



ICONDATA

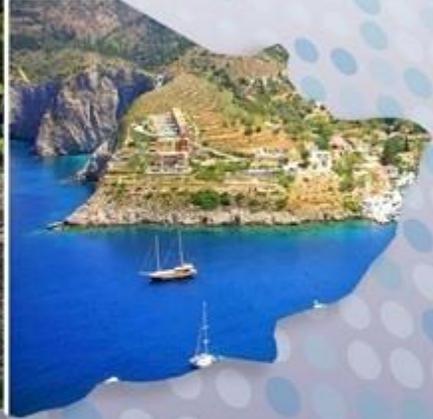
INTERNATIONAL CONFERENCE ON DATA SCIENCE & APPLICATIONS

October 3-6, 2019
Balikesir - TURKEY

PROCEEDINGS
BOOK

www.icondata.org

e-ISBN
978-605-031-662-9



KAZ DAĞLARI
EDREMIT, BALIKESİR



Editors

Murat GÖK
Emre DANDIL
Hüseyin GÜRÜLER

Volume 2
Abstract Book



GEMBA akademi
Eğitim ve Danışmanlık Hizmetleri

zenginveri

PREFACE

In the current information age, data is the basis of all intelligent systems. From the agricultural society to the industrial society, from there to the information society, the transformation process is moving towards Industry 4.0. The ability to store and process large amounts of data on computers has led to an increase in the capabilities of the products and services produced. The statistical and artificial learning studies based on meaning deriving from data have paved the way for intelligent systems in all sciences.

2st International Conference on Data Science and Applications (ICONDATA'19) has been organized in Balıkesir, TURKEY on October 3-6, 2019.

The main objective of ICONDATA'19 is to present the latest data-based researches from all disciplines. This conference provides opportunities for the different areas delegates to exchange new ideas and application experiences face to face, to cooperate between different disciplines from both natural and social sciences and to find global partners for future collaboration.

All paper submissions have been double blind and peer reviewed and evaluated based on originality, technical and/or research content/depth, correctness, relevance to conference, contributions, and readability. Each paper proposal was evaluated by two reviewers.

20 selected papers presented in the conference that match with the topics of the journals will be published in the following journals:

- Data Science and Applications,
- Veri Bilimi Dergisi,

In particular we would like to thank Prof. Dr. Suat CEBECİ, Rector of Yalova University, Prof. Dr. İbrahim Taş, Rector of Bilecik Şeyh Edebali University and Prof. Dr. Hüseyin Çiçek, Rector of Muğla Sıtkı Koçman University.

Looking forward to see you in ICONDATA 2020,
Murat GÖK, Emre DANDIL and Hüseyin GÜRÜLER
Editors

COMMITTEES

Honorable and Advisory Board

Prof. Dr. Suat Cebeci	Rector of Yalova University
Prof. Dr. İbrahim Taş	Rector of Bilecik Şeyh Edebali University
Prof. Dr. Hüseyin Çiçek	Rector of Muğla Sıtkı Koçman University

Conference Chairman

Assoc. Prof. Dr. Murat Gök	Yalova University, Turkey
----------------------------	---------------------------

Organizing Board

Assoc. Prof. Dr. Hüseyin Gürüler	Muğla Sıtkı Koçman University, Turkey
Assist. Prof. Dr. Emre Dandil	Bilecik Şeyh Edebali University, Turkey
Hasibe Candan	Yalova University, Turkey
Aykut Durgut	Balıkesir University, Turkey
Vongani Maluluke	University of Cape Town, Republic of South Africa
Banu Sayar	Yalova University, Turkey

Science Board

Prof. Dr. Abderrahmane Bouda	National Maritime Superior Institute, Algeria
Prof. Dr. Abdullah Uz Tansel	Baruch College, New York, USA
Prof. Dr. Ahmet Cumhuri Kınacı	Çanakkale Onsekiz Mart University, Turkey
Prof. Dr. Ahmet Sabri Öğütlü	Harran University, Turkey
Prof. Dr. Ayhan İstanbullu	Balıkesir University, Turkey
Prof. Dr. Ayşe Çetin	Yalova University, Turkey
Prof. Dr. Dursun Aydın	Muğla Sıtkı Koçman University, Turkey
Prof. Dr. Latif Taşkaya	Muğla Sıtkı Koçman University, Turkey
Prof. Dr. Ecir Uğur Küçükşille	Süleyman Demirel University, Turkey
Prof. Dr. Erhan Akyazı	Marmara University, Turkey
Prof. Dr. Erman Coşkun	Bakırçay University, Turkey
Prof. Dr. Gufran Ahmad Ansari	Crescent Institute of Science & Technology, India
Prof. Dr. Hacer Gümüş	Kocaeli University, Turkey
Prof. Dr. Kaka Shahedi	Sari Agricultural Sciences and Natural Resources University, Iran
Prof. Dr. Kamel Harchouche	University of Sciences and Technology of Houari Boumediene, Algeria
Prof. Dr. Mustafa Öztaş	Yalova University, Turkey

Prof. Dr. Müfit Çetin	Yalova University, Turkey
Prof. Dr. Nour El Islam Bachari	University of Sciences and Technology of Houari Boumédiene, Algeria
Prof. Dr. Ramazan Bayındır	Gazi University, Turkey
Prof. Dr. Sinan Uğuz	Süleyman Demirel University, Turkey
Prof. Dr. Sinan Şen	Yalova University, Turkey
Prof. Dr. Şakir Taşdemir	Selçuk University, Turkey
Prof. Dr. Tamer Kahveci	University of Florida, Florida, USA
Prof. Dr. Vilda Purutçuoğlu	Middle East Teknik University, Turkey
Assoc. Prof. Dr. Abdülkadir Tepecik	Yalova University, Turkey
Assoc. Prof. Dr. Akın özçift	Celal Bayar University, Turkey
Assoc. Prof. Dr. Deniz Kılınç	Celal Bayar University, Turkey
Assoc. Prof. Dr. Eray Can	Yalova University, Turkey
Assoc. Prof. Dr. Eyüp Akçetin	Muğla Sıtkı Koçman University, Turkey
Assoc. Prof. Dr. Hasan Erdiñ Koçer	Selçuk University, Turkey
Assoc. Prof. Dr. M. Gökhan Genel	Yalova University, Turkey
Assoc. Prof. Dr. Nevin Güler Dinçer	Muğla Sıtkı Koçman University, Turkey
Assoc. Prof. Dr. Murat Yabanlı	Muğla Sıtkı Koçman University, Turkey
Assoc. Prof. Dr. Orhan Kesemen	Karadeniz Teknik University, Turkey
Assoc. Prof. Dr. Sunay Türkdogan	Yalova University, Turkey
Assoc. Prof. Dr. Ali Hakan Işık	Mehmet Akif Ersoy University, Turkey
Assoc. Prof. Dr. Ufuk Bal	Muğla Sıtkı Koçman University, Turkey
Assoc. Prof. Dr. Ümit Ünver	Yalova University, Turkey
Assist. Prof. Dr. Abit Balin	Istanbul University, Turkey
Assist. Prof. Adem Tuncer	Yalova University, Turkey
Assist. Prof. Dr. Alper Kürşat Uysal	Eskişehir Teknik University
Assist. Prof. Dr. Aytaç Pekmezci	Muğla Sıtkı Koçman University, Turkey
Assist. Prof. Dr. Enis Karaaslan	Muğla Sıtkı Koçman University, Turkey
Assist. Prof. Dr. Beste Hamiye Beyaztas	Bartın University, Turkey
Assist. Prof. Dr. Burcu Okkaloğlu	Yalova University, Turkey
Assist. Prof. Dr. Duygu Yıldırım Peksen	Yalova University, Turkey
Assist. Prof. Dr. Eyüp Çalık	Yalova University, Turkey
Assist. Prof. Dr. Fahriye Zemheri Navruz	Bartın University, Turkey
Assist. Prof. Dr. Gözde Mert	Nişantaşı University, Turkey
Assist. Prof. Dr. Gül Yücel	Yalova University, Turkey
Assist. Prof. Dr. Güncel Sarıma	Muğla Sıtkı Koçman University, Turkey

