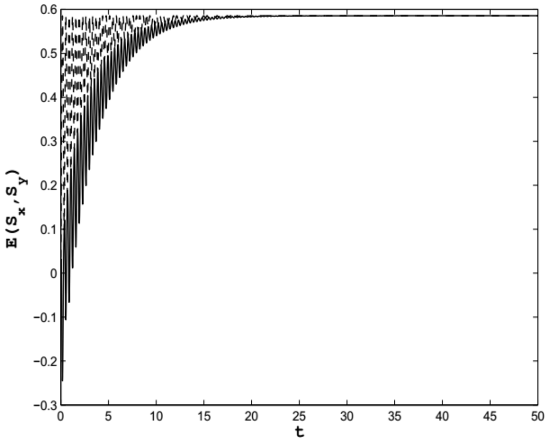


FIGURE 5. Entropy squeezing factor $E(S_x)$ for dash line and $E(S_y)$ for solid line as a function of time t. $n = 3$, $k = 2$, $a = 0.4$ and $T = 1$.

In Figure (5), we look at a longer-time behavior of the entropy squeezing factor for observing more clearly the decoherence effect of the noise. We see that both

$E(S_x)$ and $E(S_y)$ decay gradually and eventually reach the same stable value in time due to the destructive effect of the noise with $E(S_y) \leq E(S_x)$.

In summary, we have studied the entropy squeezing of a two-level atom interacting with a single-mode quantum field by a multi-photon Jaynes-Cummings Model in the presence of the two-state random phase telegraph noise. We have shown that the entropy squeezing is very sensitive to the noise. It disappears in time quickly due to the strongly destructive effect of the noise.

Acknowledgements

I am grateful to the referees for the comments and recommendations that improve this paper.

References

[1] E. T. Jaynes and F. W. Cummings, Proc. IEEE **51**, 89 (1963).

[2] H.-I. Yoo and J. H. Eberly, Phys. Rep. **118**, 239 (1985).

[3] B. W. Shore and P. L. Knight, J. Mod. Opt. **40**, 1195 (1993).

[4] M. O. Scully and M. S. Zubairy, Quantum Optics (UK:Cambridge University Press), (1997).

[5] G. Rempe, H. Walther and N. Klein, Phys. Rev. Lett. **58**, 353 (1987).

[6] X. S. Li, D. L. Lin, T. F. George and Z. D. Liu, Phys. Rev. A **40**, 228 (1989).

[7] X. S. Li, D. L. Lin and T. F. George, Phys. Rev. A **40**, 2504 (1989).

[8] P. Zhou and J. S. Peng, Phys. Rev. A **44**, 3331 (1991).

[9] M. M. Ashraf and M. S. K. Razmi, Phys. Rev. A **45**, 8121 (1992).

[10] K. Wódkiewicz, P. L. Knight, S. J. Buckle and S. M. Barnett, Phys. Rev. A **35**, 2567 (1987).

[11] I. I. Hirschman, Am. J. Math **79**, 152 (1957).

[12] J. Sanchez-Ruiz, Phys. Lett. A **201**, 125 (1995).

[13] J. Sanchez-Ruiz, Phys. Lett. A **244**, 189 (1998).

[14] M.-F. Fang, P. Zhou and S. Swain, J. Mod. Opt. **47**, 1043 (2000).

[15] M. Abdel-Aty, M. S. Abdalla and A.-S. F. Obada, J. Phys. A: Math. Gen. **34**, 9129 (2001).

[16] T. M. El-Shahat, S. Abdel-Khalek, M. Abdel-Aty and A.-S. F. Obada, Chaos, Solitons & Fractals **18**, 289 (2003).

[17] T. M. El-Shahat, S. Abdel-Khalek and A.-S. F. Obada, Chaos, Solitons & Fractals **26**, 1293 (2005).

[18] X.-Q. Yan, B. Shao and J. Zou, Chaos, Solitons & Fractals **40**, 215 (2009).

[19] H. Kayhan, Phys. Scr. **84**, 045401 (2011).

[20] S. V. Lawande, A. Joshi and Q. V. Lawande, Phys. Rev. A **52**, 619 (1995).

[21] A. Joshi, J. Mod. Optics **42**, 2561 (1995).

[22] B. Buck and C. V. Sukumar, Phys. Lett. A **81**, 132 (1981).

[23] C. C. Gerry, Phys. Rev. A **37**, 2683 (1988).

[24] V. Bŭzek *et.al.*, Phys. Rev. A **56**, 2352 (1997).

[25] A. I. Burshtein, Zh. Eksp. Teor. Fiz. **49**, 1362 (1965) [Sov. Phys.-JETP **22**, 939 (1966)].

[26] A. I. Burshtein and Y. S. Oseledchilk, Zh. Eksp. Teor. Fiz. **51**, 1071 (1966) [Sov. Phys.-JETP **24**, 716 (1967)].

[27] L. D. Zusman and A. I. Burshtein, Zh. Eksp. Teor. Fiz. **61**, 976 (1971) [Sov. Phys.-JETP **34**, 520 (1972)].

[28] H. Kayhan, Eur. Phys. J. D **48**, 443 (2008).

[29] H. Kayhan, Int. J. Quantum Inform. **9**, 1229 (2011).

E-mail address: hunkar_k@ibu.edu.tr
Current address: Department of Physics, Abant Izzet Baysal University, Bolu-14280, Turkey.
ORCID: http://orcid.org/0000-0001-6340-8933

Available online: December 20, 2017

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 59, Number 2, Pages 11-19 (2017)
DOI: 10.1501/commua1-2_0000000100
ISSN 1303-6009
http://communications.science.ankara.edu.tr/index.php?series=A2-A3

COMMUNICATIONS
SERIES A2-A3

INTERPOLATION METHODS FOR RECOVERING THE SAMPLING VALUES OF GPR DATA

MERVE ÖZKAN OKAY and REFIK SAMET

ABSTRACT. Ground Penetrating Radar (GPR) is widely used to acquire the data from near surface depth. The acquired GPR data allow the users to investigate and examine the underground structures (anomalies) easily, quickly and accurately without any excavation. In GPR studies, data collection parameters such as the profile interval and step size, which can be controlled by users, play an important role in the identification of underground structures. But search area properties such as uneven surface, the presence of archaeological and other obstacles cannot be controlled by users. The obtained accuracy depends on the completeness and resolution of acquired GPR data. Due to some research area properties the data acquired from the search area may become incomplete and inadequate. Before analyzing, visualization and interpretation of the underground structures, the incomplete GPR data should be recovered. In this paper, nonstandard interpolation method are proposed for completing the missing data. The proposed methods were implemented on the real GPR data acquired from the test area. The obtained results showed that the similarity of the produced data as quite closer to the original data.

1. INTRODUCTION

Ground Penetrating Radar (GPR) is a widely used method to investigate the underground archaeological and geological structures [3], [17]. The use of GPR in researches and applications has recently been increasing, because it can explore and detect the underground structures quickly and accurately.

There are two main factors that affect the success of GPR research and applications. These are data collection parameters and search area properties. Data collection parameters such as antenna frequency, sampling and profile range, etc. are under the control of users, and the values of these parameters can be selected according to the search area properties. On the other hand, the search area properties such as uneven surface, the archaeological and other obstacles, technical failures during data collecting, etc. are outside the control of users. Any physical and chemical changes under the ground such as, metals, dissolved salts and the presence of conductive materials, etc. affect the properties of electromagnetic waves such as speed, amplitude

Received by the editors: October 27, 2017; Accepted: December 15, 2017
Key word and phrases: Ground Penetrating Radar (GPR), Sampling Value, GPR Data, Underground Structures, Incomplete Data, Interpolation Methods

and wavelength [1], [4-5], [7]. Due to search area properties, the obtained data can be missing/inadequate, and therefore an accuracy of 2D/3D visualization of the underground structures decreases [6]. This study proposes nonstandard Mean interpolation method to produce incomplete data as close to original data. In addition to proposed interpolation method, standard Cubic, Cubic Spline, Linear, Median and Mean interpolation methods were tested on real GPR data. The obtained results proved that the proposed nonstandard Mean interpolation methods give the best results in comparison with the standard ones to produce the incomplete data.

The study is organized as follows. Firstly, the proposed methodology is described to produce the incomplete data. Secondly, proposed nonstandard and known standard interpolation methods are implemented on real GPR data and obtained results are compared. Finally, obtained results from comparing are evaluated and the contributions of the study are summarized.

2. Proposed nonstandard mean interpolation methods

GPR data consist of N parallel profiles. Each profile consists of M traces. Each trace consists of K sample values (Fig.2.1).

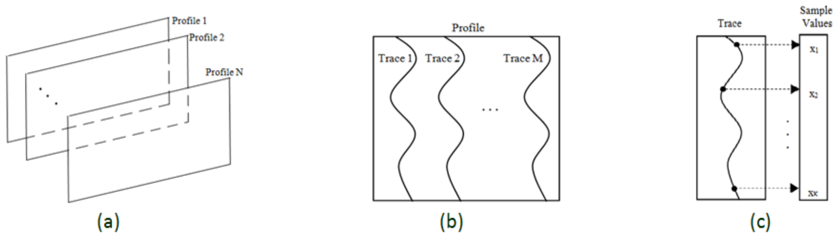


Figure 2.1 (a) Profiles of GPR data; (b) Traces of profile; (c) Sample values of trace

Due to the ruggedness and uneven surface of the search area, technical failures, etc. some sampling values may not be measured and collected. In order to recover these incomplete data with interpolation methods [12-13], the following steps are applied:

1. Obtaining data from the search area;
2. Producing new sampling value by using original sampling values;
3. Comparing the new sampling value with original sampling value;
4. Determining the appropriate interpolation method.

2.1 Obtaining the Data from the Search Area

It is necessary to use data sets obtained from the search area where the underground structure is well known for testing and verification of proposed technique. This type of data set is only possible with a created test area. The various sizes materials (such as, pipes and stone walls, metal and plastic drums) were placed in this test area and obtained data from there. In this way, the accuracy of the information about object can be determined precisely.

2.2 Producing the New Sampling Values

Trace consists of K sampling values. Some of these sampling values are randomly extracted. By applying different interpolation methods with the remaining sampling values, new values are produced instead of the extracted values. Interpolation methods used for producing sampling value are standard interpolation methods (Cubic, Spline, Linear, Median) and proposed interpolation method [9-11], [14-16].

The production of new sampling values by using the original sampling values is applied for m^{th} trace *(m=1,2,fi,M)* of each n^{th} profile *(n=1,2,fi,N)* (Fig.2.2).

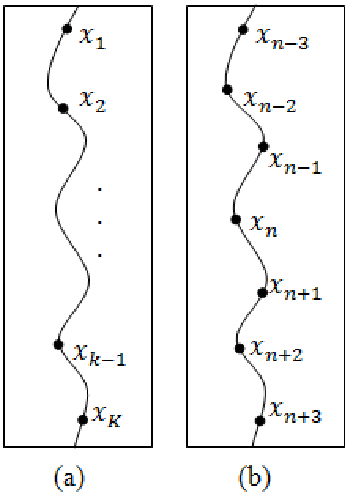


FIGURE 2.2 (a) The sampling values of m^{th}trace; (b) The sampling values used in interpolation method

2.2.1 Proposed Mean Interpolation Method

Any sampling value x_k $(k = 1,2,\dots,K)$ of the m^{th} *(m=1,2,fi,M)* trace is produced by using sampling values x_{k-i} and x_{k+i} $(i = 1,2,\dots,\mathrm{I} \Leftarrow (k-i) \geq$

1 and $(k + i) \leq K)$ (Fig. 2.2). According to standard mean interpolation method, the average of used sampling values is calculated to produce new sampling value (Eq. #2.1). On the other hand, according to the proposed nonstandard Mean interpolation method, the variable E of the increase or decrease amount (Eq. #2.2) and distance (Eq. #2.3) between the used values are taken into account during producing incomplete trace values.

$$M_{std} = \sum_{i=1}^{n/2} \frac{x_{k-i} + \cdots + x_{k+i}}{n} \tag{2.1}$$

$$E = \frac{x_{k+1} - x_{k-1}}{x_{k+1}} \tag{2.2}$$

When the incomplete sampling values are calculated using more than 2 neighbor sampling values, weight w of neighbors should be taken into account.

$$w = \frac{2\left({}^{n_v}/_{2} - i + 1\right)}{{}^{n_v}/_{2}\left({}^{n_v}/_{2} + 1\right)} \tag{2.3}$$

where n_v is a number of used neighbor sampling values. According to Eq. #2.3 the weight of the nearest neighbor sampling values is higher than others. So, using Eq. #2.2 and Eq. #2.3, x_k can be calculated as follows.

$$x_k = \sum_{i=1}^{\frac{n_v}{2}} w(x_{k-i}\,(0.5 \pm E) + x_{k+i}\,(0.5 \mp E)) \tag{2.4}$$

According to Eq. #2.4, 0.5 is the coefficient used to calculate the standard mean, that is, it is taken in the same percentage from sampling values. In addition, in the developed interpolation technique, if there is an increase between the values, the increase amount is added to 0.5 for using high neighbor value and the increase amount is subtracted from 0.5 for using low neighbor value.

2.3. Comparison and Determining Similarity Ratio

In the comparison step, the produced sampling values by using interpolation operations are compared with the original sampling values. The similarity ratio is taken into consideration while determining the ideal interpolation technique.

Comparison was done by the Pearson Correlation Coefficient (PCC) [2], [8] metric. The PCC is calculated using the following formula.

$$P_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{2.5}$$

where, X and Y are sampling values; $cov(X,Y)$ is the covariance of the two sampling values; σ_X and σ_Y are the standard deviations of corresponding sampling values. The value of PCC varies between -1 and 1. A value that is close to -1 or 1 means that the similarity between the two sampling values is strong.

2.4. Determining the Optimal Interpolation Technique

While determining the optimal interpolation technique, the results of interpolation techniques and original data are compared. According to comparing results, the technique giving the closest similarity ratio as close to the original is determined as an optimal interpolation technique. In addition to similarity ratios, the number of used neighbor values is taken into account. As a result, the optimal technique is determined by looking at similarity ratios and the number of used neighbor value.

3. Implementation

For the implementation of the proposed nonstandard Mean and standard Cubic, Cubic Spline, Linear, Median and Mean interpolation methods [9-11], [14-16], the profiles (Fig.3.1 (b)) acquired from embedded wall structure (Fig.3.1 (a)) in the test area were used.

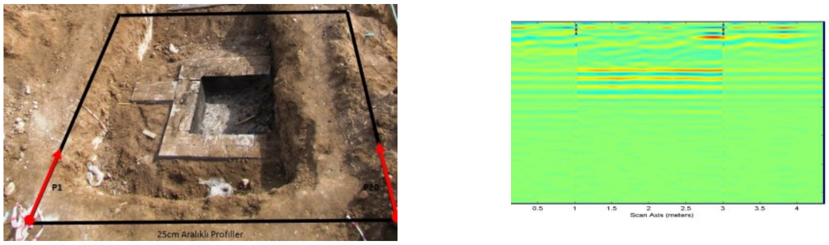


Figure 3.1 (a) The wall structure and (b) Sample profile

20 parallel profiles of the wall structure were acquired from the test area. The profile length was about 5 m. The distance between the profiles was taken as 0.25 m. The direction of the recorded profiles are shown in Fig.3(a), and the n^{th} $(n = 1,2, ..., 20)$ profile image of the implementation data set is shown in Fig.3(b). Each of $M=188$ traces consists of $K=512$ sampling values. 7 sampling values were selected from the m^{th} $(m = 1,2,\dots,188)$ trace at random intervals (e.g. [122-128], [172-178], etc.). First, the middle original sampling value (the fourth of 7) is removed, and then the new

sampling value instead of removed one was calculated using interpolation methods on the base of neighbor sampling values, and finally the sampling values of original and new sampling value were compared.

In order to decide for the appropriate interpolation method, the similarity ratio between the new sampling value and the original sampling value was used. The obtained results are given in Table 1.