Assist. Prof. Dr. Güneş Harman	Yalova University, Turkey
Assist. Prof. Dr. Gürcan Çetin	Muğla Sıtkı Koçman University, Turkey
Assist. Prof. Dr. Halit Karalar	Muğla Sıtkı Koçman University, Turkey
Assist. Prof. Dr. Hülya Akat	Muğla Sıtkı Koçman University, Turkey
Assist. Prof. Dr. İsmail Kırbaş	Mehmet Akif Ersoy University, Turkey
Assist. Prof. Dr. M.Bahar Başkır	Bartın University, Turkey
Assist. Prof. Dr. Melih Ağraz	Middle East Teknik University, Turkey
Assist. Prof. Dr. Muhammed Kürşad Uçar	Sakarya University, Turkey
Assist. Prof. Dr. Murat Canayaz	Van Yüzüncü Yıl University, Turkey
Assist. Prof. Dr. Murat Okkaloğlu	Yalova University, Turkey
Assist. Prof. Dr. Naveed Islam	Islamia College University, Pakistan
Assist. Prof. Dr. Osman H. Koçal	Yalova University, Turkey
Assist. Prof. Dr. Ali İmran Vaizoğulları	Muğla Sıtkı Koçman University, Turkey
Assist. Prof. Dr. Sait Ali Uymaz	Selçuk University, Turkey
Assist. Prof. Dr. Serdar Neslihanoglu	Eskişehir Osmangazi University, Turkey
Assist. Prof. Dr. Serdar Birogul	Düzce University, Turkey
Assist. Prof. Dr. Süleyman Uzun	Bilecik Şeyh Edebali University, Turkey
Assist. Prof. Dr. Şebnem Özdemir	Beykent University, Turkey
Assist. Prof. Dr. Ufuk Beyaztas	Bartın University, Turkey
Assist. Prof. Dr. Ufuk Çelik	Bandırma Onyediy Eylül University, Turkey
Assist. Prof. Dr. Ayşegül Akaydın	Nisantasi University, Turkey
Assist. Prof. Dr. Dilek Bayrak	Namik Kemal University, Turkey
Assist. Prof. Dr. Şerife Yılmaz	Karadeniz Technical University, Turkey
Lecturer Ali Osman Selvi	Bilecik Şeyh Edebali University, Turkey
Assist. Dr. İnanç Kabasakal	Ege University, Turkey
Dr. Guillermo Santamaria-Bonfil	Universidad Nacional Autónoma de México, Mexico
Dr. İrfan Kösesoy	Yalova University, Turkey
Dr. Murat Zorba	Çanakkale Onsekiz Mart University, Turkey
Dr. Tülay Öncü Öner	Manisa Celal Bayar University, Turkey
Dr. S. Balamurugan	Quants IS and Consultancy Services, India
Dr. Şengül Can	Manisa Celal Bayar Üniversitesi, Turkey
Dr. Vivek Kumar	National University of Science and Technology, Russia
Dr. Yunus Özen	Yalova University, Turkey

CONTENTS

Forecasting Sulfur Dioxide (SO ₂) Using Cointegration and Vector Error Correction Model: A Case Study for State of Kuwait	1
Predicting Sulfur Dioxide (SO ₂) Using Co-Integration and Vector Error Correction Model: A Case Study from State of Kuwait	2
Variability and Decadal Evolution of Temperature and Salinity in The Mediterranean Sea Surface	3
Intrinsic Attenuation Structure in Shallow Fault Zone in West Anatolia Region (Edremit Bay)	4
Detection of Environmental Changes Using Saliency Maps.....	5
Statistical and Biometric Revision of Morphometric Characteristics in Fish of The Genus <i>Diplodus</i> (<i>Sparidae</i>) of The Algerian Coast.....	6
Developing the Text Summarization API with Natural Language Processing Techniques	7
Identifying Natural Road Accidents Factors Using Spatial Big Data Mining Techniques	9
Visualization of Geographic Information System (GIS) Data with R Programming Language	10
Determination of Heavy Metal Levels in Liver and Muscle Tissues of Carp (<i>Cyprinus carpio</i> L., 1758) from Atatürk Dam Lake.....	11
Molecular Cloning and Functional Characterisation of an Ω Desaturase from The Common Ragworm <i>Hediste (Nereis) Diversicolor</i> (Müller, 1776)	12
Visualization and Animation of Temporal Turkey Data.....	13
Portfolio Risk Estimation via ARMA-GARCH Kapula Approach.....	14
Comparing Statistical Surveillance Methods for the Detection of Infectious Disease Outbreaks	15
The Calculation of the Ship Hull Biofouling Surface, As a Source of the Introduction of Invasive Species at the Port of Arzew	16
NetSim: A Tool for Network Simulation	17
Predicting Stock Prices Using Social Media Data	18
Investigation of Criteria for The Best Architecture Selection of Artificial Neural Networks.....	19
An RFM-Based Customer Segmentation Model for Targeted Offers in Mobile Gaming Industry	20
Regime Switching Fuzzy Time Series Model	21
Investigation of Ethnic and Cultural Differences in International Business in the Context of Generations	22
Classification of Eye Blinks in EEG Signals	23
Clustering of Biometric Data	24
A Hybrid Feature Selection Algorithm to Enhance Turkish News Classification.....	25
Estimation Modelling of Ames Housing Price with Machine Learning Algorithms	26
Comparison of CPU And GPU Based Acceleration Techniques in Machine Learning	27
ThesisManager: An Editor for Easy Formatting of The Graduate Thesis	28
Diabetic Retinopathy Detection Using Computer Vision	29
Triggering Diversity of Artificial Intelligence Based Art Research by using Turkish-Ottoman Art Genre	30
VGG16-CNN Ethnic Origin Classification with Deep Learning Model	31
Turkey Printing Industry of Several Variables in Terms of Future Projections	32
Real Time Estimation of Data from Color Sensor	33

Feature Extraction of P300 Signals Using Spectral Entropy Method	34
An Expert System of Price Forecasting for Second Hand Cars Using Regression Models	35
Classification of Eye Blinks in EEG Signals Using Neural Networks	36
The Estimation of Regression Coefficients in Composite Regression Models Using Heuristic Optimization Algorithms	37
The Investigation of Factors Affecting Level of Happiness Using Data Mining Methods	38
Very High Resolution Remote Sensing for The Diachronic Study of Beaches in The Central Algerian Region	39
A Change Detection Framework: Satellite Image Application	40
Contribution of Synthetic Cloudiness Images in The Classification of Meteorological Phenomena Using Msg Images	41
A Modified Type-2 Fuzzy Clustering Algorithm Using Atanassov's Intuitionistic Fuzzy Sets.....	42
Sismik Dalganın Yer İçindeki Yayılımının Zaman ve Frekans Ortamında Modellenmesi-Modeling Seismic Wave Propagation in Time and Frequency Domain	43
Fake Account Detection on Twitter Using Machine Learning Algorithms	44
An Investigation Into the Noise Filtering Performance of Quadratic Image Filters	45
R Programming Language Process Mining Package: BUPAR.....	46
A Fashion Image Product Recommendation System using Deep Learning	47
Estimation of Censored Regression in case of Skew and Bimodal Errors.....	48
An Optimization Energy Consumption of Sensor Nodes	49
Using Machine Learning Techniques to Predict the Type of Asthmatic Allergy	50
Generalized Fuzzy Rough Approximation Spaces Formed by a New Fuzzy Set Valued Function on Canonical Hypergroups.....	51
A Method Based on Natural Language Processing for Identify and Correct Errors in Text Data	52
DCC-GARCH-Copula Approach in Modelling Dependency between Stock and Government Bonds.....	53
Classification of Human Activities Using Deep Learning.....	54
Application of Fiber Bragg Grating Sensors in Helicopter Flight Test Instrumentation	55
Comparison of Methods Cox Regression and Random Forests in Survival Analysis.....	56
Supplier Selection Using Fuzzy Analytic Hierarchy Process Method: Application of an Automotive Supply Industry Company	57
Investigation of the Efficiency of Solid Waste Management in Level 1 Regional Municipalities with Data Envelopment Analysis	58
Development of Image Filter Based on Quadratic Filter for Enhancement of Brain MR Scans	59
Computer Aided Analysis of Brain MRI Images	60

Forecasting Sulfur Dioxide (SO₂) Using Cointegration and Vector Error Correction Model: A Case Study for State of Kuwait

Ahmad ALSABER¹, Pan JIAZHU¹

¹ *Strathclyde University, Glasgow, UNITED KINGDOM*

Abstract

The research is present a comparative performance of time series modelling and forecasting for the air quality in the state of Kuwait using several approaches (e.g. uni-variate time series and multivariate time series). Using the Air-quality row data (daily observations) gathered from Kuwait Environment Public Authority (K-EPA). The sample data covers from 2012 till 2017, and number of monitoring stations are four different measurement stations. A uni-variate time series analysis will performed individually for each station using the ARIMA model. On the other hand, a multivariate time series analysis will be performed based on the analysis of co-integration rank test to build the Vector Error Correction Model (VECM) and Vector Autoregression (VAR) also will be considered. Co-integration allows the identification of environmental integrated time series that exhibit similar dynamics in the long run and the estimation of their relationships, by exploiting the stationary linear combinations of these time series. For the quality of the results, K-Means with Simultaneous will be performed to deal with outliers. Finally, based on the comparison of each estimated time series model, we expect better results for Vector Error Correction Model (VECM) and Vector Autoregression (VAR) rather than the other univariate time series model. SO₂, NO₂, O₃ and Pm₁₀ were selected for modeling multivariate time-series analysis to predicate the future concentration and to predicte the probability of having pollutant chances according to Kuwait EPA standards. Those pollutants are significantly associated with health effect; and nowadays many literature studied uni-variate time-series to predict the future concentration.

Keywords: *Multivariate Time Series Model, Co-Integration, VEMC, Outliers, Missing Imputation, Air Pollution, Kuwait*

Predicting Sulfur Dioxide (SO₂) Using Co-Integration and Vector Error Correction Model: A Case Study from State of Kuwait

Ahmad ALSABER¹

¹ *Strathclyde University, Glasgow, UNITED KINGDOM*

Abstract

This research paper will present a comparative performance of time series modelling and forecasting for the air quality in the state of Kuwait using several approaches (e.g. univariate time series and multivariate time series). Using the Air-quality row data (daily observations) gathered from Environment Public Authority (EPA) in the state of Kuwait that covers the period from 2012 till 2017 based on four different measurement stations, a univariate time series analysis was performed individually for each station using the ARIMA model. On the other hand, a multivariate time series analysis was performed based on the co-integration rank test to build the Vector Error Correction Model (VECM) and Vector Autoregression (VAR). For the quality of the results, K-Means with Simultaneous was performed to deal with outliers. The missing observations have been treated by Multivariate Imputation by Chained Equations (MICE). Finally, based on the comparison of each estimated time series model, we find the results of Vector Error Correction Model (VECM) and Vector Autoregression (VAR) are better than the other univariate time series model.

Keywords: *Multivariate Time Series, Co-Integration, VEMC, Outliers, Air Pollution, Kuwait*

Variability and Decadal Evolution of Temperature and Salinity in The Mediterranean Sea Surface

Lamri NACEF¹, Nour El Islam BACHARI¹, Abderrahmane BOUDA², Imene BELGAID¹

¹*Houari Boumediene University of Science and Technology "USTHB", Algiers, ALGERIA*

²*National Maritime Superior Institute, Algiers, ALGERIA*

Abstract

Analysis of spatial and temporal variations of temperature and salinity at Mediterranean Sea surface, as well as, the search of possible trends in both parameters are the main goals of this work. Data used in this work are those of the Med-Atlas 2000 database. The used statistical techniques allowed us to obtain various climatological fields of temperature and salinity, on a period of 45 years (1955-1999) at the Mediterranean Sea surface. Spatial and temporal analysis of those fields shows that the north-south gradient is weaker than the east-west gradient. The strongest variability in both mean fields is sharper in downwelling areas than anywhere else, showing the colder and less saline surface waters. Warmer and saltiest water surface are located in southeast of the Levantine basin. The eastern Mediterranean Sea is generally more saline than the western basin, by about 1. The temperature seasonal cycle is more marked than the salinity seasonal cycle. The summer-winter thermal and saline fields are completely contrasted, especially in the northern Adriatic Sea. The annual SST anomalies chronology shows that the largest positive peak of inter-annual SST variability is encountered in 1994 and the largest negative peak in 1992. Whereas those related to salinity observed in 1983 and in 1997, respectively. The decadal averages from the 1960s to the 1990s indicate that there is a cooling of Mediterranean Sea surface in the 1970s and a northward warming since the 1980s that accelerated in the 1990s. The eastern Mediterranean Sea exhibits a higher warming rate as compared to the western basin, but the average increase in sea surface temperature is about 0.2 °C/decade. The Salinity rising trend corresponds to the cooling periods and the decreasing trend is associated with the warming ones.