Table 1. Similarity ratio

Interpolation Method	Removed Values#	Two neighbor values are used *(I=2)*	Four neighbor values are used *(I=4)*	Six neighbor values are used *(I=6)*
Standard Cubic	25	0.9415	0.8159	0.8614
	75	0.9297	0.8311	0.7703
	125	0.9302	0.8343	0.7721
	134	0.9324	0.8654	0.8162
	175	0.9239	0.7844	0.7761
Standard Spline	25	0.9415	0.7212	0.2199
	75	0.9297	0.7645	0.2729
	125	0.9302	0.7870	0.3074
	134	0.9324	0.8469	0.3871
	175	0.9239	0.7058	0.2391
Standard Linear	25	0.9415	0.7712	0.7607
	75	0.9297	0.7845	0.7456
	125	0.9302	0.8270	0.6864
	134	0.9324	0.8502	0.7844
	175	0.9239	0.7542	0.7243
Standard Median	25	0.9421	0.8793	0.9015
	75	0.9700	0.9527	0.9414
	125	0.9768	0.9538	0.9500
	134	0.9868	0.9743	0.9703
	175	0.9538	0.9137	0.9035
Standard Mean	25	0.9421	0.9226	0.9122
	75	0.9700	0.9567	0.9478
	125	0.9768	0.9662	0.9550
	134	0.9908	0.9787	0.9690
	175	0.9538	0.9295	0.9146
Proposed Mean	25	0.9742	0.9423	0.9316
	75	0.9823	0.9717	0.9524
	125	0.9910	0.9856	0.9821
	134	0.9967	0.9903	0.9814
	175	0.9798	0.9539	0.9418

Each trace of the wall structure consists of *K=512* sampling values. The 25th, 75th, 125th, 134th and 175th sampling values were assumed as incomplete ones and removed from traces. Instead of removed sampling values, the new sampling values were produced by interpolation methods using two, four and six neighbor sampling values of the complete sampling values. The similarity ratios of the new sampling values with the original sampling values are given in Table 1. Based on these results, according to PCC ratio the highest similarity is obtained by the mean interpolation method using two neighbors. Looking at the other results (four and six neighbor values), the further away from the neighbors of the values to be produced, the similarity ratio more likely falls. Therefore, choosing the two nearest neighbors gives the highest similarity ratio.

4. Results

There are two main factors that affect success of GPR research and applications. These are data collections parameters and search area properties. Data collection parameters which are under the control of users and it can arrange optionally. Search area properties which are outside the control of users are important factors for analyzing, processing, visualization and interpretation of underground structures. Factors which are not in the user s control such as failures occurred during data acquisition stage, obstacles on and under the surface, etc. may lead to incomplete sampling values or missing sampling values in that region. In this context, various interpolation methods have been applied to the data acquired from the concrete test area to investigate the accuracy and the completeness of produced missing data. The results and findings are listed below.

When determining the optimal interpolation method, the produced sampling value is compared with the original sampling value extracted from the data. In determining step, the technique produced closest result to original is taken into account. While the standard and proposed interpolation methods are applied, interpolation is applied with different number of neighbor values (two, four, six). Interpolation technique by using the two closest neighbor values gives higher similarity ratio than other techniques. Similarity ratios of proposed method with two neighbor values are approximately 93-99%. On the other hand, in interpolation with more value, although samples are taken from most of the collected data, the similarity ratio reduces. Because the data values can change, as the distance between the data increases. In summary, when new data is produced instead of missing or uncollected data in archaeological research and applications, proposed mean interpolation technique with the two nearest neighbor values gives the highest similarity ratio.

In future works, the Kriging, IDW and other interpolation techniques will be investigated to produce the incomplete GPR data.

Acknowledgement. This work has been funded by The Scientific and Technological Research Council of Turkey - TÜBİTAK under grant 5130012 project and 2211-E research fellowship program.

References

[1] Annan, A.P. 2009. Electromagnetic Principles of Ground Penetrating Radar. In Ground Penetrating Radar: Theory and Applications, edited by Harry M. Jol, pp. 3-40. Elsevier, Amsterdam.

[2] Benesty, J., Chen, J., Huang, Y. and Cohen, I. 2009. Pearson correlation coefficient. Noise reduction in speech processing. Springer Berlin Heidelberg, 1-4.

[3] Bristow, C.S. and Jol H.M. 2003. GPR in sediments: advice on data collection, basic processing and interpretation, a good practice guide. Geological Society: London, Special Publication 211; 9-28.

[4] Cassidy, N.J. 2009b. Electrical and Magnetic Properties of Rocks, Soils and Fluids. In: Ground Penetrating Radar: Theory and Applications, edited by Harry M. Jol, pp. 41-72. Elsevier, Amsterdam.

[5] Conyers, L.B. 2004. Ground-Penetrating Radar for Archaeology. AltaMira Press, Lanham. Dojack, L. 2012. Ground Penetrating Radar Theory, Data Collection, Processing, and Interpretation: A Guide for Archaeologists, 7-9.

[6] Dojack, L. 2012. Ground Penetrating Radar Theory, Data Collection, Processing, and Interpretation: A Guide for Archaeologists, 7-9.
[7] Leckebusch, J. 2003. Ground-penetrating Radar: A Modern Three-dimensional Prospection Method. Archaeological Prospection. Vol 10; 213-240.

[8] Levinson, N. 1947. The Wiener RMS (root mean square) error criterion in filter design and prediction.

[9] Maeland, E. 1988. On the comparison of interpolation methods. Medical Imaging. IEEE Transactions 7(3), 213-217.

[10] McKinley, S. and Megan, L. 1998. Cubic spline interpolation. College of the Redwoods 45(1), 1049-1060.

[11] Meijering, E. and Michael, U. 2003. A note on cubic convolution interpolation. IEEE Transactions on Image Processing 12, 477-479.

[12] Ozkan, M. and Samet, R. 2017. Interpolation Techniques to Recover the Incomplete GPR Data. In: The 16th International Conference Geoinformatics, Kiev, Ukraine.

[13] Safont, G., Salazar, A., Rodriguez, A., Vergara, L., 2014. On Recovering Missing Ground Penetrating Radar Traces by Statistical Interpolation Methods. Remote Sensing 6, 7546-7565.

[14] Samet, R., Çelik, E., Şengönül, E., Tural, S. and Özkan, M. 2015. Interpolation approach to search hidden result in GPR data. In: The 5th International Conference on Control and Optimization with Industrial Applications, Baku, Azerbaijan. 422 –425.

[15] Samet, R. and Özkan, M. 2016. Incomplete Data Production Methods in GPR Research and Applications. In: The 15th International Conference Geoinformatics, Kiev, Ukraine.

[16] Samet, R., Çelik, E., Tural, S., Şengönül, E., Özkan, M., Damcı, E. 2017. Using interpolation techniques to determine the optimal profile interval in ground-penetrating radar applications, Journal of Applied Geophysics (140), pp. 154-167.

[17] Zhao, W., Tian, G., Forte, E., Pipan, M., Wang, Y., Li, X., Shi, Z., Liu, H., 2015. Advances in GPR data acquisition and analysis for archeology. Geophysical Journal International 202, 62-71.

Current Address: Merve ÖZKAN: Ankara University, 50.Yıl Campus, Computer Engineering Department, Ankara TURKEY
E-mail Address: *merveozkan@ankara.edu.tr*
ORCID: *https://orcid.org/ 0000-0002-1071-2541*

Current Address: Refik SAMET: Ankara University, 50.Yıl Campus, Computer Engineering Department, Ankara TURKEY
E-mail Address: *samet@eng.ankara.edu.tr*
ORCID: *https://orcid.org/0000-0001-8720-6834*

Available online: December 20, 2017

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 59, Number 2, Pages 21-28 (2017)
DOI: 10.1501/commua1-2_0000000101
ISSN 1303-6009
http://communications.science.ankara.edu.tr/index.php?series=A2

COMMUNICATIONS
SERIES A2-A3

DETECTION OF MUSCLE FATIGUE: RELATIVE STUDY WITH DIFFERENT METHODS

MOHAMED REZKI, ISSAM GRICHE, ABDELKADER BELAIDI AND MOULOUD AYAD

ABSTRACT. This paper aims to investigate problem of muscle fatigue through the application of a comparative study by different processing techniques in order to see the effect of physical exercise on Electromyography characteristics. Indeed, electromyography is the best physiological examination to study muscle activity and it is translated by an electromyogram (EMG). The analysis of the biomedical signal "EMG" before and after physical exercise allowed us to quantify the physical effort and give some diagnostic elements that can help the practitioner. We applied some signal processing techniques to quantify the physical effort and therefore we were able to identify and detect muscle fatigue.

1. INTRODUCTION

EMG is the process by which an examiner puts a needle (or electrode) into a particular muscle and studies the electrical activity of that muscle (this electrical activity comes from the muscle itself) [1].

The EMG is a signal that must be recorded, so there are several electronic acquisition platforms (Arduino, PowerLab from ADInstruments Ltd, Data Acquisition card “DAQ” from National Instruments, etc.). We have chosen as an acquisition platform the Arduino mega card for its availability (see figure 01). This card must be connected through electrodes on the surface of the skin because the EMG is a non-invasive examination (Figure 2).

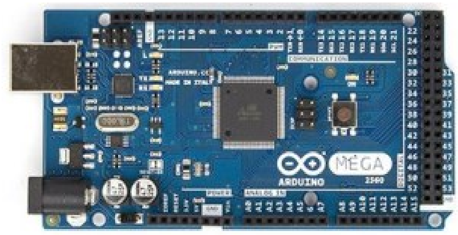


Figure 1. Model of Arduino Mega 2560 R3 Front [2].

Received by the editors: November 11, 2017; Accepted: December 17, 2017.
Key word and phrases: Muscle fatigue, EMG, FFT, Wavelet

Figure 2. EMG Electrodes with electrodes.

To study the effect of physical exertion on the characteristics of the EMG, there are some typical physical exercises that take place in the research laboratory. We chose a physical exercise that causes the muscular fatigue of the hand (see figure 03); the reference being open hand.

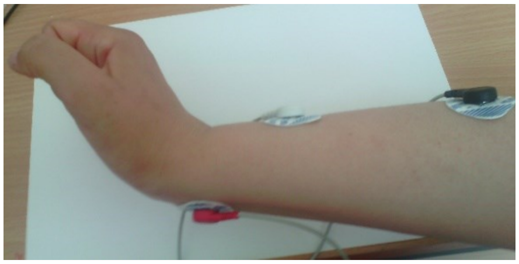


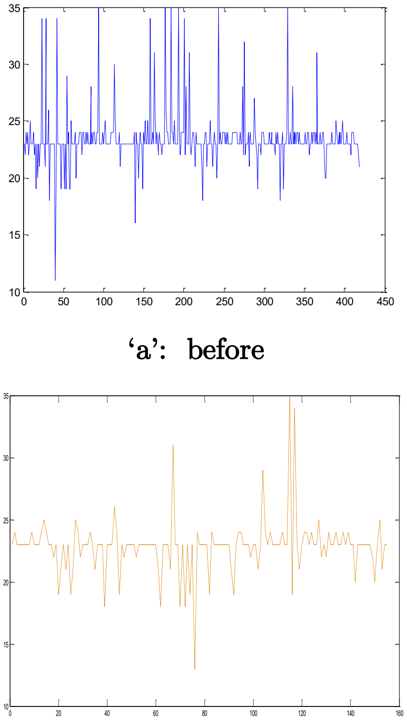
Figure 3. Physical exercise provoking hand's fatigue.

2. Materials and Methods

2.1 Data acquisition

After performing physical exercise (see Figure 03), we obtain the following results (Figure 04):

As a first observation, one sees from the shape of the acquired signals that physical exercise has caused a dilation of the signal, that is to say the peaks (representing the activity of the muscular fibers) become more and more spaced.



‘a’: before

‘b’: after

FIGURE 4. Acquired data before and after doing hard exercise.

2.2 Processing techniques

2.2.1 Fast Fourier Transform. The Fast Fourier Transform (FFT) is is commonly used in analyzing the spectral content of any deterministic biosignal, it’s a numerical approach for quick computation of the Discrete Fourier transform (DFT). The equation of DFT is as follows [3]:

$$X(k)=\frac{1}{N}\sum_{n=0}^{N-1}x(n)e^{-j\left(\frac{2\pi kn}{NT}\right)} \tag{2.1}$$

where: $X(n)$ is the input biosignal whose sampling period is T. The spectrum $X(k)$ is estimated at multiples of: fs/N, where fs is the sampling frequency.

2.2.2 Wavelet Transform technique. Wavelet is a wave-like oscillation with amplitude that begins at zero, increases, and then decreases back to zero. We have two major types of wavelet: continuous wavelet transform and discrete wavelet transform [4]. The continuous wavelet transform (CWT) is defined as [5]:

$$T_x(t,a;\Psi) = \int_{-\infty}^{\infty} x(s)\Psi^*_{t,a}(s)ds \qquad (2.2)$$

where $\Psi^*_{t,a}(s)$ is obtained from the mother wavelet $\Psi(s)$ by time translation t and dilation by scale a:

$$\Psi_{t,a}(s) = \frac{1}{\sqrt{|a|}}\Psi\left(\frac{s-t}{a}\right) \qquad (2.3)$$

a: is the dilation parameter.
s : is the translation parameter.

There are numerous analyzing wavelet called mother wavelets such as: Haar, Mexican hat, Morlet and Daubechies. This latter seems (Duabechies) the most used in the analysis of biological signals [6]. Its shape can be seen in the following figure:

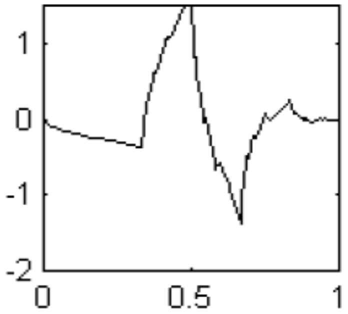


Figure 5: Diagram of the wavelets mother function “Daubechies”

3. Results and Discussion

3.1. Fast Fourier Transform

By applying the Fast Fourier Transform technique on the primary signals (raw data), we get the following curves:

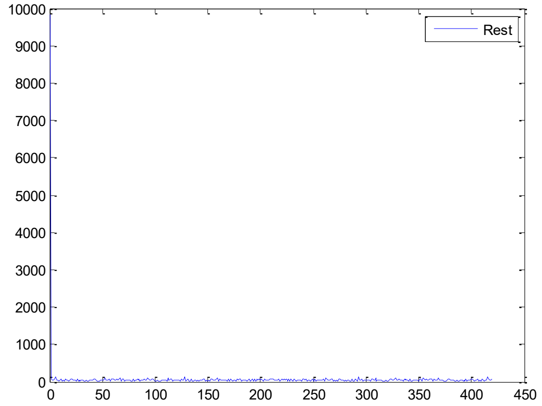


FIGURE 6. FFT of Rest's case (without effort).

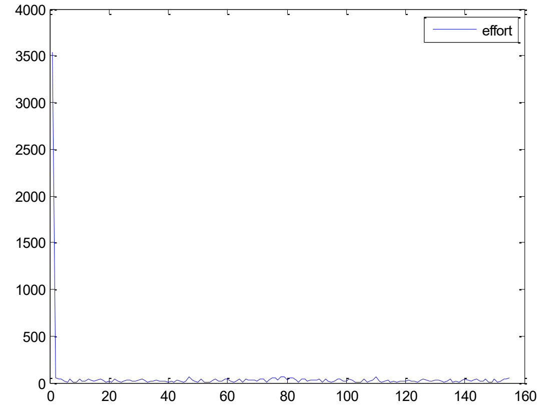


FIGURE 7. FFT of effort's case (with effort)

From the figures, we note that the effort incite an extension of the signal intensity (to 4000) compared to the signal of rest. But in general the FFT technique gives little information for the analysis of biomedical signals.

3.2. Technique of wavelet

We have chosen the Daubechies continuous wavelet for treating our raw data. After running the Matlab program, we obtained these signals (Figures 8 and 9):

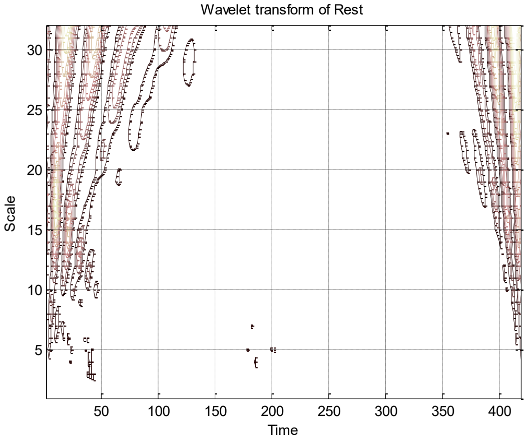


FIGURE 8. The wavelet transforms of an EMG signal (scalogram of Rest's situation).

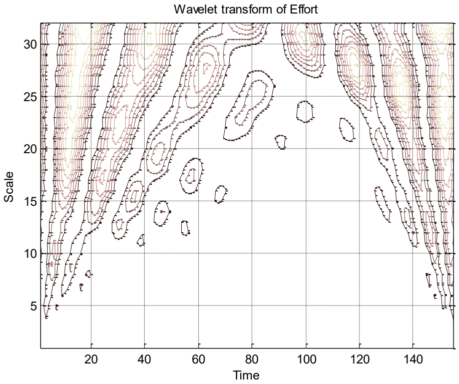


FIGURE 9. The wavelet transforms of an EMG signal (scalogram of Effort's situation).

The first result is concentration of wavelet coefficients on the peaks of the EMG signal.
Another finding, we see clearly that the signal of effort's wavelet is very rich in coefficients and in energy. So we can deduct from the shape of the signal confirms or not if there is fatigue.

3.3. Complementary study “Entropy”

There are complementary techniques to wavelet for processing, based in general on Shannon wavelet entropy. The modified Shannon wavelet entropy is computed by [7]:

$$H^k(f_b) = -\sum_{i=1}^{M} P_i^k \log P_i^k, \quad \sum_{i=1}^{M} P_i^k = 1, \qquad (3.1)$$
$$f_c = k \in [J, K]$$

where, P^k_i is the distribution sequence obtained from wavelet coefficients. P^k_i i is calculated by [7]:

$$P_i^k(f_b) = |W_x(m,n)| / \sum_{j=1}^{M} |W_x(m,n)| \qquad (3.2)$$

By calculating the maximum of Shannon entropy coefficient for the two cases (before and after getting fatigue ‘effort’), we get this table:

Table 1. Maximum of Shannon entropy

Rest	**Fatigue**
-1.4791e+006	-5.1385e+005

We can deduct from the table that the fatigue’s muscle signal’s entropy is higher than the rest case. This signal of fatigue contains a lot of different information (very rich in information).

4. Conclusion

The analysis of EMG signal through the processing tools offered by Matlab is very useful to detect muscle fatigue. In this paper, we applied some signal processing techniques to quantify the physical effort and therefore detect muscle fatigue.

The application of wavelets especially those of the Daubechies seems the best. A complementary study based on calculation of Shannon entropy was done and it has given us some explanations to wavelet results.

We hope to be able in the future to apply other digital processing techniques and to derive other results in order to identify the problem of muscular fatigue.

References

[1] J. M. Weiss, L.D Weiss and J.K. Silver, *Easy EMG* . Elsevier Inc., First edition, Philadelphia, USA, (2004) 4.

[2] R Arduino card. Retrieved: July, 12, 2017. Available at: http://arduino.cc/.

[3] E. M. Kutz. *Standard Handbook of Biomedical Engineering and Design.* McGraw-*H*ill Companies,Inc., First edition, New York, USA, (2003) 447-448.

[4] M.Rezki & al., *De-noising a Signal's ECG sensor using various Wavelets Transforms and other analyzing techniques*. International Journal of Applied Engineering Research, (2013) 589-599.

[5] K. J. Blinowska, J. Zygierewicz, *Practical Biomedical Signal Analysis Using MATLAB*. CRC Press, Taylor & Francis Group, LLC, First edition, New York, USA, (2011) 54.

[6] J. Rafiee & al., *Wavelet basis functions in biomedical signal processing*. Expert Systems with Applications, vol.38 , (2011) 6190–6201.

[7] M.Ayad , D.Chikouche , N.Boukezzoula N and M.Rezki, *Search of a robust defect signature in gear systems across adaptive Morlet wavelet of vibration signals*. IET Signal Processing, 8/9, (2014) 918 –926.

Current Address: MOHAMED REZKI, *ISSAM GRICHE*[1], *MOULOUD AYAD*:
Department of Electrical Engineering, University of Bouira, 10000 Bouira, ALGERIA
E-mail Address: M. REZKI (Correspoding author) m.rezki@univ-bouira.dz
ORCID: https:// orcid.org/0000-0003-2406-1644

Current Address: ABDELKADER BELAIDI: [2]*Department of electrical engineering, ENPO-ORAN -ALGERIA*

Available online: December 21, 2017

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 59, Number 2, Pages 29-40 (2017)
DOI: 10.1501/commua1-2_0000000102
ISSN 1303-6009
http://communications.science.ankara.edu.tr/index.php?series=A2

COMMUNICATIONS
SERIES A2-A3

3D VISUALIZATION APPROACH TO GPR DATA

MERVE ÖZKAN OKAY and REFIK SAMET

ABSTRACT. Ground Penetrating Radar (GPR) is used to acquire data from near-surface depth for archeological, infrastructural, etc. researches and applications. Acquired data allow users to visualize and interpret the underground structures with high accuracy. The 3D visualization of the underground structures is one of the most problem for GPR research and applications. Usage of the suitable approach for 3D visualization will increase the accuracy of visualization and interpretation of underground structures. In order to contribute to this problem, an approach is proposed. Firstly, GPR data are acquired from the search area and data preprocessing steps are applied to GPR data. Secondly, the incomplete or missing data are recovered using interpolation techniques. Thirdly, the GPR data corresponding to the underground structures or anomalies are extracted and placed in a 3D cube. Finally, the extracted GPR data are visualized in 3D environment. The proposed approach was implemented on the real GPR data acquired from the test area. The results showed that created 3D models of the underground structures are very close to real model.

1. INTRODUCTION

In general, in GPR research and applications, the search area is squared and scanned to acquire the raw data [3], [6]. The acquired raw data are preprocessed for interpreting the underground structures. One of the most important problems GPR research and applications is 3 dimensional (3D) visualization of hidden structures (anomalies) in GPR data [2]. This study proposes a method for contributing the mentioned problem. The proposed method consists of four steps. In the first step of proposed method, raw data is acquired from search area. This acquired raw data were previously processed using standard data processing techniques (trace editing, spectral analysis and band-pass filtering, highpass filtering, background removal, gain and migration etc.). In the second step, interpolation techniques are used to recover the missing data [11-13]. After completing the missing data, filter is applied without disturbing the resolution of the profiles to remove meaningless spots or the noises caused by electromagnetic waves. In the third step, the underground structures are extracted from GPR data and placed in a 3D cube. In the final step, the extracted

Received by the editors: October 27, 2017; Accepted: December 18, 2017.
Key word and phrases: Ground Penetrating Radar (GPR), GPR Data, Underground Structure, Anomalies, Profile, Trace, Sampling Value, Preprocessing, Interpolation, 3D Visualization.

part of underground structure (anomaly) are visualized in 3D environment by adding volume obtained 2D model. The 3D model of underground structures viewed from different angles.

This study is organized as follows. In Section 2, the proposed approach to 3D visualization and interpretation of the GPR data is described. The implementation results using real GPR data are discussed in Section 3. The obtained results are summarized in final section.

2. PROPOSED APPROACH

GPR data consist of N parallel profiles. Each profile consists of M traces. Each trace consists of K sample values (Fig.2.1).

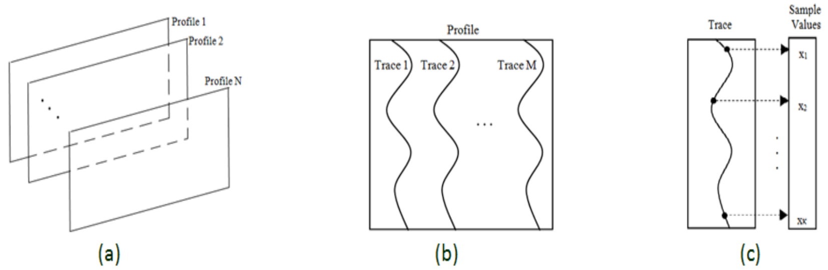


Figure 2.1 (a) Profiles of GPR data; (b) Traces of profile; (c) Sample values of trace

The proposed approach for visualizing and interpreting the underground structures (anomalies) in the 3D environment consists of four steps:

1. Obtaining GPR data and data processing;
2. Recovering the missing GPR data;
3. Extraction of underground structures (anomalies);
4. 3D Visualization of underground structures.

2.1. Obtaining GPR data and Data Processing

It is necessary to use data sets obtained from the search area where the underground structure is well known for testing and verification of proposed technique [15]. The test area is created for obtaining this type of data set. The various sizes materials (such as, pipes and stone walls, metal and plastic drums) were placed in this test area and obtained data from there.

Generally, wall structures are investigated in GPR research, especially in archeological applications. The wall structure used in this study was placed to lie within an area of 4x4 square meters. A wall structure with dimensions of 2x2 square meters is included. The depths of the upper and lower surfaces of the object are defined to be 0.7 meters and 1.2 meters, respectively. A 3D view of the model is provided below (Fig. 2.2). The wall structure was scanned parallel lines for obtaining raw data.

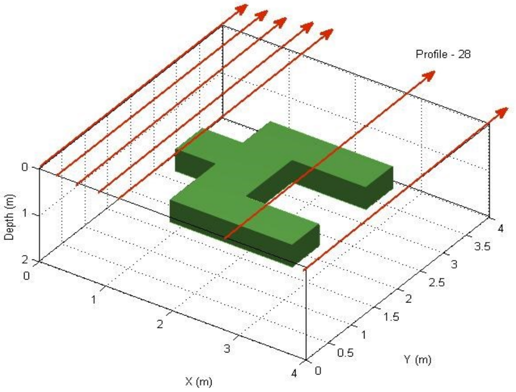


Figure 2.2. 3D view of the wall structure model used in this study

Software such as GPRMax 2D/3D and MatGPR [1], [5] can be used for the processing of data. In this study, the GPRMax 2D/3D software package was used to obtain the appropriate data from raw data. During data collection, the main concern should be to select the profile orientation such that the profiles will intersect the structure perpendicularly. Different users will apply different data processing stages. Generally, certain data processing steps, such as “dewow”, “gain”, “filters”, and “background removal”, are applied to all data. If hyperbolas appear in the radargrams, then “migration” processing should be applied. In this study, the MatGPR software package was used for data processing, and “gain”, “background removal” and “migration” procedures were applied to the profiles.

2.2. Recovering the Missing GPR Data

According to the size of the search area and the data acquisition parameters, different numbers of profiles are acquired. The acquired GPR data may be incomplete and insufficient, due to the some properties of the search area (rugged surface, obstacles, characteristics of underground structures, etc.). The search area properties are outside of the control of users and it is affected by any physical or chemical change in search area. In order to visualize underground structures (anomalies) more clearly, either data should be acquired at small intervals or interpolation methods are applied [7], [9], [11-13]. The advantage of collecting data with a small profile interval is that doing so allows the geometry of underground structures to be visualized at high resolution. However, this benefit comes with the disadvantage of increasing the cost and processing time. Because of mentioned problem, interpolation methods are applied for producing or recovering missing data.

2.3. Extraction of Underground Structures (Anomalies)

After data processing and recovering, in regions without anomalies, the sampling values in the profiles consist of values close to each other according to the properties of the environment. The anomalies are extracted from the profiles by using sampling value feature. The sampling value in GPR data very close to each other without anomalies, otherwise, these values are quite different [10]. The averages and standard deviations of the sampling values of the profiles are calculated to extract anomalies. The threshold value is calculated by using these average and standard deviation values (Eq. (2.1)).

$$t^n = mean(P^n) + std(P^n),\ n = 1,2,\dots,N \tag{2.1}$$

where, t is threshold value for extracting anomalies, N is the total number of obtained profiles from the search area and P is the examining profile to extract anomalies. Sampling values above or under the threshold value are extracted and placed in a 3D cube.

2.4. Visualization of Underground Structures in 3D Environment

The sampling values representing the anomalies extracted from the profiles by using sampling values properties (Eq. #1) are placed in a 3D cube. Thus, 2D visualization of underground structure is obtained. During the extraction operation, some meaningless points can be extracted from profiles and because of this the geometry of underground structures is affected negatively. In order to solve mentioned problem meaningless points are removed from image. After that, smoothing is applied to soften sharp edges and regions. The 3D image is created by adding the volume to the whole 2D image [8], [10], [14]. After the 3D image is created, the colors of isosurface are calculated using the color values and the image is colored.

3. IMPLEMENTATION OF PROPOSED APPROACH

3.1. Used Data Set

In this section, the results of applying the proposed methodology to real data are presented. For this purpose, the data acquired from a test area specifically designed and created. The test area is located on the Golbasi Campus of Ankara University. The layout plan for the objects embedded in the test area is illustrated in Fig. 3.

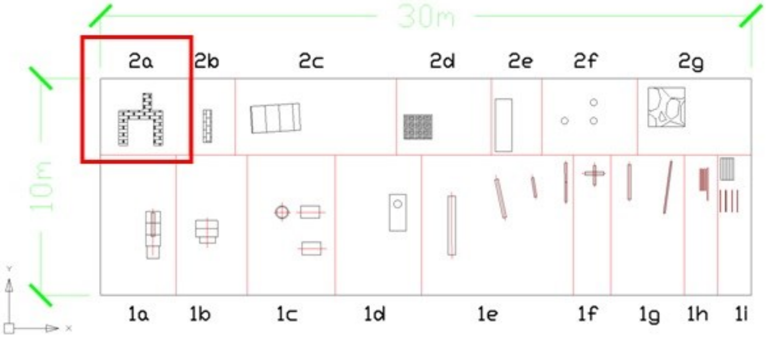


FIGURE 3.1. Object layout plan in the test area

A real image of the wall structure represented by the object numbered 2a in Fig. 3.1 is shown below (Fig. 3.2). 20 parallel profiles of the wall structure were acquired from the test area. The profile length was about 5 m. The distance between the profiles was taken as 0.25 m. The direction of the recorded profiles is shown in Fig. 3.2.

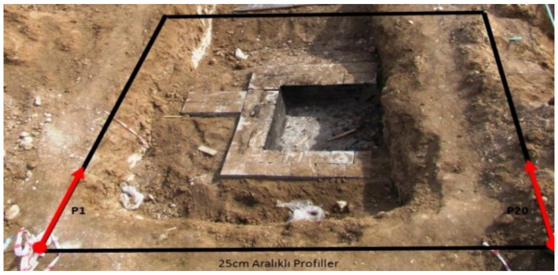


Figure 3.2. The wall structure and directions of profiles

The wall structure has the following properties:
- Width: 200 cm
- Length: 200 cm
- Thickness: 60 cm
- Embedded base depth: 120 cm
- Embedded ceiling depth: 70 cm.

First, marker alignment and direction editing operations were applied to the acquired data. The remaining data processing operations such as DC shift, static correction, linear gain, band-pass filter, background removal, were performed.

3.2. Implementation Results

According to the proposed approach in this study, firstly, data preprocessing steps and interpolation were applied to recovering the missing data. The sharp regions on the anomaly were softened and gaps due to data loss were filled. After recovering the data, filter is applied to remove outliers and noises. Figure 3.3 (a) and (b) show the wall structure slice before and after preprocessing and interpolation, accordingly.

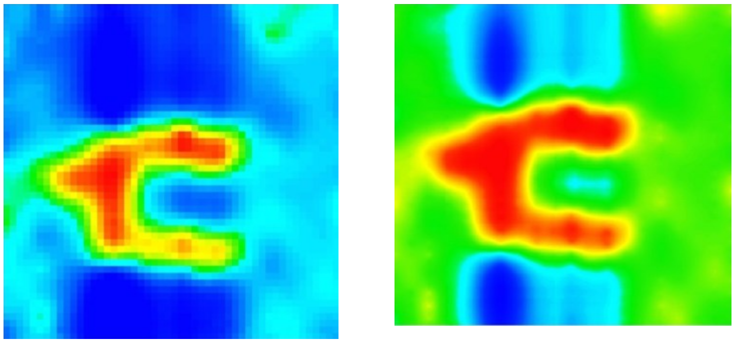

(a) (b)

Figure 3.3. a) The wall structure slice before preprocessing and interpolation; b) The wall structure slice after preprocessing and interpolation.

Secondly, regions with anomalies extracted from the profiles by using the calculated threshold value according to Eq. #1. The extracted regions were placed in 3D cube. After this process, it can be understood that the 2D view is the wall structure (Fig. 3.4).

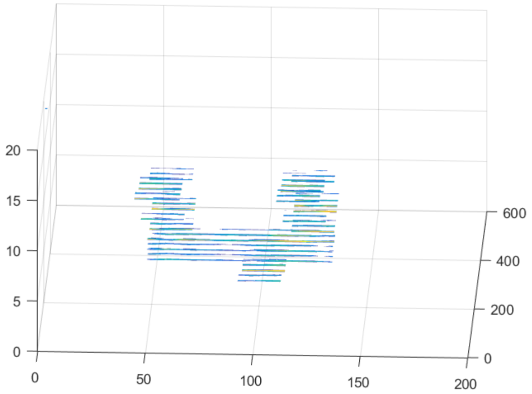


Figure 3.4. Anomalies in the cube

The 2D wall regions obtained from each profile were placed in a 3D cube. The obtained 2D image was combined as a whole by patch graphics. A patch graphics object is composed of one or more polygons that may or may not be connected. Patches are useful for modeling objects and for drawing 2- or 3-D polygons of arbitrary shape.