Keywords: *Mediterranean Sea Surface, Temperature, Salinity, Spatio-Temporal Variability, Decadal Variations*

Intrinsic Attenuation Structure in Shallow Fault Zone in West Anatolia Region (Edremit Bay)

Bedri SEVINÇ¹

¹ Mugla Sıtkı Kocman University, Mugla, TURKEY

Abstract

Earthquake records in the fault zones are the most useful data in seismic quality factor (Q) research. In this study, we have investigated the causes of signal energy loss in the fault region from reflection data using an artificial source. We have defined the characteristics of the fault structures in shallow sediments at a few km depths in the Gulf of Edremit by calculating the absorption values of seismic signals. In the study airgun has been used as seismic source, and a single channelled zero offset data has been acquired along a marine line in Edremit bay of west Anatolia. The airgun source was fired 2789 times at intervals of 2 seconds. In the scope of this research effective Q value of fault and adjacent zones have been reviewed. The effective-Q values were obtained from reflected seismic data. Spectral ratio (SR) and optimization methods (OM) were used to estimate seismic Q values. Q values are the differences in the lithology leading to the change and fractures and cracks in the layer and the accumulation of liquid and gas in these cracks. In the Edremit Gulf, Quality factor values were analyzed and internal changes in the fault regions were investigated. The mean $Q_p = 135$, $Q_p = 83.5$ in the fault region and $Q_p = 159.6$ after the fault region were obtained before the F1 fault region. $Q_p = 159.6$ before the F2 fault zone, $Q_p = 90$ at the fault zone, and $Q_p = 119.7$ after the fault zone. The low Q values show that the fault area is full of liquid and gas in fractures and cracks.

Keywords: *Intrinsic Attenuation, Sseismic Quality Factor, Spectral Ratio Method, Optimization Method*

Batı Anadolu'da Edremit Körfezi'nde Sığ Fay Bölgelerindeki Sismik Veri Zayıflamasının Araştırılması

Özet

İçsel zayıflamayı tanımlayan çoğu sismik kalite faktörü (Q) araştırmalarında veri kaynağı, fay hattında olmuş deprem kayıtlarıdır. Bu araştırmada yapay bir kaynak kullanılarak Fay bölgesinde sismik yansıma sinyalindeki enerji kaybının nedenlerini araştırdık. Edremit Körfezinde birkaç km derinliğe kadar sığ çökellerde sismik sinyaldeki enerji emilimi hesaplanarak fay yapısının özelliklerini tanımlamaya çalıştık. Araştırmada sismik kaynak olarak hava tabancası kullanıldı. Edremit Körfezinde deniz üzerinde bir hat boyunca, sıfır ofset tek kanallı veri toplandı. 2sn aralıklarla ile patlayan hava tabancası ile 2789 adet atış kaydı alındı. Araştırma kapsamında fay bölgesi ve yakın bölgedeki efektif Q değerleri gözden geçirildi. Sismik Q değerleri tahmininde spektral oran (SR) ve optimizasyon yöntemleri (OM) kullanıldı. Yer içi homojen bir yapıda olmadığı için tahmin edilen Q değerleri çok farklı olabilir. Q değerleri değişimine yol açan litolojideki farklılık ve tabaka içindeki kırık ve çatlaklar ile bu çatlaklardaki sıvı ve gaz birikimidir. Edremit Körfezi'nde Q kalite faktörü değerlerinin analizi ile fay bölgelerindeki içsel değişim incelenmiştir. F1 fay bölgesi öncesinde P dalgası kalite faktörü ortalaması (Q_p) = 135, fay bölgesinde $Q_p = 83.5$, fay bölgesinden sonra $Q_p = 159.6$ değerleri elde edildi. F2 fay bölgesinden önce $Q_p = 159.6$, fay bölgesinde $Q_p = 90$, fay bölgesinden sonra ise $Q_p = 119.7$ değerleri hesaplandı. Q değerlerinin düşük olması bize incelenen fay bölgesinde kırık ve çatlaklar içinde sıvı ve gaz dolu olduğunu göstermektedir.

Anahtar Kelimeler: *İçsel Zayıflama, Sismik Kalite Faktörü, Spektral Oran Yöntemi, Optimizasyon Yöntemi*

Detection of Environmental Changes Using Saliency Maps

Burak ÇİÇEK¹, Mehmet DİKMEN¹

¹*Baskent University, Ankara, TURKEY*

Abstract

Change detection has been widely studied to evaluate many issues such as urban growth, assesment of natural disasters urban and regional planning. Moreover, it is useful for updating maps of agricultural lands and organising associated resources. The input data used in common techniques used for change detection studies can be grouped into three categories: color, height, and hybrid of both. Starting with color, the information it represents is sensitive to noise. Additionally, illumination conditions of different satellite or aerial images of the same study area may not be similar, therefore, this makes it difficult to compare images in an efficient way. On the other hand, the height information is inadequate alone and getting such data involves complex procedures and requires some additional cost. In consideration of aforementioned issues, this paper introduces a novel approach based on saliency maps which analyze true color (RGB) images. Saliency map is an image which represents the global measure of visual attention on the input image. Generation of a saliency map involves creating maps for different features (hue, intensity, orientation etc.) which contribute to the determination of conspicuous regions. These feature maps are further combined into a single map which indicates the conspicuity level of each pixel. The advantage of using a saliency map is that it can eliminate some types of noise and enable the researcher to compare two images regardless of their illumination levels. In the first step of the presented approach, saliency maps of two different aerial images of a particular area are generated. Any region that is salient for one map but not salient for the other is an important cue of change. From this point of view, at next step, the difference of two images is computed. The results have presented some interesting points which makes this study a potential choice for relevant applications.

Keywords: *Environmental Change Detection, Saliency Map, Urban Planning, Remote Sensing*

Statistical and Biometric Revision of Morphometric Characteristics in Fish of The Genus *Diplodus* (Sparidae) of The Algerian Coast

Kamel HARCHOUCHE¹

¹ *Houari Boumediene University of Science and Technology, Algiers, ALGERIA*

Abstract

The species in the sparid family of the genus *Diplodus* (*D. vulgaris*, *D. sargus sargus*, *D. annularis*, *D. cervinus cervinus* and *D. puntazzo*) are a small group of fish whose systematics have long been controversial. The reexamination of a number of specimens on the basis of biometric, morphometric, meristic and osteological criteria and the use of additional multivariate techniques to the traditional morphometrical methods could bring to the search for new characters of recognition to distinguish the different species, in particular among the young individuals of this group very similar morphologically. The biometric criteria as the multivariate numerical analyzes, in a general way, support the grouping of *D. cervinus cervinus* and *D. puntazzo* into a single species. It is the height of the caudal peduncle that brings it out of the other species. The osteological and meristic characters are concordant to regroup respectively *D. sargus sargus* and *D. puntazzo* as well as *D. annularis* and *D. vulgaris*.