The wall structure was created, by specifying the coordinates of anomalies vertices/edges and some form of color data. After patching operation, wall image was completed as a whole. The obtained patched image is shown in Fig. 3.5.

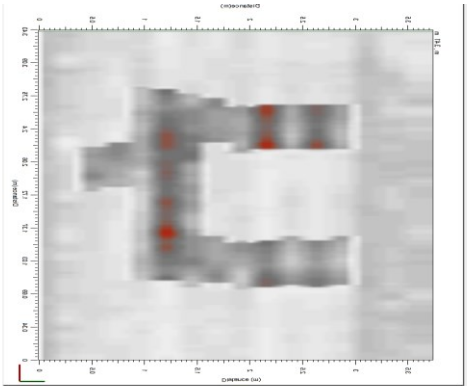


Figure 3.5. The patched wall structure image

After the patching operation, image resolution was affected negatively. In order to solve this problem, filter and smoothing is applied to obtained wall structure image. Finally, by adding volume and colored to the image processed, 3D image of underground structure was created. Different views of the created wall structure are given in Fig. 3.6.

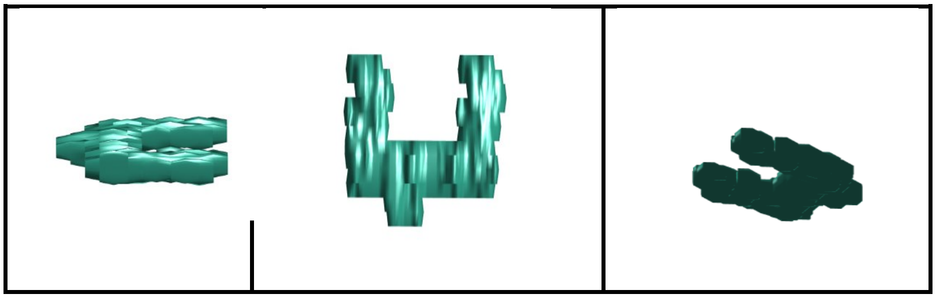


Figure 3.6. 3D view of wall structure from different angles

4. RESULTS

GPR is a widely used method to investigate the underground structures in near surface depth. The use of GPR method has been increasing in recent years as it has detected underground structures quickly and accurately. In GPR studies, 3D visualization of GPR data play a vital role in the identification of underground structures accurately. GPR data are used to investigate underground structures (anomalies). Thanks to this data, different underground structures (anomalies) are detected, examined and visualized. In the considered context, 3D visualization approach is proposed and the obtained results are summarized below.

One of the most important problems in GPR research and applications is visualization and interpretation of the underground structures with high accuracy. In order to contribute to this problem, four-step approach is proposed to visualize and interpret the underground structures. Firstly, the real data acquired from created test area and preprocessing operations are applied. Secondly, the missing data in the GPR profiles, traces and sampling values were recovered and the anomalies were clarified with interpolation techniques. Subsequently, the anomalies were extracted from the profiles and placed 3D cube. Finally, the anomalies were visualized in 3D environment with added volume. In summary, 3D visualization was performed to visually investigate and interpret the information of anomalies such as type of structure and its depth. By inspecting the 3D model of underground structures from different angles users can interpret anomalies with high accuracy. The obtained results have showed that visualized 3D model of the underground structures are very close to real model.

Acknowledgement

This work has been funded by The Scientific and Technological Research Council of Turkey - TÜBİTAK under grant 5130012 project and 2211-E research fellowship program.

REFERENCES

[1] Arkedani, M.R., Neyt, X., Benedetto, D.,Slob, E.,Wesemael, B., Bogaert, P., Craeye, C., Lambot, S., 2014. Soil moisture variability effect on GPR data. 15th International Conference on Ground Penetrating Radar, Brussels, Belgium, p. 214-217.

[2] Conroy, J.P. and Radzeviciu, S.J. 2003. Compact MATLAB code for displaying 3D GPR data with translucence. Computers & Geosciences 29, 679–681.

[3] Conyers, L.B. 2004. Ground-Penetrating Radar for Archaeology. AltaMira Press, Lanham

[4] Dojack, L. 2012. Ground Penetrating Radar Theory, Data Collection, Processing and Interpretation: A Guide for Archaeologists, 7-9.

[5] Giannopoulos, A., 2005. Modelling ground penetrating radar by GprMax. Construction and Building Materials 19, 755–762.

[6] Goodman, D., Piro, S., Nishimura, Y., Schneider, K., Hongo, H., Higashi, N., Steinberg, J., Damiata, B., 2009. GPR archaeometry, In: Jol, H.M. (Ed.), Ground Penetrating Radar: Theory and Applications. Elsevier, Amsterdam, pp. 479-508.

[7] Intyas, I., Suksmono, A.B., Munir, A. 2016. Image Quality Improvement for GPR Acquisition Using Interpolation Method. In: The 22nd Asia-Pacific Conference on Communications, Yogyakarta, Indonesia.

[8] Kadioglu, S. and Ulugergerlic E.U. 2012. Imaging karstic cavities in transparent 3D volume of the GPR data set in Akkopru dam, Mugla, Turkey. Nondestructive Testing and Evaluation, 27(3), 263–271.

[9] Maeland, E., 1988. On the comparison of interpolation methods. IEEE Transactions on Medical İmaging. 7, 213-217.

[10] Nuzzo, L., Leucci, G., Negri, S., Carrozzo, M.T. and Quarta,T. 2002. Application of 3D visualization techniques in the analysis of GPR data for archaeology. Annals Of Geophysıcs, 45(2), 231-337.

[11] Ozkan, M. and Samet, R., 2017. "Interpolation Techniques to Recover the Incomplete GPR Data". In: The 16th International Conference Geoinformatics, Kiev, Ukraine.

[12] Samet, R. and Özkan, M. 2016. Incomplete Data Production Techniques in GPR Research and Applications. In: The 15th International Conference Geoinformatics, Kiev, Ukraine.

[13] Samet, R., Çelik, E., Tural, S., Şengönül, E., Özkan, M., E., Damcı, "Determining the optimal profile interval by using interpolation techniques in GPR applications", J.Appl. Geophysics, 2017.

[14] Sun, W., Xu, Q., Zhang, H., Yao, Z. 2012. Research on Detection and Visualization of Underground Pipelines. In: The 2nd International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE), Nanjing, Jiangsu, China.

[15] Zhao, W., Tian, G., Forte, E., Pipan, M., Wang, Y., Li, X., Shi, Z., Liu, H., 2015. Advances in GPR data acquisition and analysis for archeology. Geophysical Journal International 202, 62-71.

Current Address: Merve ÖZKAN: Ankara University, 50.Yil Campus, Computer Engineering Department, Ankara TURKEY
E-mail Address: merveozkan@ankara.edu.tr
ORCID: https://orcid.org/ 0000-0002-1071-2541

Current Address: Refik SAMET: Ankara University, 50.Yıl Campus, Computer Engineering Department, Ankara TURKEY
E-mail Address: samet@eng.ankara.edu.tr
ORCID: https://orcid.org/0000-0001-8720-6834

Available online: December 21, 2017

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 59, Number 2, Pages 41-54 (2017)
DOI: 10.1501/commua1-2_0000000103
ISSN 1303-6009
http://communications.science.ankara.edu.tr/index.php?series=A2

A NEW SIMILARITY COEFFICIENT FOR A COLLABORATIVE FILTERING ALGORITHM

OZGE MERCANOGLU SINCAN and ZEYNEP YILDIRIM

Abstract. Recommender systems give the opportunity to present automatically personalized content across many digital marketing channels to visitors depending on visitor movements on the site. In recent years, there has been a lot of interest in e-commerce companies in order to offer personalized content. So, recommender systems become very popular and many studies have been done in this regard. New works are being done day by day to improve the results. In this paper, we propose a new memory-based collaborative filtering algorithm. Calculation of similarities between items or users is a critical step in memory-based CF algorithms. Therefore, we proposed a new function for calculation of similarities based on user ratings. In this study the more similar the user s pleasures are, the more similar it is to the products the users choose, is adopted. The adopted idea in this study is that the more similar the user s pleasures are, the more similar products are chosen. We estimate the degree which a user is interested in X product. To do this, we find other users who are interested in product X and calculate the similarity ratios of those users to the user. We tested our algorithm in MovieLens 100K dataset and compared to other similarity functions. We used MAE and RMSE measures in our experiments.

1. Introduction

Recommendation systems (RSs) are intelligent information filtering engines that shorten the decision-making process. These systems are the leading parts of the user experience head of our favorite platforms. RSs are fed from both open and closed interactions. Open interactions are your preferences on your platform, your scorecards, your comments, or the information you give when you create a profile. Closed interactions are clicks, purchases, and searches. Recommendation engines predict the user or client s online interests by looking at the data below the interactions in these two concepts. Correct personalization always evaluates content and real user claims together. The ultimate goal of these systems is to maximize the user experience by influencing "moment of truths". These important and short

Received by the editors: October 18, 2017; Accepted: December 20, 2017.
Key word and phrases: Recommendation systems, collaborative filtering, user similarity.

moments of decision depend on what the user really needs, the company s communication and the information he or she can reach.

Nowadays, we can always encounter a recommendation mechanism on the internet. Suggestions are highly consistent suggestions that typically show up by recommendation mechanisms, such as Google, Youtube, where you guess what you re looking for when you re doing a search, or what other people download the same song while downloading a song from iTunes. Within e-commerce firms, advice mechanisms based on estimations and ease of travel needs will contribute to customer loyalty. Today, Amazon.com s content changes dynamically based on recommendation systems. Even Amazon.com sends the potential product to logistics centers close to the delivery address, prior to the purchase, by looking at previous orders, product searches, basket movements, to ensure orders are delivered much sooner. Therefore, RSs have become very popular in recent years. It s so popular that even site ads are now user-specific. When you visit a site, you can easily reach the prices of similar products or different sales places you have searched for before. They are applied in various applications.

RS can be considered under 3 main headings: collaborative filtering (CF), content-based filtering (CBS), and hybrid recommender systems.

CF is one of the most successful RS techniques. The main purpose of CF systems is to determine which product a particular user likes, using the user s knowledge of the products. CF systems generally use data sets that contain user information and users interest in products. There are many challenges for CF tasks; problems with very rare data, scaling with an increasing number of users and items, satisfying short-term recommendations and being able to cope with other problems such as synchronicity, data breaches and privacy issues. Also CF systems cannot recommend for new users and items. Early-generation CF systems use the product ratings of users to calculate the similarity between users or products. Then, based on these calculations, they are predicting. Memory-based CF systems are the most preferred system by companies. The reason for this is that its application is simple and largely effective. Thanks to CF systems, the effort that users spend searching for a product is diminishing. This brings the firm; customer loyalty, high sales, more ads. However, CF systems are inadequate. CF systems do not work efficiently if the user in the dataset has very low product rating data or if the target user has limited common products available to other users. Model-based CF approaches have been proposed to solve the inadequacy of memory-based CF systems and to increase the efficiency even further. Model-based CF approaches use product rating data. Each user s score is calculated according to their individual product ratings. Machine learning is often used for calculations.

Along with collaborative filtering, content-based filtering is also commonly used in recommendation systems. The features of the items are used to make a suggestion in content based recommendation systems. In such systems, the user is advised of new items that have the common features of the user s past preferences.

Hybrid recommender systems is the another main headings of RS. In this approach, content-based filtering and collaborative filtering methods are used together. The goal is to get rid of as much as possible the disadvantages of having a single method and to combine the advantages of the methods. Content-based and cooperative filtering methods can be used together in different ways.

In this paper, we applied memory-based CF system by using user-item ratings data. In memory-based CF systems, similarity calculations between users or products are very important. While the similarity between product a and product b is calculated for product-based CF systems, two of these products have user evaluations. On the contrary, for a user-based CF algorithm, first, a similarity value between u and v users that grades the same items is computed. There are many calculation methods to calculate similarity between users or products. The most commonly used similarity calculation methods are: Pearson correlation-based similarity, vector cosine-based similarity, distance-based similarity.

In this study, we improve the algorithm of our preliminary work. In [1], we estimated the rate by multiplying weight and users rating. In this work, we used weighted average method for getting better results and compare our algorithm with others.

The rest of this paper is organized as follows. Section 2 reviews previous studies regarding recommendation systems. Section 3 explains the proposed method. Section 4 shows the experimental results. The last section concludes our paper.

2. RELATED WORK

Interest in RS is increasing day by day. The main reason for this is the increase in the use of social media. Along with social media, different solutions are needed because the data that can be used for RSs are increased substantially. In this section, a brief summary of the work on RS will be presented.

After the CF method was found, a number of recommendation systems were created using this method. Tapestry [2], is the best known of these. This work was done in order to allow users to see the titles that attracted only the interest of users, in their e-mails. After Tapestry, GroupLens [3] was proposed in 1994. The GroupLens study was designed to make it easier for users to read news from the Internet. While users were reading the news, they could see the predicted values that they could give and then change it according to their own values. Thus, they contributed to the operation of the system. Until now, this method has been applied to many areas.

Breese et al. divided CF into two classes: memory-based, model-based. Similarity ratios, correlation coefficients and statistical methods were applied in this study. Correlation coefficients and statistical methods gave good results [4]. Herlocker et al., proposed an estimation method that is based on weighting according to the similarity coefficient of the degree of co-products between the target user and other users. [5]. In [6], a probabilistic structure was used. The solution to the problem of "new user" of CF systems is provided by the structure used. In addition, they have reduced the operating cost because they work on carefully selected user data. They took their results on two different datasets. In [7], they do not consider the general consistency relationship between users or products of existing memory-based CF systems. They suggest a self-learning system based on solving the problems that arise as a result of this approach. In this system, rational individual prediction is made by looking at the preferences and ratings of the users. The results were higher when compared to other methods available. Adamopoulos et. al. [8] proposed a new method for estimating prospective opportunities based on unknown ratings and weighted percentages. The proposed approach demonstrates the practical application of classical KNN in the context of neighborhood models that adapt the near neighbor method. In addition, they have conducted an empirical study that shows that the proposed method is better than the standard user-based collaborative filtering approach with a wide range of ratings in areas such as item forecast accuracy and discounted cumulative earnings normalized on F basis. Bulut et al. [9] proposed two new methods for the estimation step, which is the last step of collaborative filtering algorithms. The sparsity of rating matrix is always the major challenge which restricts the performance of collaborative filtering [10]. The cause of this problem is that the vector dimension of users or items is always very large. As the developing of machine learning algorithms, a method called matrix factorization is now the major method to decrease the sparsity of the matrix. Luo et. al. [11] focuses on non-negative matrix factorization (NMF)-based CF development with a single-element-based approach. The main idea is to replace the missing function of the standard distance with the sum of the square-errors and search for a non-negative update process depending on each relevant property parameter, instead of all property matrices. The experimental results in the four large industrial datasets show that their method can take advantage of the computational efficiency over NMF-based CF model. Hernando et. al. [12] presents a new technique for collaborative filtering based proposal systems.

As with the classical matrix multipliers, a vector of the K component is associated with each user and each item. However, unlike the classical matrix multipliers, the components of this vector vary in the range [0, 1], and this change provides significant advantages in the probabilistic sense. It is also the level at which techniques can compete with classical matrix multiplier separation techniques in terms of accuracy in estimates and recommendations. The works in [13,14] can be examined for detailed information.

Another common method is content-based filtering (CBS) algorithms. In this method, when a new product is proposed to the user, the similarity between the user s other products and the new product is checked. Content-based methods use information about the product to make suggestions. This, in turn, makes a great contribution to the proposal of the new product. In [15], a content based book recommendation system was developed. Machine learning has been used for word groups. It has been observed that this approach has the right recommendation. In [16], a method for solving problems arising from natural language ambiguity is proposed. In this method, it is suggested to classify semantic approaches from top to bottom and from bottom to top. This method has not been able to fully solve the problem of the words used according to synonyms and specialization areas. The work in [17] can be examined for detailed information.

The last method we will examine in this section is the hybrid systems. In this system, it is aimed to obtain more efficient results by using existing methods together. In [18], a hybrid system was created by using CF and CBS methods together. [19] suggests a new content-collaborative hybrid system. In the study, the similarity between users is calculated according to the content-based profiles of users. Machine learning is used when semantic profiling is being done. From the results, it is understood that the proposed system made successful estimates. The work in [20] can be examined for detailed information.

3. THE PROPOSED METHOD

In collaborative filtering approaches, data which includes m users ($u_1$to $\mathrm{u_m}$) and n items (i_1 to i_n) are converted to a user-item matrix. Table 1 shows a movie rating-matrix concerning five users and five items. As seen in the table, some values are missing. Here, CF estimates the missing values in these tables and recommend the users the items which the user can like.

TABLE 1. An example of a user-item matrix.

	i_1	i_2	i_3	i_4	i_5
u_1	2	?	1	1	4
u_2		2		3	5
u_3	5	4			1
u_4			3		2
u_5	1	3		2	4

Collaborative filtering algorithms can be analyzed in three stages: similarity function that used, neighborhood selection and estimation of rate.

3.1. SIMILARITY FUNCTIONS

In our preliminary work [1], we proposed a new similarity coefficient, sim(a, u)[1]. The formula of similarity coefficient between the active user u and the other user v, can be defined as equation (1) where $|uv|$ denotes the average absolute differences of the ratings for common rated items for both users, k denotes a constant. $|uv|$ and k can be calculated as in the equation (2) and (3) where C denotes the set of common rated items, $R_{u,i}$ and $R_{v,i}$ denote the ratings of users u and v on item i, U denotes set of users who rate for missing item of u and have common rated items with u. The similarity coefficient is inversely proportional to the difference between the persons. The greater sim(u, v)[1], the more similar the users are to each other.

$$\mathrm{sim}(\mathrm{u},\mathrm{v})^{[1]} = \frac{\mathrm{k}}{|\mathrm{uv}|} \tag{1}$$

$$|uv| = \sum_{i \in C} |u_i - v_i| \tag{2}$$

$$\mathrm{k} = \frac{1}{\sum_{\mathrm{v} \in \mathrm{U}} \frac{1}{|\mathrm{uv}|}} \tag{3}$$

In this study, besides sim(u, v)[1], we used Pearson correlation coefficient, cosine similarity and distance similarity. They are defined as equation (4), (5) and (6), where R_{u_mean} , R_{v_mean} denote the averages of all $R_{u,i}$ and all $R_{v,i}$ respectively; n denotes the number of items commonly rated by both users. Pearson correlation coefficient takes a range of values -1 to 1. The closer the value to 1 shows the more similarity between users. Cosine similarity takes a value between 0 and 1.

$$\mathrm{sim(u,v)_{Pearson}} = \frac{\sum_{i=1}^{n}\left(R_{u,i}-R_{u_{mean}}\right)\left(R_{v,i}-R_{v_{mean}}\right)}{\sqrt{\sum_{i=1}^{n}\left(R_{u,i}-R_{u_{mean}}\right)^2}\sqrt{\sum_{i=1}^{n}\left(R_{v,i}-R_{v_{mean}}\right)^2}} \tag{4}$$

$$\mathrm{sim(u,v)_{Cosine}} = \frac{\sum_{i=1}^{n}\left(R_{u,i}\right)\left(R_{v,i}\right)}{\sqrt{\sum_{i=1}^{n}\left(R_{u,i}\right)^2}\sqrt{\sum_{i=1}^{n}\left(R_{v,i}\right)^2}} \tag{5}$$

$$\mathrm{sim(u,v)_{Distance}} = \frac{1}{1+\sqrt{\sum_{i=1}^{n}\left(R_{u,i}-R_{v,i}\right)^2}} \tag{6}$$

3.2. NEIGHBORHOOD SELECTION

Thresholding and k nearest neighbors (KNN) are the most used neighborhood selection methods [9]. We applied KNN algorithm to choose most similar users to the active user. In order to choose similar users to the active user, we sort the users in the set of U according to similarity coefficients in ascending order. We choose $k = 5$, $k = 10$ and $k = max$. We mean that the number k is maximum, we have not limited the k to any number. It is the element number of in U set (all users who rate for missing item i and have common rated items with the active user).

3.3. ESTIMATION OF RATE

In order to estimate missing item's rate for a user, we first find the user set U which described in Section 3.1. Then, we calculate similarity coefficients $\mathrm{sim}(u,v)^{[1]}$ between the active user and these users. After calculating similarity coefficients, we estimate the missing value by using equation (7) in or preliminary work [1].

$$\mathrm{R_{u,i}}^{[1]} = \sum_{\mathrm{v} \in \mathrm{U}} \mathrm{R_{v,i}}\ \mathrm{sim(u,v)}^{[1]} \tag{7}$$

In this study, we estimate the rates with weighted average as in equation (8). In this calculation, the users' evaluation criteria are also considered [9].

$$\mathrm{R_{u,i}} = \mathrm{R_{u_{mean}}} + \frac{\sum_{\mathrm{v} \in \mathrm{U}} \left(\mathrm{R_{v,i}} \text{- } \mathrm{R_{v_{mean}}}\right) \mathrm{sim(u,v)}}{\sum_{\mathrm{v} \in \mathrm{U}} \mathrm{sim(u,v)}} \tag{8}$$

4. RESULTS AND DISCUSSION

In this paper, we use MovieLens [21] 100K Dataset. MovieLens data sets were collected at the University of Minnesota by the GroupLens Research Project. There are three sets of data with different number of ratings, i.e. 100K data set contains 100000 points for 1682 movies by 942 users; The 1M dataset includes grading from 3,0209 to 6040 users in 3900 movies; In the 10M data set, 71567 users have a rating of 100,00054 for 106,000 film in. All ratings in the three data sets range from 1 to 5. For the 100K data set, 106 votes are awarded per average user; each user will rate at least 20 movies and each movie is rated 59 times on average. For 1M data, the average vote order for each user and each movie is 166 and 256, 936 per movie in the 140 and 10M data sets per user, respectively. As the data size increases, more ratings are included, but, the densities are 0.063 for a 100K data set, 0.042 for a 1M data set, and 0.013 for a 10M data set. In this respect, 100K data is better than the other two.

We estimate the score of 1300 videos. Then we compare the estimated and the actual rates with the most popular evaluation metrics; Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) [13]. RMSE and MAE are calculated by equations (9) and (10), where N is the number of all data, $R_{u,i}$ denotes the real rating value and $R_{u,i'}$ denotes predicted rating value for item i by user u. MAE is the average absolute difference between the real and predicted ratings. RMSE is the square root of the average square of all errors. It amplifies the large absolute difference between the real and predicted ratings. We measure the estimation accuracy by using both of them. The lower MAE/RMSE means the predicted ratings are closer to the real rating values.

$$\text{RMSE}= \sqrt{\frac{1}{\text{N}}\sum_{\text{i}=1}^{\text{N}} (\text{R}_{\text{u,i}}\text{- R}_{\text{u,i}})^2} \tag{9}$$

$$\text{MAE}= \frac{1}{\text{N}}\sum_{\text{i}=1}^{\text{N}} |\ \text{R}_{\text{u,i}}\text{- R}_{\text{u,i}}\ | \tag{10}$$

We first compare rate estimation functions by using our similarity coefficient. In rate estimation, method 1 uses the equation (7) and method 2 uses the equation (8). In method 1, the active user has no contribution to rate estimation, only similar users are used for calculation. In method 2, besides the contribution of similar persons, the weight of the active user is added. Table 2 shows MAE and RMSE values of these methods. It is shown that weighted average calculation (method 2) gives better results. Therefore, we decided to work on with this rate calculation function.

Table 2. MAE/RMSE values when rate estimation function changes

	MAE			**RMSE**		
	KNN with K=5	KNN with K=10	KNN with K=Max	KNN with K=5	KNN with K=10	KNN with K=Max
Method1	0.8309	0.8086	0.8179	1.0769	1.0458	1.0473
Method2	0.7794	0.7723	**0.7618**	0.9937	0.9844	**0.9800**

After choosing the rate estimation function, we compare the similarity coefficients and effect of the choosing k similar users with KNN algorithm. Table 3 shows the results. It is seen that our similarity coefficient outperforms distance similarity and it gets closer performance to other coefficients. It can be concluded that choosing similar users to active user has positive effect as seen in Figure 1 and 2, because when the number of used similar user increases, the lowest MAE and RMSE are obtained. Figure 1 gives the MAE values of similarity coefficients with different number of nearest neighbor on the Movielens-100K data set.

Table 3. Comparison of similarity coefficients and effect of the choosing k similar users

	MAE			**RMSE**		
	KNN with K=5	KNN with K=10	KNN with K=Max	KNN with K=5	KNN with K=10	KNN with K=Max
Cosine	0.7882	0.7593	**0.7415**	0.9953	0.9607	**0.9486**
Pearson	0.7889	0.7581	0.7897	1.0023	0.9748	1.0285
Distance	0.8476	0.8148	0.7619	1.0875	1.0307	0.9683
Proposed	0.7794	0.7723	**0.**7618	0.9937	0.9844	0.9800

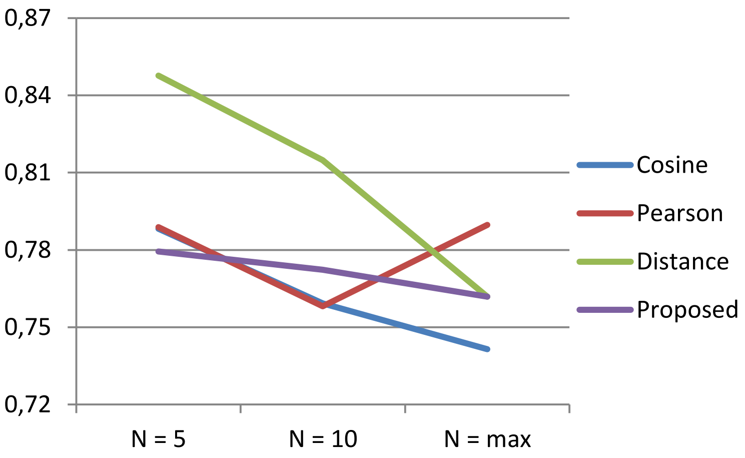


Figure 1. The MAE values of similarity coefficients with different K-neighbors

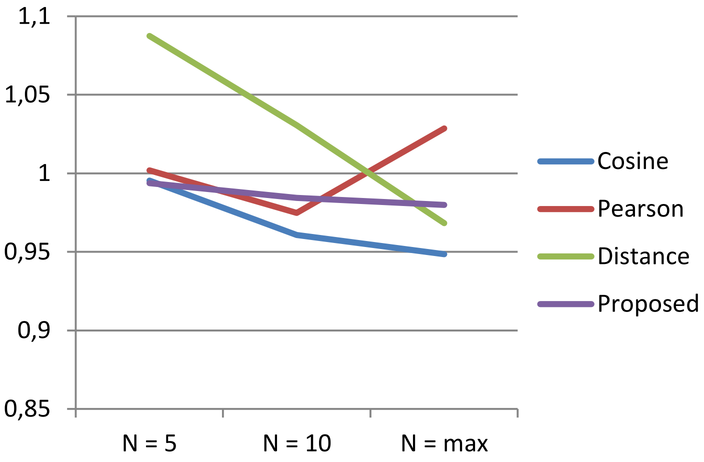


Figure 2. The RMSE values of similarity coefficients with different K-neighbors

5. CONCLUSION

In recommendation systems, collaborative filtering is one of the most widely used methods. In our preliminary work, we proposed a new similarity function to calculate similarities between users. In this study, to get better results, we extended our preliminary work. We included the user weights together with the active user weight. We used MAE and RMSE evaluation metrics to evaluate our study. We observed that our similarity coefficient outperforms distance similarity and it gets closer performance to other coefficients. We also observed that when the number of nearest neighbor increases, it gives better results.

REFERENCES

[1] Sincan, O.M., Yildirim, Z., “Video recommendation system using collaborative filtering”, International Conference on Advances in Science and Arts ICASA 2017, (2017).

[2] Goldberg, D., Nichols, D., Oki, B. M., Terry, D. “Using collaborative filtering to weave an information tapestry.” Communications of the ACM, 35/12 (1992) 61-70.

[3] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J. “GroupLens: an open architecture for collaborative filtering of netnews.” In Proceedings of the 1994 ACM conference on Computer supported cooperative work, (1994), p. 175-186.

[4] Breese, J. S., Heckerman, D., Kadie, C. “Empirical analysis of predictive algorithms for collaborative filtering.” In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, (1998), p. 43-52.

[5] Herlocker, J. L., Konstan, J. A., Riedl, J. “Explaining collaborative filtering recommendations.” In Proceedings of the 2000 ACM conference on Computer supported cooperative work, (2000), p. 241-250.

[6] Yu, K., Schwaighofer, A., Tresp, V., Xu, X., Kriegel, H. P. “Probabilistic memory-based collaborative filtering.” IEEE Transactions on Knowledge and Data Engineering, 16/1(2004) 56-69.

[7] Yang, J. M., Li, K. F. “Recommendation based on rational inferences in collaborative filtering.” Knowledge-Based Systems, 22/1 (2009) 105-114.

[8] Adamopoulos, P., Tuzhilin, A. “Recommendation opportunities: improving item prediction using weighted percentile methods in collaborative filtering systems.” In Proceedings of the 7th ACM conference on Recommender systems, (2013), p. 351-354).

[9] Bulut, H., Milli, M. “İşbirlikçi filtreleme için yeni tahminleme yöntemleri.” Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 22/2(2016) 123-128.

[10] Gogna, A., Majumdar, A. “A comprehensive recommender system model: Improving accuracy for both warm and cold start users.” IEEE Access, 3(2015) 2803-2813.

[11] Luo, X., Zhou, M., Xia, Y., Zhu, Q. “An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems.“ IEEE Transactions on Industrial Informatics, 10/2(2014) 1273-1284.

[12] Hernando, A., Bobadilla, J., Ortega, F. “A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model.” Knowledge-Based Systems, 97(2016), 188-202.

[13] Yang, X., Guo, Y., Liu, Y., Steck, H. “A survey of collaborative filtering based social recommender systems.” Computer Communications, 41(2014) 1-10.

[14] Yang, Z., Wu, B., Zheng, K., Wang, X., Lei, L. “A Survey of Collaborative Filtering-Based Recommender Systems for Mobile Internet Applications.” IEEE Access, 4(2016) 3273-3287.

[15] Mooney, R. J., Roy, L. “Content-based book recommending using learning for text categorization.“ In Proceedings of the fifth ACM conference on Digital libraries, (2000), p. 195-204.

[16] Elahi, M., Ricci, F., Rubens, N. “A survey of active learning in collaborative filtering recommender systems.” Computer Science Review, 20(2016) 29-50.

[17] Lops, P., De Gemmis, M., Semeraro, G. (2011). “Content-based recommender systems: State of the art and trends.” In Recommender systems handbook, (2011), p. 73-105

[18] Semeraro, G., Lops, P., Degemmis, M. “WordNet-based user profiles for neighborhood formation in hybrid recommender systems.” In Hybrid Intelligent Systems, 2005. HIS 05. Fifth International Conference on, (2005), p. 6-pp.

[19] Degemmis, M., Lops, P., Semeraro, G. “A content-collaborative recommender that exploits WordNet-based user profiles for neighborhood formation.“ User Modeling and User-Adapted Interaction, 17/3(2007) 217-255.

[20] Burke, R. “Hybrid recommender systems: Survey and experiments.” User modeling and user-adapted interaction, 12/4(2002) 331-370.

[21] Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A., Riedl, J. “MovieLens unplugged: experiences with an occasionally connected recommender system.” In Proceedings of the 8th international conference on Intelligent user interfaces, (2003), 263-266.

Current Address: Ozge MERCANOGLU SINCAN: Ankara University, Faculty of Engineering, Department of Computer Engineering, 06830, Gölbaşı, Ankara, TURKEY
E-mail Address:: *omercanoglu@.ankara.edu.tr*
ORCID: *https:// orcid.org/0000-0001-9131-0634*

Current Address: Zeynep YILDIRIM: Ankara University, Faculty of Engineering, Department of Computer Engineering, 06830, Gölbaşı, Ankara, TURKEY
E-mail Address: *yildirimz@.ankara.edu.tr*

Available online: December 21, 2017

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 59, Number 2, Pages 55-68 (2017)
DOI: 10.1501/commua1-2_0000000104
ISSN 1303-6009
http://communications.science.ankara.edu.tr/index.php?series=A2

PERSON IDENTIFICATION USING FUNCTIONAL NEAR- INFRARED SPECTROSCOPY SIGNALS USING A FULLY CONNECTED DEEP NEURAL NETWORK

OZGE MERCANOGLU SINCAN, HACER YALIM KELES, YAGMUR KIR,
ADNAN KUSMAN, and BORA BASKAK

Abstract. In this study, we investigate the suitability of functional near-infrared spectroscopy signals (fNIRS) for person identification using data visualization and machine learning algorithms. We first applied two linear dimension reduction algorithms: Principle Component Analysis (PCA) and Singular Value Decomposition (SVD) in order to reduce the dimensionality of the fNIRS data. We then inspected the clustering of samples in a 2d space using a nonlinear projection algorithm. We observed with the SVD projection that the data integrity associated with each person is high in the reduced space. In the light of these observations, we implemented a random forest algorithm as a baseline model and a fully connected deep neural network (FCDNN) as the primary model to identify person from their brain signals. We obtained %85.16 accuracy with our FCDNN model using SVD reduction. Our results are in parallel with the neuroscience researches, which state that brain signals of each person are unique and can be used to identify a person.