Keywords: *Diplodus*, *Morphometric Characters*, *Algeria*

Developing the Text Summarization API with Natural Language Processing Techniques

Güncel SARIMAN¹

¹Mugla Sıtkı Kocman University, Faculty of Technology, Information Systems Engineering, Mugla, TURKEY

Abstract

With the progress of technology, people are born with various needs. As computer technology provides us with useful solutions at every moment of our lives, expectations from technological products are also diversifying. Intensifying lives have brought people's virtual assistants to find solutions to themselves with faster and more practical solutions. Texts, signs, menus, e-mails, SMS, web pages, advertisements and other text, images and sound that can be found anywhere can reveal how important natural language is. People can handle all their work during the day with almost voice and text. Natural language processing is the analysis of texts in natural languages such as Turkish and English by analyzing the sound waves in the software program and transferring them to the computer environment. Since natural language processing is an automatic calculation of human languages, there are algorithms that take the text produced by man as input and algorithms that produce natural-looking text as output. Natural Language Processing chat bots are used in many intelligent applications such as text summarization, language translation and data definition. A natural language processing process includes steps such as pre-processing, entity extraction, measurements of word frequencies. In this study, an application was made on text summarization. In this study, an application interface (API) has been developed which can be used with different platform and software languages. System selected web page or loaded pdf, word and so on. It is designed to summarize the text in the documents and send the summary text to the requesting platform. Thus, mobile, web and so on. will be able to use abstract text on platforms and even provide a more comfortable usage for the user. For software that uses the API, there is no need to rewrite a program for summarization. The developed text summarization system has been developed by using Django web application framework using different natural language processing libraries (NLTK, Gensim, spaCY, TextBlob) in python language. Statistical text summary techniques were used for summarizing. The apine test was developed on the website developed over English articles and news sites, and the results were compared with the performances of the libraries. The pythonanywhere platform is used to test the Python Web API.

Keywords: Natural Language Processing, Text Summarization, Python

Doğal Dil İşleme Teknikleriyle Metin Özetleme API' sinin Geliştirilmesi

Özet

Teknolojinin giderek gelişmesi beraberinde insanlar için çeşitli ihtiyaçlarında doğmasına sebep olmaktadır. Bilgisayar teknolojisinin hayatımızın her anında bizlere faydalı çözümler üretmesiyle birlikte teknolojik ürünlerden beklentiler de giderek çeşitlenmektedir. Yoğunlaşan hayatlar insanların daha hızlı ve pratik çözümlerle kendilerine çözüm üreten sanal asistanları ortaya çıkarmıştır. Herhangi bir yerde rastlanabilecek metinler, işaretler, menüler, e-postalar, SMS, internet sayfaları, reklamlar ve daha farklı yazı, görüntü ve ses doğal dilin ne kadar önemli olduğunu ortaya çıkarmaktadır. İnsanlar neredeyse ses ve metin ile gün içerisinde tüm işlerini halledebilmektedirler. Doğal dil işleme, Türkçe, İngilizce gibi doğal dillerdeki metinlerin, ses dalgalarının bilgisayar tarafından algılanarak yazılım programında çözümlenmesi ve bilgisayar ortamına aktarılmasıdır. Doğal dil işleme, insan dillerinin otomatik hesaplaması olduğundan, süreçte insan tarafından üretilen metni girdi olarak alan algoritmalar ve çıktı olarak doğal görünümlü metinler üreten algoritmalar bulunmaktadır. Doğal Dil İşleme sohbet botları, metin özetleme, dil çeviri ve veriden görüş tanımlama gibi birçok akıllı uygulamada kullanılmaktadır. Bir doğal dil işleme süreci ön-işleme, varlık çıkarımı, kelime frekanslarının ölçümleri gibi aşamaları içerir. Bu çalışmada metin özetleme üzerine bir uygulama gerçekleştirilmiştir. Çalışmada farklı platform ve yazılım dilleriyle kullanılabilir bir uygulama ara yüzü (API) geliştirilmiştir. Sistem seçilen web sayfası veya yüklenen pdf, word vb. dokümanlardaki metni özetleyerek istekte bulunan platforma özet metni gönderen bir yapıda tasarlanmıştır. Böylece mobil, web vb. platformlarda özet metin kullanılabilir hatta seslendirilerek

kullanıcıya daha rahat bir kullanım sunabilecektir. API'yi kullanan yazılımlar içinde özetleme için tekrar bir program yazmaya gerek kalmayacaktır. Geliştirilen metin özetleme sistemi python dilinde farklı doğal dil işleme kütüphaneleri kullanılarak (NLTK, Gensim, spaCY, TextBlob) Django web uygulama çatısıyla geliştirilmiştir. Özetleme için istatistiksel metin özetleme teknikleri kullanılmıştır. Geliştirilen apinin testi ise İngilizce makaleler ve haber siteleri üzerinden geliştirilen web sitesinde yapılmış, elde edilen sonuçlar ile kütüphanelerin performansları karşılaştırılmıştır. Python Web API' sini test edebilmek için pythonanywhere platformu kullanılmıştır.

Anahtar Kelimeler: *Doğal Dil İşleme, Metin Özetleme, Python*

Identifying Natural Road Accidents Factors Using Spatial Big Data Mining Techniques

Tahar Mehenni¹, Baligh Belkorchia¹

¹ *Mohamed Boudiaf University, Computer Science Department, M'sila, ALGERIA*

Abstract

Road accidents are one of the most imperative factors of death among people. Hence, prevention measures must be taken to overcome the accident rate. Different countries have different geographical and environmental conditions and hence the accident factors diverge in each country. For avoiding possible accidents in rural and urban areas, useful information has to be extracted from the road accident data, using big data mining approaches, such as decision trees and association rules. We propose a Spatial Big Data mining approach using association rules in order to identify the most frequent accident factors related to geographical and environmental conditions. The entire datasets have been imported from different sources. The information extracted from these huge datasets shows that the developed models could provide new several predictions that can assist the authority to improve road safety.

Keywords: *Road Accident, Spatial Big Data, Association Rules, ArcGis*

Visualization of Geographic Information System (GIS) Data with R Programming Language

İbrahim Edibali ATALAY¹, Serdar NESLİHANOĞLU¹

¹ Eskişehir Osmangazi University, Department of Statistics, Eskişehir, TURKEY

Abstract

Geospatial information systems that collect, store, update, process, analyze, and include hardware, software and personnel are commonly referred to as Geographic Information Systems (GIS). In other words, GIS is an information system that allows to map geographically identifiable literary and vector data in the same environment and to query and analyze these data for the specified purposes. Along with the developing technology, GIS can be described as a computer-based database management system that generally analyzes, stores, manages and visualizes geographic information on a map by digitalizing it. After the data collected by GIS is transferred to computer environment, the importance of using visualization methods to reveal, understand, explain and interpret the existing structures in the process of discovery and analysis of these data is increasing day by day. For this reason, in our study, we focused on making graphical visualization and animations of GIS data more efficient and efficient. For this purpose, graphical visualization and animations of GIS data were made by using ggplot library and ganimate package in R programming language and applied on GIS data obtained from European Union countries. As a result, graphical visualizations and animations developed in the R program language provided a different perspective to the visualizations in the field of GIS. **Keywords:** Animation, Geographical Information Systems (GIS), ggplot, Data Visualization