1. Introduction

fNIRS is a non-invasive optical imaging technique that is used to measure the blood flow in the brain. It measures oxyhemoglobin (OxyHb) and deoxyhemoglobin (Hb) concentrations in blood. To accomplish this, it sends 2 wavelengths (695 and 830 nm) of infrared light. This light is emitted and reflected back. According to Beer-Lambert law, OxyHb and Hb changes can be calculated by the amount of light absorbed in brain tissue.

fNIRS studies have contributed to making progress for understanding the human brain. In this method, neuronal activity is determined by measuring changes in

Received by the editors: October 18, 2017; Accepted: December 20, 2017.
Key word and phrases: functional near-infrared spectroscopy (fNIRS), person identification, PCA, SVD, fully connected deep neural network, random forest.

OxyHb and Hb concentrations. During a task, the brain activity increases, and blood flow also increases. When the activity stops, the blood flow is decreases and thus the concentration of OxyHb and Hb is decreases [1]. In recent years, fNIRS has gained a lot of attention because it is non-invasive, inexpensive, portable and easy to use. It is used in many diagnoses and researches including Alzheimer, schizophrenia, depression, epilepsy, Parkinsonism, dementia, brain computer interface (BCI), pain, emotion, sleep research etc. [2]. Because previous studies have shown that brain signals of every individual is unique [3], hence fNIRS data can be used for person identification [4,5,6].

In this study, we aim to identify people using their brain signals measured by fNIRS. We organize the rest of the paper as follows: In Section 2, we review the related work. In Section 3, we present the details of the proposed approach for person identification. In Section 4, we provide the performance of our approach. Finally, we conclude the paper in Section 5.

2. RELATED WORK

Ferrari and Quaresima [2] reviewed the fNIRS studies from the discovery of fNIRS (1992) until 2012. In this study, important events were detailed in chronological order. fNIRS is used in the main areas of psychiatry, neurology, psychology, education and BCI. It was reported that approximately 700 fNIRS units have been used worldwide for human brain cortex fNIRS studies on adults and infants.

In the field of psychiatry, schizophrenia is the most widely reported issue in fNIRS applications [7]. Koike et al. [7] reviewed the fNIRS studies in schizophrenia patients. Verbal fluency task (VFT) is a popular task in neuropsychological tests and neurological imaging measurements. Studies show that schizophrenic patients have worse performance on VFT than healthy people.

BCI is a system that controls computers or other external devices through brain activity. Naseer and Hong [8] reviewed fNIRS-based BCI studies between 2004 and 2014; the studies are evaluated in terms of tasks which are applied to subjects during measurements, the utilized noise removal methods, feature extraction methods and classification methods. The two most common brain areas in which brain signals are obtained are the primary motor cortex and the prefrontal cortex. Various experiments are performed regarding the motor cortex, including a motor execution and imagination of a motor execution; and regarding the prefrontal cortex, such as mental arithmetic, music imagery, emotion induction and object rotation [8]. Between 2004 and 2014 in fNIRS based BCI studies, the most common noise removal method was bandpass filtering; while mean, slope, variance, peak, skewness and kurtosis features are utilized widely. The most widely used classification methods in this domain are Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Hidden Markov Models (HMM), and Artificial Neural Networks (ANN).

Hiroyasu et al. [9] proposed a gender classification method using the human brain s blood flow change data that are measured by fNIRS. Firstly, numerical memory task was applied to 22 subjects during the measurements. Secondly, features were extracted from time series and they were labeled as male or female. Then, a learning model was constructed. At rest period, average classification of men and women was %62 while at memory task period it was %81. It can be deduced that at memory task period, blood flow is triggered by brain activities and there is a difference between men and women cerebral blood flow.

Heger et al. [4] investigated the suitability of fNIRS data for person identification. Although electroencephalogram (EEG) has been used for biometric identification [3,5], the authors report that fNIRS data has not been used for biometric identification before. In the study, mental tasks were applied two times at two different days to 5 subjects, because mental states can change over time. Best average identification accuracy (%80) was obtained when low frequency band and longer window were used.

McDonald and Solovey [6] aimed to identify subjects using only brain data. In Boyer et al. [10], the first 30 minutes of long time fNIRS data was used. Subjects were in resting state at the first 30 minutes. Multilayer perceptron was used as the classifier; they obtained 63% accuracy. Because the probability of random identification in 30 people is 3.3%, %63 is a significant rate. This shows that even though the brain is at rest, each individual may have its unique brain signal.
Deep learning, popular in recent years, is a machine learning method used to train artificial neural networks. It is a promising approach because it does not require expert knowledge and feature extraction. After successful results in areas such as object recognition, natural language processing, and voice classification, deep learning has also been used in the classification of fNIRS data [1, 11-14].

3. MATERIALS AND METHODS

3.1. Data Acquisition

Cerebral blood flow changes were measured by the 52-channel fNIRS device (ETG-4000; HitachiMedicalCo., Tokyo, Japan) located at the Brain Research and Applications Center of the Ankara University. 32 healthy subjects were recruited for the experiment. The subjects sat in front of a computer screen, which displayed the experimental tasks.

Reading the Mind From the Eyes Test [15] was adopted for the fNIRS environment by the Matlab Psychophysiology Toolbox software as the activation paradigm (Figure 1). The neuroimaging task consisted of two conditions. During the ToM condition the subjects were expected to guess the correct mental state presented in the eye photographs (A blocks). The subjects were allowed to respond to as many photographs as they could / during the 30 s intervals. There was no predetermined time limit per photograph. During the control condition the participants were presented the same eye photographs as the ToM condition, but, they were expected to guess whether the eyes presented belong to a man or a woman during 30 s (B blocks). Again, they were allowed to respond to as many photographs as they could during the 30 s intervals of the control task. Both the ToM and the control conditions were presented in four blocks.

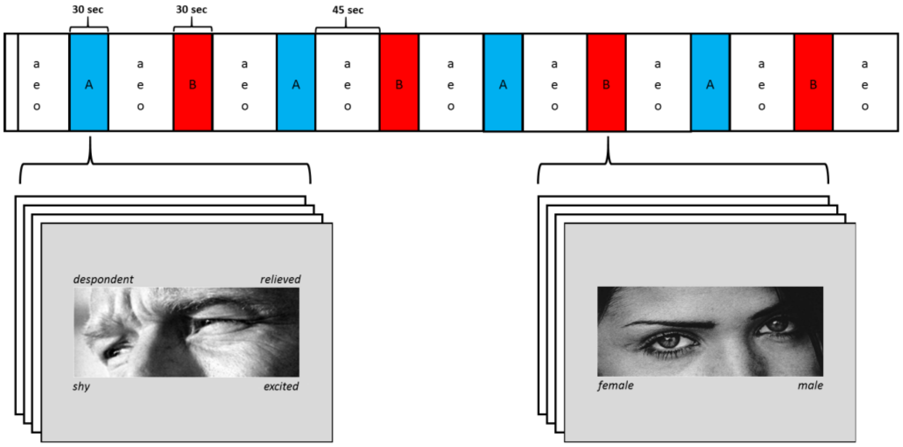


Figure 1. Reading the Mind from the Eyes Task as the Cortical Activation Paradigm

The two task conditions, A and B were presented consequently and were preceded and followed by 45 s rest periods. Since the participants responded verbally to the task, they were stipulated to repeat Turkish vowels (/a/,/e/,/o/) during those rest periods in order to control the effect of articulation.

3.2. Dimension Reduction

The signal set that are produced with the fNIRS device data acquisition has a non-stationary character that makes them hard to analyze in high dimensions due to high redundancy. Since, we only have a limited amount of data for training a machine learning algorithm, we need to reduce the number of dimensions to a space where the variation of the data is preserved sufficiently in the reduced space. For this purpose, we applied two linear dimension reduction algorithms: Principle Component Analysis (PCA) and Singular Value Decomposition (SVD). Although they are both computing the eigenvalues of the data distributions, SVD better handles sparse data distribution since it works directly using the data matrix, while PCA uses the covariance matrix of the data. In addition to reducing the data dimension using one of these methods, we want to verify if the reduction preserves the data integrity in the reduced space. For this purpose, we utilized t-Distributed Stochastic Neighbor Embedding (tSNE) [16] to visualize our dataset in the reduced dimension.

We analyzed the effectiveness of the PCA and SVD dimension reduction techniques on our dataset using tSNE algorithm; in our experiments, we used the Python *sklearn.manifold* package implementation. We observed that SVD reduction provides better separation between different subjects. This is visible in Figure 2.

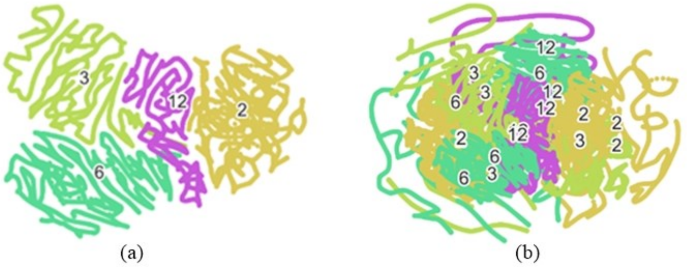


FIGURE 2. tSNE visualization of 4 persons (with person-ids: 2,3,6,12) after (a) SVD reduction (b) PCA reduction to 5-dimensional space.

In Figure 2, a non-linear projection of the data set on a 2-dimensional space is depicted. In Figure 2-(a), SVD projection of the four subjects in the original dataset to 5-dimensional space is re-projected to a 2-dimensional space; while in Figure 2-(b), PCA projection of the same dataset is depicted. Each person in the dataset is given an identification number. After the projections, the spatial position of each subject's data projection is depicted with a different color and the corresponding person-id is placed at the centroid of the projected space. When the data cluster of a subject is split by another cluster, the person-id may appear in multiple spatial locations, i.e. Figure 2-(b). The tSNE algorithm finds a manifold in the reduced dimensional space that has similar distribution with the original data distribution in the high dimensional space. Considering this, Figure 2 gives a good intuition about the data distribution in the SVD reduced space and the PCA reduced space, assuming that optimum manifolds are found by the tSNE algorithm. These visualizations depict that SVD provides more clear boundaries than the PCA reduction between the clusters of person data samples. We observed the similar phenomenon for the reductions into 10 and 20 dimensional spaces. The integrity of the 5-dimensional space representation of the data is still high, hence we decided to work on this very reduced dimensional space to identify person from its fNIRS signals.

To validate the success of the projections with the SVD method, we extended the visualizations of more data samples from the dataset by including 6 randomly selected subjects, 8 randomly selected subjects and 32 subjects, which corresponds to all the samples in the dataset; the corresponding visualizations are depicted in Figure 3-(b), (c) and (d), respectively. The resultant visualizations show that we can utilize the projected data for person identification using a classification technique that can find non-linear class boundaries. We selected two such classification methods for this purpose. The details of our classification approaches are provided in the next section.

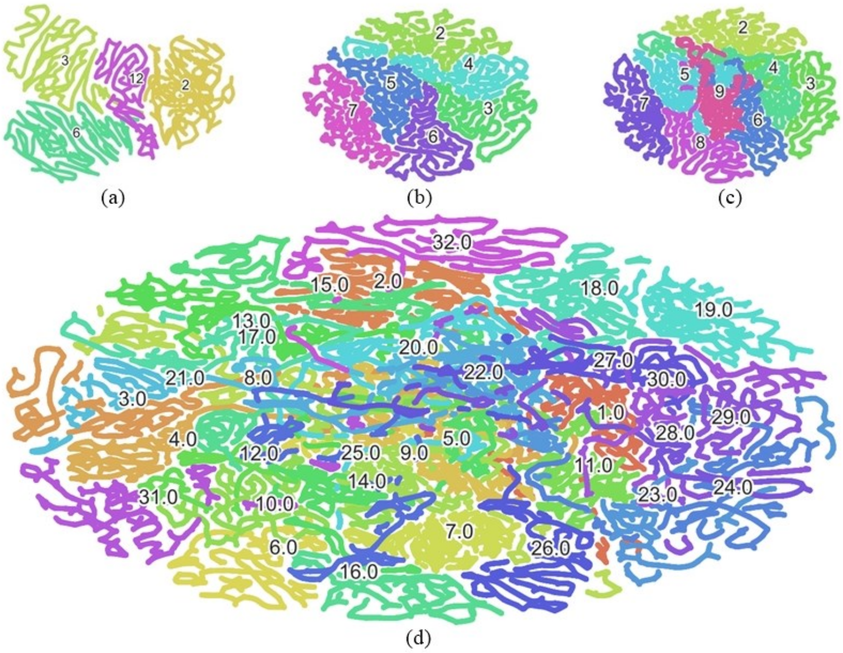


FIGURE 3. tSNE visualization of the 5-dimensional space of our fNIRS dataset. SVD is used for dimension reduction. The samples are selected randomly in the upper row; the number of different subjects in (a) is 4, (b) is 6, (c) is 8 and (d) is 32.

3.3. Classification

Our aim in this study is to identify subjects using their fNIRS signals that are gathered during the whole session of different tasks. Since we have a limited number of subjects and limited number of data samples, we first reduced the data samples approximately 10%, from 52-dimensional space to a 5-dimensional space to reduce *the curse of dimensionality* problem. In this space, we trained two different classifiers to identify subjects from their samples: (1) a fully connected deep neural network, (b) an ensemble of 25 decision trees, using random forest algorithm. We will provide the details of each classifier in this section.

3.3.1. Fully connected deep neural network (FCDNN)

In the machine learning literature, a neural network that has more than one hidden layer is called a deep neural network. We designed a FCDNN with 3 hidden layers, one input and one output layer. It is a feed forward, fully connected network; each layer is connected to all the neurons in the next layer and no connections exist among the units in the same layer. The architecture of our FCDNN is depicted in Figure 4. The architecture parameters are determined after extensive number of experiments using a subset of the dataset.

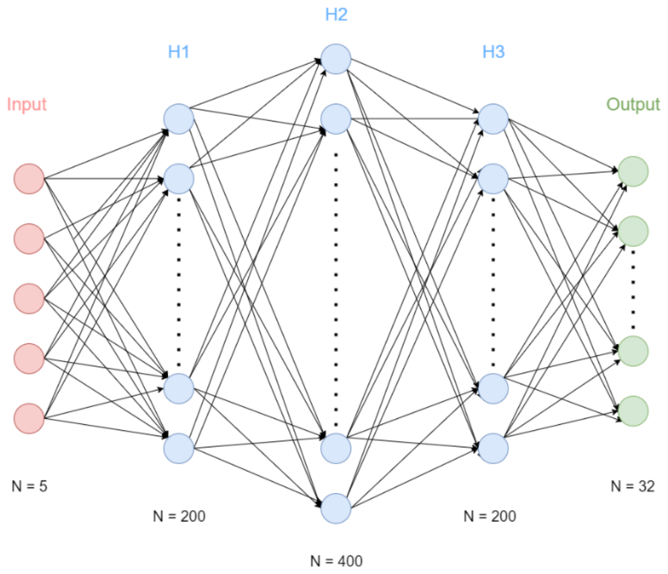


Figure 4. The architecture of our FCDNN.

As it is shown in Figure 4, input layer has 5 units, the first hidden layer (H1) has 200 units, the second hidden layer has 400 units (H2), the third hidden layer (H3) has 200 units and the output layer has 32 units. As the number of output layers (32) tell, the network is designed as a multi-class classifier; each output unit corresponds to a subject category in our dataset. Each unit in the hidden layers is implemented with a rectified linear unit (ReLU) activation function. The scores that are generated at the last layer is used by a softmax function to approximate the log likelihoods of the classes for each given training sample. During training, we used categorical cross-entropy loss for computing the gradients in the back-propagation algorithm; we used the Adam optimizer [17], which is one of the most effective optimization algorithms that is used to reduce the loss function in this domain.

In order to reduce the overfitting to the training data we applied dropout regularization after each hidden layer with 50% probability. Moreover, we also included L_2 weight regularization in some of our experiments. The details of our experiments are provided in Section 4.

3.3.2. Random Forest Model (RF)

The second classifier that we trained for our person identification task is a Random Forest model. It is based on an algorithm that fits a decision tree classifier on some subsets of the given training samples, and then uses the ensemble of these decision trees' scores to generate an average score for each category in classification. RF model is an easily trained ensemble method that effectively eliminates overfitting by averaging the scores of many decision trees. This model is selected as the baseline model to assess the performance of the FCDNN model. In our implementation we used Python (*sklearn.ensemble package*) implementation of the algorithm, which is publicly available and well documented. Interested readers may refer to the detailed documentation of the algorithm, which is considered beyond the scope of this paper. During the training of the RF model, we set the number of decision trees to a range of values while hyperparameter tuning, i.e. 25, 100, 200. We did not observe much difference between the classification accuracies with the validation dataset, hence we set it to 25 for computational efficiency in the final model generation that is used during testing. The test results are discussed in the following section.

4. RESULTS AND DISCUSSION

We performed extensive experiments with the FCCNN and RF model to assess the performance of two data dimension reduction methods, namely the PCA and SVD, by reducing the dimension by more than 10% in the person identification task.

The test data is prepared as follows: for each subjects all the samples in a given time-sequence data is split by the first 75% as the training samples and the remaining 25% as the test samples. Considering that in each part of this time-varying data, a different task is processed by the subjects, such a division can be considered as a random division when person identification problem is considered. The idea behind this data splitting is that the last 25% of the fNIRS signals belong to a different time and different sub-task, hence are not related directly to a part of the training samples, hence can be considered as a valid test set in this problem. It is important to note that if the samples are randomly selected from the whole time-varying stream, overfitting to the training samples would yield almost the same high accuracy with the test set. Since in this case the samples belong to the very same pattern with the training data, hence is not suitable for assessing the model performances.

Utilizing the above experiment setting in the data training and testing, we trained an RF model as our baseline model to better assess the performance of our FCDNN model. We did not make extensive experiments with this model, yet only changed the number of estimators in the RF algorithm, i.e. 25, 100, 200. As we stated in the previous section, we did not observe a significant improvement as we increased the number of samples with validation data, we set it to 25 and obtained the baseline accuracies that are shown in Table 1. The results show that with a standard RF classifier, we can identify the subjects by more than 70% even if we reduce the data dimension by 10%. In RF experiments, SVD reduction performed slightly better than the PCA reduction method, i.e. 0.14%.

Table 1. Person identification accuracies with Random Forest algorithm

Reduction Technique	Test Accuracy (%) # estimators: 25
SVD	73.19
PCA	73.05

Using the above-mentioned training and test datasets, we also trained our FCDNN model. The training and test accuracies of different experiment settings are summarized in Table 2. The results show that the FCDNN model performs better by more than 10% when we compare it with the baseline (RF) model. Moreover, SVD reduction method has better generalization capability than the PCA reduction, although the accuracies of the two methods are only slightly different.

At the beginning, we first overfit to the training data to observe the capacity of the network. When it is sufficiently high, where with both reduction methods we get more than 99%, we concluded that this model architecture is suitable in our problem setting. The test accuracy when we the model overfits to the data is better in PCA than SVD, i.e. 84.14%. Then we applied dropout regularization after all hidden layers with 50% dropout probability and L_2 weight regularization in the loss function with different weights. We observe in the third row of Table 2 that SVD based model accuracy increases by 1.56%, while PCA based model accuracy increases by only 0.22%. In this setting the weight decay parameter is 0.02.

Table 2. Person identification accuracies when SVD and PCA dimension reduction techniques are used

	SVD (d = 5)		**PCA (d = 5)**	
	Train Accuracy (%)	Test Accuracy (%)	Train Accuracy (%)	Test Accuracy (%)
No regularization	99.39	83.60	99.98	84.14
Dropout L2:0.01	98.85	84.37	98.82	84.28
Dropout L2:0.02	98.85	**85.16**	98.91	**84.36**
Dropout L2:0.03	98.90	83.97	98.83	84.30

5. CONCLUSION AND FUTURE WORKS

fNIRS is an optical brain imaging technique used for investigation of the brain functions. Being cheap compared to other brain imaging devices, being portable and non-invasive make fNIRS prevalent.

In this study, we analyzed two data dimension reduction techniques, namely the PCA and SVD, using a very successful data visualization approach, i.e. tSNE. Our observations show that fNIRS data distribution is highly coherent; we observe isolated clusters of samples for distinct subjects. The clustering is slightly better when SVD is used. We observed that when we reduce the dimension to 5, we can only keep the 49.07% of total variance of the data and yet still could preserve the coherency in the tSNE projection. Hence, decided to work on this *very* reduced dimensional space.

We implemented a baseline model (RF) and a FCDNN model to classify fNIRS data samples in the reduced dimensional space for person identification. We obtained around 73% accuracy with the baseline (RF) model and improved that result with our FCDNN architecture to around 85%. These results show that we can identify subjects with an acceptable accuracy using a simple fully connected neural network architecture. Although the data has a time-varying character, we did not model the change of the samples with time in our solution. We evaluate each sample in an isolated manner.

We also observe that although the test accuracies of both reduction methods (SVD and PCA) are very close to each-other, SVD displays better generalization capability than the PCA in our experiments. We also observe better clustering in tSNE projections with SVD. This idea needs further exploration by increasing the test datasets that are gathered from the same subjects, which is scheduled as a future work in this research. As a future work, we also aim to analyze the time-characteristics of the data for each person using a time sequence model.

REFERENCES

[1] Huve, G., Takahashi, K., Hashimoto, M., “Brain activity recognition with a wearable fNIRS using neural networks”, In Mechatronics and Automation (ICMA), (2017), 1573-1578.

[2] Ferrari, M., Quaresima, V., “A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application”, Neuroimage, 63/2 (2012), 921-935.

[3] Marcel, S., Millán, J. D. R., “Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation”, IEEE transactions on pattern analysis and machine intelligence, 29/4 (2007).

[4] Heger, D., Herff, C., Putze, F., Schultz, T., “Towards biometric person identification using fNIRS”, In Proceedings of the Fifth International Brain-Computer Interface Meeting: Defining the Future, (2013).

[5] Campisi, P., La Rocca, D., “Brain waves for automatic biometric-based user recognition”, IEEE transactions on information forensics and security, 9/5 (2014), 782-800.

[6] McDonald, D. Q., Solovey, E., “User identification from fNIRS data using deep learning”, In The First Biannual Neuroadaptive Technology Conference, (2017) 156.

[7] Koike, S., Nishimura, Y., Takizawa, R., Yahata, N., Kasai, K., “Near-infrared spectroscopy in schizophrenia: a possible biomarker for predicting clinical outcome and treatment response”, Frontiers in psychiatry, 4 (2013).

[8] Naseer, N., Hong, K. S., “fNIRS-based brain-computer interfaces: a review”, Frontiers in human neuroscience, 9 (2015).

[9] Hiroyasu, T., Hanawa, K., Yamamoto, U., “Gender classification of subjects from cerebral blood flow changes using deep learning”, In Computational Intelligence and Data Mining (CIDM), IEEE Symposium, (2014), 229-233.

[10] Boyer, M., Cummings, M. L., Spence, L. B., Solovey, E. T., "Investigating mental workload changes in a long duration supervisory control task", Interacting with Computers, 27/5 (2015), 512-520.

[11] Hennrich, J., Herff, C., Heger, D., Schultz, T., "Investigating deep learning for fNIRS based BCI", In Engineering in Medicine and Biology Society (EMBC), 37th Annual International Conference of the IEEE, (2015), 2844-2847.

[12] Hiwa, S., Hanawa, K., Tamura, R., Hachisuka, K., Hiroyasu, T., "Analyzing brain functions by subject classification of functional near-infrared spectroscopy data using convolutional neural networks analysis", Computational intelligence and neuroscience, (2016), 3.

[13] Pham, T. T., Nguyen, T. D., Van Vo, T., "Sparse fNIRS feature estimation via unsupervised learning for mental workload classification", In Advances in Neural Networks, (2016), 283-292.

[14] Trakoolwilaiwan, T., Behboodi, B., Choi, J. W., "Convolutional neural network for functional near-infrared spectroscopy in brain-computer interface", (2017), 423-424.

[15] Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I., "The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism", The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42/2 (2001), 241-251.

[16] L.J.P. van der Maaten, Hinton, G. E., "Visualizing high-dimensional data using t-SNE", Journal of Machine Learning Research 9 (2008), 2579-2605.

[17] Kinga, D., Adam, J. B., "A method for stochastic optimization", International Conference on Learning Representations (ICLR), (2015).

Current Address: Ozge MERCANOGLU SINCAN: Ankara University, Faculty of Engineering, Department of Computer Engineering, 06830, Gölbaşı, Ankara, TURKEY
E-mail Address: omercanoglu@.ankara.edu.tr
ORCID: *https:// orcid.org/0000-0001-9131-0634*

Current Address: Hacer YALIM KELES: Ankara University, Faculty of Engineering, Department of Computer Engineering, 06830, Gölbaşı, Ankara, TURKEY
E-mail Address: *hkeles@.ankara.edu.tr*

Current Address: Yagmur KIR: Ankara University, School of Medicine, Department of Psychiatry, Ankara University Brain Research Center, TR06590, Dikimevi, Ankara, Turkey

E-mail Address: *yagmurucan@windowslive.com*

Current Address: Adnan KUSMAN: Ankara University, School of Medicine,
Department of Psychiatry, Ankara University Brain Research Center, TR06590,
Dikimevi, Ankara, Turkey
E-mail Address: *kusman@ankara.edu.tr*

Current Address: Bora BASKAK: Ankara University, School of Medicine,
Department of Psychiatry, Ankara University Brain Research Center, TR06590,
Dikimevi, Ankara, Turkey
E-mail Address: baskak@medicine.ankara.edu.tr

Available online: December 27, 2017

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 59, Number 2, Pages 69-79(2017)
DOI: 10.1501/commua1-2_0000000105
ISSN 1303-6009
http://communications.science.ankara.edu.tr/index.php?series=A2

COMMUNICATIONS
SERIES A2-A3

SENTIMENT ANALYSIS USING A RANDOM FOREST CLASSIFIER ON TURKISH WEB COMMENTS

NERGIS PERVAN and HACER YALIM KELEŞ

Abstract. Sentiment analysis is an active research area since early 2000s as a field of text classification. Most of the studies in this field focus on the analysis using the text in English language, where the Turkish and the other languages have fallen behind. The purpose of this research is to contribute to the text analysis in Turkish language using the contents that we access through web sites. In particular, we deduce the sentiment behind noisy product reviews and comments in a highly popular commercial web page. In this context, we generate a unique dataset that includes 9100 product review samples for training our classification model. There are different word representation methods that are utilized in sentiment analysis, such as bag-of-words and n-gram models. In this work, we generated our word models using the word2vec algorithm. In this model, each word in the vocabulary is represented as a vector of 300 dimensions. We utilize 70% of our dataset in the training of a Random Forest Model and make binary classification of sentiments as being positive or negative, utilizing the ratings of the user for the product as classification labels. In the highly noisy and unfiltered comments, we achieve an accuracy of 84.23%.

1. Introduction

Sentiment analysis has emerged as an important topic with the increase of social media interactions, use of forums and blogs, sales comments and ratings through e-commerce websites. Sentiment analysis is a field of Natural Language Processing (NLP), which is also referred to as, opinion mining, sentiment classification, opinion extraction, etc. in the literature. The research in this field has started in the early 2000s with the works of [1], [1], [3], [4], [5] and [6]. The use of the term *sentiment analysis* first appeared in [7].

Sentiment analysis is a way to determine the writer's opinions polarity as positive or negative in a piece of text, about a particular topic, and product etc. The field applications contributed to many tasks on the evaluation of consumer products, understanding the impacts of some social events, and evaluating movie reviews. In

Received by the editors: November 11, 2017; Accepted: December 27, 2017.
Key word and phrases: Sentiment Analysis, Word2Vec, Random Forest, Turkish Web Comments.

addition to these, there are even studies that make stock market predictions using some sentiment states on tweets [8].

Feature extraction on texts is a challenging problem; the challenge is on converting the characters, words, sentences or documents to computational units which is useful for sentiment classification. There are recent studies based using deep learning algorithms based on character level text representations [9] [10]. Different word-level algorithms that are effective representations of words for feature extraction on texts are proposed, such as *word2vec* [11], *glove* [12] and *fasttext* [13]. The word vectors have semantic and syntactic meaning of words. Using the pre-trained word vector models are beneficial in terms of time in classifying sentences.

In this work, we first collected customer reviews including misspelling words and meaningless character combinations from e-commerce websites. We used the entire dataset for word embedding and a part of the dataset for classification. The obtained word representation model was used for classifying Turkish reviews as positive or negative with a Random Forest (RF) classifier.

The paper is organized as follows: In Section 2, we explain the materials and the proposed method, in Section 3, we describe our experiments and discuss the results. We conclude the paper in Section 4.

2. MATERIALS AND METHODS

The proposed model is outlined in Figure 1. The first step is data generation that is main contribution to this domain because of lack of Turkish labeled data set. Data preprocessing is the second step is required for noisy refinement to obtain both word embedding vector and training samples. The last step is training classification model using training samples are obtained with the transformation of labeled data set conjunction with word embedding to review vectors.

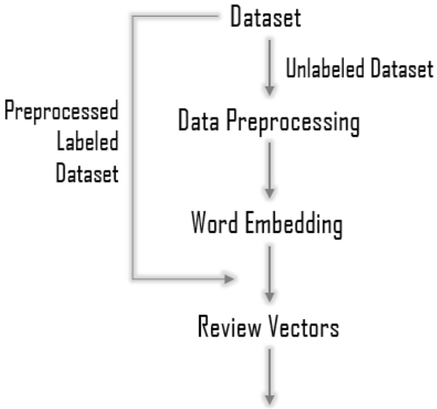


Figure 1. Flow of the method

Dataset Preparation

We collected the dataset from a Turkish e-commerce web site, which includes costumer product reviews in electronics category. This category has many sub-categories such as computers, cell phones, TV, video games, etc. Collected comments are not in ideal shape all the time; they usually contain some misspelling words and meaningless word structures, therefore consist of highly noisy sentences. Some sample sentences from our original dataset collection are shown in Table 1.

The dataset includes a total of 93922 consumer reviews. For supervised model generation, we need a label for each review that represents the sentiment behind the review. We automatically determined labels by using the information encoded with the star ratings that is provided by the user together with each review. For example, if a customer likes a product very much, he/she gives a rating using five or four stars. We associate the labels of the comments with their star ratings. There are up to five star categories, i.e. from one to five, for each comment but in training, we used only certain sentiments in the reviews that have either one star, i.e. a negative sentiment, or five star, i.e. a positive sentiment. We discarded in between ratings in our dataset. We approximately balanced the number of positive and negative reviews in the dataset.

TABLE 1. Sample review sentences from our original dataset. * indicates correct spelling of reviews.

Reviews
*bu telefon **mutish** kendime aldim oneririm* * bu telefon müthiş kendime aldım öneririm
***Tlfnumdan** cok memnunum kaldim teşekkür ler* *Telefonumdan çok memnun kaldım teşekkürler
***tlf** çok güzel **herkeze** tavsiye ederim.bu paralara alınabilecek en iyi **tlfn**..* * Telefon çok güzel herkese tavsiye ederim bu paralara alınabilecek en iyi telefon.
*Geçikmeli kargo ama üründe sıkıntı yok yine**de tşkr*** * Gecikmeli kargo ama üründe sıkıntı yok yine de teşekkürler
***hoparlor eywallah ta** basa **glince** bu ne ya çok kötü **bi** bas sistemi son sese açınca berbat beklentimin altında **bi** ürün beğenmedim* * hoparlör eyvallah da basa gelince bu ne ya çok kötü bir bas sistemi son sese açınca berbat beklentimin altında bir ürün beğenmedim

Data Preprocessing

Data preprocessing is an important and necessary step before training models and requires different attention and care in different languages, especially for Turkish language. In opinion mining problems, we need to remove uninformative characters, words, phrases, etc., hence redundancy, from the dataset. The data refinement helps reducing the dimensionality and noise in the data in terms of word representation; hence helps easy generation of word embedding and training classification models.