Coğrafi Bilgi Sistemi (CBS) Verilerinin R Programlama Dili ile Görselleştirilmesi

Özet

Coğrafi bilgiyi toplayan, saklayan, güncelleyen, işleyen, analiz eden ve içerisinde donanım, yazılım ve personel bulunan konumsal bilgi sistemleri genel olarak Coğrafi Bilgi Sistemleri (CBS) olarak adlandırılmaktadır. Başka bir deyişle CBS, coğrafi olarak tanımlanabilen yazınsal ve vektörel verileri aynı ortamda eşleme ve belirlenen amaçlar doğrultusunda bu verilerin sorgulanmasını ve analiz edilmesini sağlayan bilgi sistemidir. Gelişen teknolojiyle birlikte CBS, coğrafi bilgileri dijital yapıya kavuşturarak genellikle bir harita üzerinde analiz eden, saklayan, yöneten ve görselleştiren bilgisayar temelli bir veri tabanı yönetim sistemi olarak da nitelendirilebilir. CBS ile toplanan veriler bilgisayar ortamına aktarıldıktan sonra bu verilerin keşifi ve analizi sürecinde var olan yapıların kolaylıkla ve daha anlaşılır bir şekilde ortaya çıkartılması, anlaşılması, anlatılması ve yorumlanması için görselleştirme yöntemlerinin kullanılmasının önemi günden güne artmaktadır. Bu sebepten, çalışmamızda CBS verilerinin grafiksel görselliğinin ve animasyonlarının daha etkin ve verimli yapılması konusuna odaklanılmıştır. Bu amaçla, çalışmamızda yenilik olarak CBS verilerinin grafiksel görselliği ve animasyonları R programlama dilinde ggplot kütüphanesi ile ganimate paketi kullanılarak yapılmış ve Avrupa Birliği ülkelerinden elde edilen CBS verileri üzerinde uygulanmıştır. Sonuç olarak, R programa dilinde geliştirilen grafiksel görsellikler ve animasyonlar CBS alanındaki görselleştirmelere farklı bir bakış açısı sağlanmıştır.

Anahtar Kelimeler: Animasyon, Coğrafi Bilgi Sistemleri (CBS), ggplot, Veri Görselleştirme

Determination of Heavy Metal Levels in Liver and Muscle Tissues of Carp (*Cyprinus carpio* L., 1758) from Atatürk Dam Lake

Ayşe Gül ŞAHİN¹, Rıdvan TEPE¹, Tunay ŞEKER¹, Mehmet KÜÇÜKYILMAZ¹, Gürel Nedim ÖRNEKÇİ¹, Özden Barım ÖZ², Mine ERİŞİR², Hülya ŞAHİN², Sema TÜRKOĞLU², Özgür FIRAT³, Fulya BENZER⁴

¹ Elazığ Water Product Research Institute, Elazığ, TURKEY

² Fırat University, Elazığ, TURKEY

³ Adıyaman University, Elazığ, TURKEY

⁴ Munzur University, Elazığ, TURKEY

Abstract

In the present study, it was aimed to determine concentrations of some heavy metals in liver and muscle tissue of carp (*Cyprinus carpio* L., 1758) from Atatürk Dam Lake. For this purpose, three different locations on Atatürk Dam Lake, Gerger selected as a reference point, Hosirge, Samsat and Sitalce which are exposed to pollution, were investigated. To assess the current state of this heavy metal accumulation, water and fish samples were taken from each station. In addition, physico-chemical and mineral analyses of the water samples were carried out during the study. Mineral concentrations as copper (Cu), iron (Fe), zinc (Zn), chromium (Cr), nickel (Ni) and cadmium (Cd) were measured using atomic absorption spectroscopy (AAS) for fish species from each sex. Regarding the results from physico-chemical analyses, Atatürk Dam Lake was considered the second class in terms of general parameters. Considering the values, the reservoir can be hence expressed as high-quality water. The water samples taken seasonally from Atatürk Dam Lake and metal levels in carp obtained by hunting were determined by analysis of the heavy metal concentrations depending on height, weight and sex. As a result, the concentrations of heavy metals in water, muscle and liver tissues of *Cyprinus carpio* were found below the permissible heavy metal level by the Environmental Protection Agency (EPA), the World Health Organization (WHO), the Food and Agriculture Organization (FAO) and the Turkish Food Codex (TGK).

Keywords: Atatürk Dam Lake, *Cyprinus Carpio*, Heavy Metal, Muscle, Liver

Atatürk Baraj Gölü'ndeki Sazan (*Cyprinus carpio* L., 1758)'nın Kas ve Karaciğer Dokularında Ağır Metal Düzeylerinin Belirlenmesi

Özet

Bu çalışmada, ülkemizin en önemli baraj göllerinden olan Atatürk Baraj Gölü'nde yaşayan sazan balığı (*Cyprinus carpio* L., 1758)'nın kas ve karaciğer dokularında bazı ağır metal düzeylerinin belirlenmesi amaçlanmıştır. Bu nedenle Atatürk Baraj Gölü üzerinde biri referans olarak seçilen Gerger, diğerleri kirliliğe maruz kalan Sitalce, Samsat ve Hosirge olmak üzere dört farklı istasyon belirlenmiştir. Bu ağır metal birikiminin mevcut durumunu belirlemek için her istasyondan su ve balık örnekleri alınmıştır. Ayrıca çalışma süresince baraj gölünün gerek yerinde gerekse alınan su numunelerinin mineral madde düzeylerinin yanı sıra fizikokimyasal analizleri yapılmıştır. Balık örnekleri her bir cinsiyet için ayrı olmak üzere kas ve karaciğer dokusunda mineral madde seviyeleri bakır (Cu), demir (Fe), çinko (Zn), krom (Cr), nikel (Ni) ve kadmiyum (Cd) konsantrasyonları atomik absorpsiyon spektrometre (AAS) cihazı kullanılarak tayin edilmiştir. Atatürk Baraj Gölü fizikokimyasal analizleri bakımından değerlendirildiğinde genel parametreler bakımından II. sınıf olarak belirlenmiştir. Bu değerler bakımından baraj gölü, yüksek kaliteli su olarak ifade edilebilir. Atatürk Baraj Gölü'nden mevsimsel olarak alınan su örnekleri ve avlanarak elde edilen sazan balığının boy, ağırlık ve cinsiyete bağlı olarak ağır metal derişim düzeyleri analizlerle tespit edilmiştir. Sonuç olarak, hem su numunelerinde hem de *Cyprinus carpio*'nun kas ve dokusunda tespit edilen ağır metal konsantrasyonlarının Çevre Koruma Ajansı (EPA), Dünya Sağlık Örgütü (WHO), Gıda ve Tarım Örgütü (FAO) ve Türk Gıda Kodeksi (TGK)'nin önerdiği kabul edilebilir ağır metal değerlerinin altında olduğu belirlenmiştir.