Since we collected the dataset from the web, each review contains html tags. We started data preprocessing by removing these tags from the data. We also removed all auxiliary characters except for the Turkish letters. In some cases the characters may express the sentiment behind the review when the combinations of characters correspond to some emoticons, like smiley, laughing, crying, etc. To unify the word representations with a simple format, we converted all the remaining letters to lowercase representation. This helps us interpreting the content in a case-insensitive way. We also apply word reduction, i.e. the word lists in a review contains a set of unique words. Tokenization is the task of splitting text based on a character (blank character, punctuation) to tokens that is a piece of character sequences. In preprocessing, we removed punctuations before tokenization, and then we generated each sentence to list of words. In addition, lemmatization is a commonly used process in data preparation; yet for Turkish language lemmatization is not necessary because it is a morphologically rich language. Turkish language consist of suffixes which changes the meaning of a word negatively, hence we did not use any lemmatization. Moreover, a recent study shows that lemmatization is not necessary for English language sentiment analysis neither [14].

Word Embedding

There are a few different algorithms for word embedding. We used word2vec representation that is proposed by Google researchers in 2013. Word2vec model training is fast and efficient. The model generates a feature vector representation for each word in a vector space. There are two ways to obtain the model; first is using a pre-trained model which is generated by Google using 100 billion words in English language. These features have 300 dimensions and in total there are 30 million unique words. The second way is training a model from scratch using your own dataset in a language you want. Since we want to generate Turkish word models, we trained word2vec model using all the data in our Turkish dataset. We used *Gensim (Generate Similar)* library of Python to implement vector-space modelling and topic modelling. The word2vec model takes sentences, which are the reviews for the obtained data set, as the input and learns word vector representations. We set some parameters for model implementation as follows. The word2vec algorithm supplies continuous bag-of-words (CBOW) and skip gram architecture for producing word vector representations. In *Gensim* default architecture is CBOW and we used it; we generated the dimensionality of the feature vectors for each learned unique words to *300*, set the maximum distance between target word and words around the target word within a sentence to *10* and ignored all words with an occurrence less than *40* times.

Classification

For classification, we first need to generate training samples in a suitable form. We used 9100 training and 3700 test samples for total of 12800 samples. Words in all samples are represented by our word2vec model with a 300-dimensional vector corresponding to that word. Each review becomes a two dimensional tensor, i.e. $N_i \times 300$, here N_i is the number of words in review *i*. We need to create a fixed size vector representation for each review for training, yet the number of words in each review, $N_{i,}$ are different, For this purpose, we computed the averages of the vectors for each review to represent as a point in the feature space of 300 dimensions. The subset of the attained vectors that we prepared for training with labels as negative and positive sentiments are utilized by a RF classifier in training. The generated model is then used in testing to classify a given, i.e. unseen, Turkish sentiment as a positive or negative sentiment.

3. EXPERIMENTAL RESULTS

This section shows the results of our word2vec model and RF model.

Training word2vec Model Results

First, we used 65525 unlabeled comments in Turkish to initialize word embeddings. After we tuned parameters required for *Gensim* tool, as a result of unsupervised learning, we attained 4328-word vector representations. When we upgrade the number of word samples to 93922 and regenerate the word vector representations, a total of 5711 unique word vectors are generated in about 7 seconds. When we use these word models, our RF classification accuracy also increases by 1.59%.

In Figure 2, we show the produced word vectors in a 2d space. Although individual words are not readable in this figure, we can see some word clusters. When we zoom in, the clusters contain similar words in terms of semantic in Figure 2 and syntactic in Figure 3 (The content is best read in the digital form). In Figure 3 the conjugations of word 'yapmak' ('do' in English) are close to each other such as 'yapmam'('I do not'), 'yapmayı' ('doing'), 'yapacağım' ('I will do'), 'yaparım' ('I do'), 'yapan' ('who do'), 'yapamayacağım' ('I will not be able to). Although they seem far away on the graph, the values in the two dimensions are very close to each other.

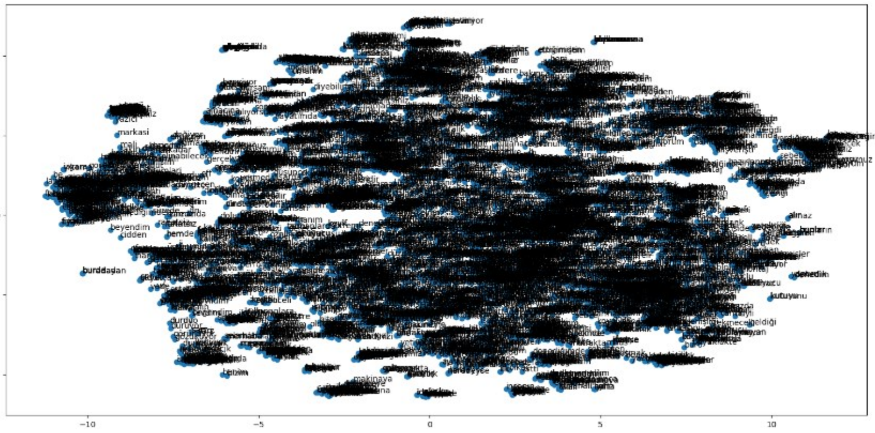


Figure 2. Distribution of the Turkish word2vec model words in two-dimensional vector space.

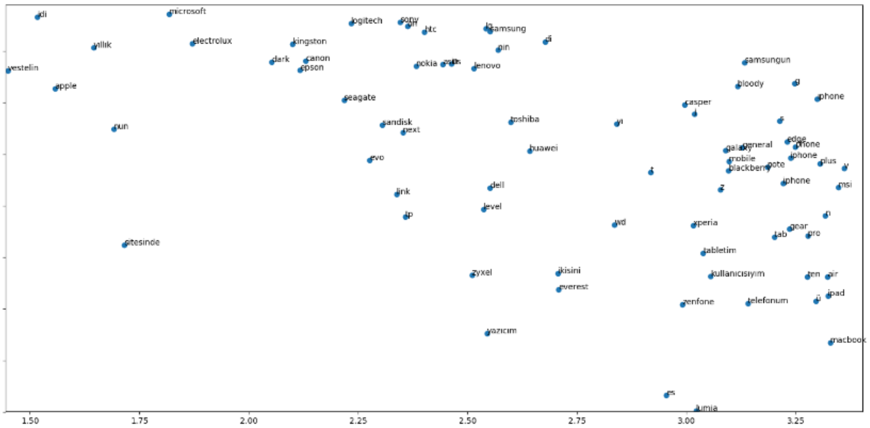


Figure 3. Projections of words that have similar semantics.

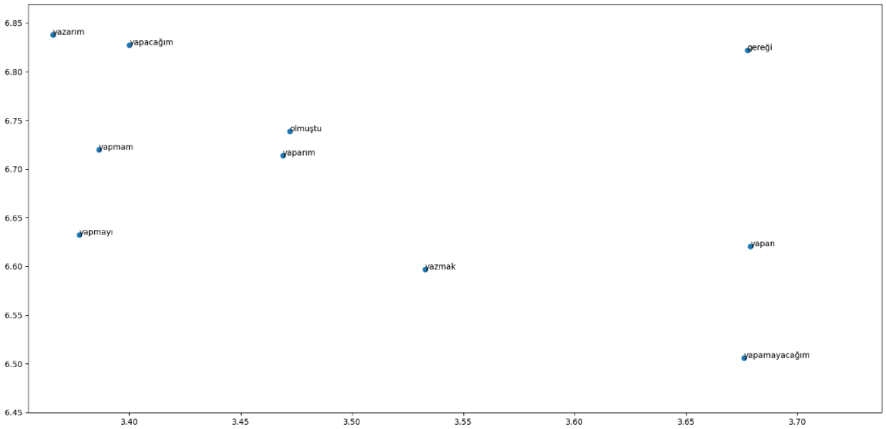


FIGURE 4. Projections of words that have similar syntax.

One of the problems of the word embedding model is that some opposite semantic words are found to be similar to each other in their word vector representations. There are recent studies that aims to project not only semantic and syntactic but also sentiment content of text before creating a model [15], [16]. [17] emphasizes the same problem with an approach distinctly using existing word embedding model.

Classification Results

We used 200 decision trees in the training of our random forest classifier. We tested our model with 3100 test samples and obtained 84.23% accuracy.

The sentiment of the text in the obtained dataset is semantically noisy. In Table 2 we show some semantically conflicting samples from the test set. Despite the positive opinion it implies hidden a negative meaning, the review is ranked as negative, and our model classified it correctly as negative. Similarly, in the second example, the word ‘inanılmaz’(‘incredible’) is misleading, i.e. have a positive meaning, and resulted in an incorrect estimation. In third review, the beginning of the sentence has positive sentiment and the rest of the sentence goes on negative; yet in this case our model could infer the true meaning. The last sample is misclassified although it clearly contains a positive sentiment. This sentence has an implied semantic meaning that exaggerates the usefulness of the product by saying it very implicitly, i.e. telling that the previous products are not products at all. So, RF model is having difficulty to mine the implied meaning behind the sentiments.

Table 2. Test samples.

Reviews	Target	Prediction
program yüklenince ram doluyor fazla program kurulmazsa ideal (ram is full when program is loaded, ideal if no more programs are installed)	Negative	Negative
İnanılmaz gürültülü bir yazıcı. (An incredibly noisy printer.)	Negative	Positive
İşinizi görecek bir alet fakat mouse la normal mouse gibi iş yapamıyoruz alışık deyilim açıkçası (a tool that serves your needs but with this mouse we can not work like (we do with) a normal mouse, actually I am not used to.)	Positive	Positive
Bunu kullandıktan sonra önceden kullandıklarım ne idi diye insan düşünmekten kendini alamıyor kalite mükemmel (After using this, people can not get away from thinking what I used before quality is perfect)	Positive	Negative

4. Conclusion

In this paper, our purpose is to contribute to Turkish language sentiment analysis using a machine learning approach. First, we collected a unique dataset, which was a primary challenge in Turkish language analysis, since there were no publicly available dataset. We applied a set of preprocessing techniques to create a useful training and test samples. We generated word vector representations using our dataset with an unsupervised learning algorithm, i.e. word2vec model generation algorithm. These representations are used in training an RF model for Turkish sentiment classification. We achieved 84.23% classification accuracy with our test samples.

The obtained word vector models and the dataset will be used in our future researches and will be made publicly available. Our future direction is training a sequence based deep learning model that can reveal complex and hidden semantic meanings behind the sentences.

REFERENCES

[1] Wiebe, J. "Learning Subjective Adjectives from Corpora", *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth*

Conference on Innovative Applications of Artificial Intelligence, July 30- August 03 (2000): 735-740.

[2] Das, S.R. and Chen, M. Y. 2001. “Yahoo! for Amazon: Extracting Market Sentiment from Stock Message Boards”. *In Proceedings of the 8th Asia Pacific Finance Association Annual Conference,* (2001).

[3] Morinaga, S., Yamanishi, K., Tateishi, K. and Fukushima, T. “Mining Product Reputations on the Web”. *In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* (2002).

[4] Tong, R. M. “An Operational System for Detecting and Tracking Opinions in On-Line Discussion”. *In Proceedings of SIGIR Workshop on Operational Text Classification,* (2001).

[5] Pang, B., Lee, L. and Vaithyanathan. S. “Thumbs up? Sentiment Classification Using Machine Learning Techniques”. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP),* (2002): 79–86.

[6] Turney, P. 2002, “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews”. *In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics,* (2002): 417–424.

[7] Nasukawa, T. and Yi, Jeonghee. “Sentiment analysis: Capturing Favorability Using Natural Language Processing”. *In Proceedings of the KCAP-03, 2nd Intl. Conf. on Knowledge Capture,* (2003).

[8] Bollen, J., Mao, H. and Zeng, X. 2010. “Twitter Mood Predicts the Stock Market”. *Journal of Computational Science*, (2010): 2(1), 1–8.

[9] Kim, Y., Jernite, Y., Sontag, D. and Rush, A. “Character-Aware Neural Language Models”. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16),* (2016).

[10] Zhang, X., Zhao, J. and LeCun, Y. “Character-level Convolutional Networks for Text Classification*”. In Proceedings of NIPS,* (2015).

[11] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. “Distributed Representations of Words and Phrases and their Compositionality”. *In Proceedings of NIPS,* (2013).

[12] Pennington, J., Socher, R., and Manning, C. D. 2014. “Glove: Global Vectors for Word Representation”. *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP),* (2014): 12.

[13] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T. “Enriching Word Vectors with Subword Information”. *arXiv preprint,* (2016): 1607.04606.

[14] Camacho-Collados, J. and Pilehvar, M.T. “On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis”. *arXiv preprint,* (2017): *1707.01780*

[15] Lan, M., Zhang, Z., Lu, Y., and Wu, J. 2016. “Three Convolutional Neural Network-Based Models for Learning Sentiment Word Vectors towards Sentiment Analysis”. *In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN-16),* (2016): 3172-3179.

[16] Tang, D., Wei, F., Qin, B., Yang, N., Liu, T., and Zhou, M. 2016. “Sentiment Embeddings with Applications to Sentiment Analysis”. *IEEE Trans. Knowl. Data Eng.*, (2015): 28 (2), 496-509.

[17] Yu, L.-C., Wang J., Lai, K. R. and Zhang X. “Refining Word Embeddings for Sentiment Analysis”. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (2017): 545-550.

Current Address: Nergis PERVAN: Department of Computer Engineering, Ankara University, Ankara 06830, TURKEY
E-mail Address: *nergispervan@gmail.com*
ORCID: *https://orcid.org/0000-0003-3241-6812*

Current Address: Hacer YALIM KELEŞ : Department of Computer Engineering, Ankara University, Ankara 06830, TURKEY
E-mail Address: *hkeles@ankara.edu.tr*
ORCID: *https://orcid.org/0000-0002-1671-4126*

INSTRUCTIONS TO CONTRIBUTORS

Communications Faculty of Sciences University of Ankara Series A2-A3: Physical Sciences and Engineering is a peer reviewed journal which has been published since 1948 by Ankara University, accepts original research articles written in English in the fields of Physics, Engineering Physics, Electronics/Computer Engineering, Astronomy and Geophysics. Review articles written by eminent scientists can also be invited by the Editor.

Manuscripts should be submitted as a single PDF file attached to an e-mail with a covering letter. In the covering letter, authors should nominate three potential reviewers and e-mailed the file to the most appropriate Area Editor of the research. The editorial office may not use these nominations, but this may help to speed up the selection of appropriate reviewers.

Manuscripts should be typeset using the LATEX typesetting system. Authors should prepare the article using the COMMUNICATIONS style before submission by e-mail. Manuscripts written in DOC form are also acceptable. A template of manuscript can be reviewed in **http://communications.science.ankara.edu.tr/index.php?series=A2A3&link=300**. After the acceptance of manuscripts for publication, we will ask you to submit the **TeX** form of the manuscript prepared in accordance with the style of the Journal. Authors are required to submit their Open Researcher and Contributor ID (**ORCID**) 's which can be obtained from http://orcid.org as their URL address in the format http://orcid.org/xxxx-xxxx-xxxx-xxxx. Acknowledgements should be given as short as possible at the end of the text. Formulas should be numbered consecutively in parentheses ( ). Footnotes should be avoided if possible, but when necessary, should be short and never contain any important part of the work and should be numbered consecutively by superscripts. All illustrations not including tables (photographs and other films, drawings, graphs, etc) must be labeled as "Figure". The proper position of each table and figure must be clearly indicated in the paper.

All tables and figures must have a number (Table 1, Figure 1) and a caption or legend. References including comments must be numbered consecutively in order of first appearance in the text. The reference number should be put in brackets [] where referred to in the text. References should be listed at the end of the manuscript in the numbered order in which they appear in the text as follows:

[1] Bairamov, E, Ozalp N., Uniform convergence and numerical computation of the Hubbell radiation rectangular source integral, *Radiation Physics and Chemistry*, 80 (2011) 1312–1315.

[2] Kelley, J. L., General Topology, Van Nostrand, 1970.

It is a fundamental condition that articles submitted to COMMUNICATIONS have not been previously published and will not be simultaneously submitted or published elsewhere. After the manuscript has been accepted for publication, the author will not be permitted to make any new additions to the manuscript.

Before publication, the galley proof is sent to the author for correction. Thus, it is solely the author's responsibility for any typographical mistakes occur in their article as it appears in the Journal. The contents of the manuscript published in the COMMUNICATIONS are the sole responsibility of the authors.

The PDF copies of accepted papers are free of charges, but hard copies of the paper, if required, are due to be charged for the amount of which is determined by the administration each year.

Editor in Chief
Commun. Fac. Sci. Univ. Ank. Ser. A2-A3.
Ankara University, Faculty of Sciences
06100, Besevler - ANKARA TURKEY

COMMUNICATIONS

FACULTY OF SCIENCES
UNIVERSITY OF ANKARA

DE LA FACULTE DES SCIENCES
DE L'UNIVERSITE D'ANKARA

Series A2-A3: Physical Sciences and Engineering

Volume 59 **Number : 2** **Year :2017**