Anahtar Kelimeler: Atatürk Baraj Gölü, *Cyprinus Carpio*, Ağır Metal, Kas, Karaciğer

Molecular Cloning and Functional Characterisation of an Ω_3 Desaturase from The Common Ragworm *Hediste (Nereis) Diversicolor* (Müller, 1776)

İbrahim GÜR¹, Oscar MONROIG², Naoki KABEYA³

¹ Water Product Research Institute, Elazığ TURKEY

² Stirling University, Institute of Aquaculture, Stirling, SCOTLAND

³ Marine Science and Technology University, Tokyo, JAPAN

Abstract

The goal of this study was the cloning and functional characterisation of HD ω_3 desaturase from *Hediste diversicolor*. PCR techniques were used to isolate two cDNA sequences from RNA library, namely HD ω_3 desaturase, to clone and functionally characterise this enzyme through expression in *Saccharomyces cerevisiae*. This enzyme is involved in the ω_3 LC-PUFA biosynthetic pathway and introduces the addition of a double bond into this biosynthetic pathway. The full-length of HD ω_3 consisted of 1167 base pairs (bp) encoding a putative protein of 388 aa. The sequence and phylogenetic analysis provided strong evidence of the identity of this gene and subsequent protein as orthologues of ω_3 desaturase as it shares a variety of protein sequences among annelid and mollusc species. Functional characterization of HD ω_3 demonstrated it possessed ω_3 desaturase activity. Functional characterisation of HD ω_3 in yeast showed its ability to convert 18:2n-6 into 18:3n-3, and 18:3n-3 into 18:4n-3, confirming the activity towards C18 fatty acid substrate possessing $\Delta 15$. As a consequence, the functional activity of HD ω_3 demonstrated that the gene enables this species for de novo ω_3 biosynthesis.

Keywords: *Hediste Diversicolor*, De Novo, Omega-3 Biosynthesis, Desaturase

Halkalı Solucan *Hediste (Nereis) Diversicolor* (Müller, 1776)'dan ω_3 Desatürazın Moleküler Klonlanması ve Fonksiyonel Karakterizasyonu

Özet

Bu çalışmanın amacı halkalı deniz solucanı *Hediste diversicolor*'dan HD ω_3 desatürazın moleküler klonlanması ve fonksiyonel karakterizasyonudur. PCR tekniği kullanılarak total RNA kütüphanesinden cDNA dizilimi, yani HD ω_3 desatüraz, başarılı bir şekilde hedef türden izole edilmiş ve enzimin fonksiyonel olarak karakterizasyonu maya kültürü *Saccharomyces cerevisiae* gerçekleştirilmiştir. HD ω_3 'nın tam uzunluğu 1167 nükleik asit çiftinden oluşmuş olup, toplamda 388 amino asidi kodlamaktadır. İzole edilen ve PCR tekniği kullanılarak başarılı bir şekilde klonlanan bu enzimin uzun zincirli ω_3 doymamış yağ asitlerinin de novo sentezlenmesinde rol almaktadır. DNA dizilim ve filogenetik analiz sonuçları izole edilen HD ω_3 desatüraz diğer türlerle, özellikle yumuşakçalar ve halkalı solucanlar ile güçlü benzerlik ilişkisi içinde olduğu belirlenmiştir. Maya kültüründe HD ω_3 'ın fonksiyonel karakterizasyonu bu genin ω_3 aktiviteye sahip olduğunu göstermiştir. HD ω_3 'ın maya kültüründe fonksiyonel karakterizasyonu enzimin 18:2n-6 yağ asidini 18:3n-3 yağ asidine aynı zamanda 18:3n-3 yağ asidini 18:4n-3 yağ asidine dönüştürebilmiştir. Bu çalışma ile birlikte hedef türden izole edilen ve fonksiyonel olarak karakterizasyonu gerçekleştirilen HD ω_3 desatüraz $\Delta 15$ aktivite göstermiş, C18 uzun zincirli yağ asitlerin sentezinde rol aldığı ispatlanmıştır.

Anahtar Kelimeler: *Hediste Diversicolor*, De Novo Omega-3 Biyosentez, Ω_3 Desatüraz

Visualization and Animation of Temporal Turkey Data

Serdar NESLİHANOĞLU¹, Ömer ÖZDEN¹

¹ *Eskişehir Osmangazi University Department of Statistics, Eskişehir, TURKEY*

Abstract

The graphical representations of the data first started with visual descriptions and map making and continued intensively in natural sciences, engineering, medicine and other fields of science. The main purpose of data visualization methods; it is easier to comprehend, understand and interpret the complex and diffuse information obtained from classical analysis methods. Data visualization continued until today, starting from the estimated data table in Egypt in the second century. Data visualization in the early 1990s, data visualization methods were used intensively in parallel with the fact that the computer was used more intensively and effectively in many fields of science. In the 21st century, data visualization studies in various scientific fields are increasing day by day with the contribution of the rapid development of technology. Data visualization is inspired by the continuous rise in the importance of work in the field. The course focuses on an overview of the history of data visualization from the beginning to the present and the visualization of the statistical data in the various science fields in the R programming language. In this study; Using the ggplot library in the R programming language effectively, a different perspective is placed on the visualization and animations of the data. In addition, visualization of temporal data on Turkey map of Turkey and their animations combined with fast, easy and effective way to be perceived, are provided allowing understanding and interpretation. As a result, a different perspective to the literature visualization of temporal data gained Turkey.

Keywords: *Animation, ggplot, Data Visualization*

Zamansal Türkiye Verilerinin Görselleştirilmesi ve Animasyonları

Özet

Verilerin grafiksel gösterimleri ilk olarak görsel betimlemeler ve harita yapımı ile başlamış ve doğa bilimleri, mühendislik, tıp ve diğer bilim alanlarında yoğun bir şekilde kullanılarak devam etmiştir. Veri görselleştirme yöntemlerinin temel amacı; klasik analiz yöntemlerinden elde edilen karmaşık ve dağınık bilgileri derleyip görselleştirerek daha kolay olarak algılanabilir, anlaşılabilir ve yorumlanabilir hale getirmektir. Veri görselleştirme Mısır'da 2. yüzyılda oluşturulduğu tahmin edilen veri tablosundan başlamak üzere günümüze kadar devam etmiştir. Veri görselleştirme 1990'lı yılların başlarında bilgisayarın birçok bilim alanında daha yoğun ve etkin olarak kullanılmaya başlamasına paralel olarak, veri görselleştirme yöntemleri de bu alanlarda yoğun olarak kullanılmaya başlanmıştır. 21. yüzyılda ise teknolojinin hızlı gelişiminin de katkısıyla çeşitli bilim alanlarında veri görselleştirme çalışmaları gün geçtikçe artarak devam etmektedir. Veri görselleştirme alandaki çalışmaların önemini sürekli olarak yükselişinden esinlenilerek bu çalışmada; başlangıcından günümüze veri görselleştirme tarihine genel bir bakışa ve çeşitli bilim alanındaki istatistiksel verilerin R programlama dilinde görselleştirilmesine odaklanılmıştır. Bu çalışmada; R programlama dilindeki ggplot kütüphanesi etkin bir şekilde kullanılarak verilerin görselleştirilmesine ve animasyonlarına farklı bir bakış açısı getirilmiştir. Buna ek olarak, Türkiye haritası üzerinde zamansal Türkiye verilerinin görselleştirilmesi ve bunların animasyonlar ile birleştirilerek hızlı, kolay ve etkin bir şekilde algılanmasına, anlaşılmasına ve yorumlanmasına imkân sağlanmıştır. Sonuç olarak, zamansal Türkiye verilerinin görselleştirilmesi literatürüne farklı bir bakış açısı kazandırılmıştır.

Anahtar Kelimeler: *Animasyon, ggplot, Türkiye, Veri Görselleştirme*

Portfolio Risk Estimation via ARMA-GARCH Kapula Approach

Emre YILDIRIM¹, Mehmet Ali CENGİZ²

¹ Ondokuz Mayıs University, Samsun, TURKEY

Abstract

One of the most important issues in portfolio risk estimation is the accurate modeling of the dependency structure between the assets that construct the portfolio. Classical portfolio approaches use multivariate normal distribution in portfolio risk estimation for the related financial assets. However, the experimental results in many studies in the literature have shown that the assumption of multivariate normality fails to model the joint distribution of financial variables. Copula approach, which models flexibly the dependency structure between variables and provides effective results in constructing new multivariate distributions, is employed in this study to create the distribution of relevant financial assets. Risk estimates are performed at 95% and 99% confidence level for the portfolio consisting of USD / TRY and TRY / TRY weekly exchange rates.

Keywords: Risk Estimation, Portfolio Management, Financial Assets, Copula

ARMA-GARCH Kapula Yaklaşımıyla Portföy Risk Tahmini

Portföy risk tahmininde en önemli konulardan biri de portföyü oluşturan varlıklar arasındaki bağımlılık yapısının doğru bir şekilde modellenmesidir. Klasik portföy yaklaşımları ilgili finansal varlıklara ilişkin portföy risk tahmininde çok değişkenli normal dağılım kullanılmaktadır. Ancak literatürdeki yapılan pek çok çalışmadaki deneysel sonuçlar, çok değişkenli normal varsayımının finansal değişkenlerin ortak dağılımını modellemede başarısız olduğunu ortaya koymuştur. Değişkenler arasındaki bağımlılık yapısını esnek bir şekilde modelleyen ve yeni çok değişkenli dağılımları oluşturmada etkili sonuçlar sunan kapula yaklaşımı bu çalışmada ilgili finansal varlıklara ilişkin dağılımı oluşturmada kullanılmaktadır. USD/TRY ile YEN/TRY haftalık döviz kurlarından oluşan portföye ilişkin % 95 ve % 99 güven seviyesinde risk tahmini yapılmaktadır.

Anahtar Kelimeler: Risk Tahmini, Portföy Yönetimi, Finansal Varlıklar, Kapula

Comparing Statistical Surveillance Methods for the Detection of Infectious Disease Outbreaks

Zeynep Filiz EREN DOĞU¹

¹*Mugla Sitki Kocman University, Mugla, TURKEY*

Abstract

In public health monitoring, disease outbreak detection has gained more importance in the recent years. Early detection of emerging infectious diseases has been proven to provide timely information for effective control and intervention of outbreaks and diminish the threat of those diseases. A variety of approaches have been applied to public health data to detect outbreaks by automated and prospective surveillance of diseases and syndromes. International, and for some countries nation-wide automated surveillance systems have been widely applied to detect epidemics and pandemics. The objective of this study is to compare the performance of different statistical surveillance methods to detect respiratory disease outbreaks. Residual CUSUM and EWMA charts are used to detect disease outbreaks retrospectively with different types of simulated time series data. The performance of algorithms are then investigated to evaluate the best practice. Finally, the results of this study is compared with the findings of other researchers for further evaluation and conclusions.

Keywords: *Surveillance, Respiratory Infections, Quality Control, Time Series*

The Calculation of the Ship Hull Biofouling Surface, As a Source of the Introduction of Invasive Species at the Port of Arzew

Abderrahmane Bouda¹, Adel Kacimi², Billel Bensari³

¹*Ecole Nationale Supérieure Maritime ENSM, Bou-Ismaïl, ALGERIA*

²*National School of Marine Sciences and Coastal Management ENSSMAL, ALGERIA*

³*Houari Boumediene University of Science and Technology, Algiers, ALGERIA*

Abstract

Historically, the biofouling of ships' hulls has always been considered amongst the oldest mechanisms causing the introduction of marine invasive species. However, in the last few decades, studies have been oriented mainly towards the introduction of non-indigenous species via ballast water, at the expense of biofouling. In this study, an alternative approach that deals with this latter parameter as the main factor of the introduction of marine organisms is proposed. The biofouling surface area of all vessels which have made call at Arzew port in Algeria is calculated over a period of one year. This parameter indicates the importance of the pressure of propagule that corresponds to the successful introduction effort of a species, since it is this submerged surface of the ship that hosts organisms and spreads them all over the world. We associate to each calculated biofouling surface a bioregion of origin, as well as other parameters such as the number of vessels, their maritime routes and the environmental similarity of Arzew port's seawater and the other different bioregions. We highlight according to the bioregion, the surfaces that constitute a critical risk of the species introduction and those that constitute a minor risk. The study showed that over a period of 1 year, 1313 ships from 28 different bioregions that have made call at Arzew port have accumulated a total wet surface corresponding to about nine million square meters. It is worth to mention that this field of research requires special attention from academics as it may lead to more effective methods to prevent future exotic species introductions.

Keywords: *Surface of Biofouling, Invasive, Species, Ship, Arzew Port*

NetSim: A Tool for Network Simulation

Aziz KANDEMİR¹, Zeynep Filiz EREN DOĞU¹, Cihat CETINKAYA¹

Muğla Sıtkı Kocman University, Muğla, TURKEY

Abstract

Modeling network traffic is an important subject which is widely studied by researchers and engineers. For network developers, establishing a real network to test different scenarios is very difficult, time-consuming and expensive. An effective network simulator can allow them i) to check how a newly established networking protocol will work and ii) to test how an existing protocol will be affected by a change, in a controlled and reproducible environment. The aim of our study is to develop an easy-to-use GUI-based network simulation tool for modeling and simulating and analyzing a network.

Keywords: *Network, Simulation, Modelling, Throughput*

Predicting Stock Prices Using Social Media Data

İbrahim SABUNCU¹, Melih KÖKTÜRK², Aykut KURUL³, Mustafa Eray ÜNAL⁴
Yalova University, Faculty of Engineering, Industrial Engineering, Istanbul, TURKEY

Abstract

In today's world, the internet is used to help individuals and companies in marketing their products, services, and ideas. In this respect, social media is considered as an important source of data for sharing these activities. In this study, Brand24, a social media data collection tool, was used on the social media platforms ("Facebook", "Twitter", "Instagram", "Forums and Blogs".) in order to collect social media mentions of 8 companies that selected ("Samsung", "Apple Inc.", "Best Buy", "Alibaba Group Holding Limited", "Walmart Inc.", "Amazon.com Inc.", "Tesco", "Costco Wholesale Corporation"). Also daily stock market values between 24 February and 10 May, total of 76 days were collected. A "Sentiment Analysis" was applied to the collected social media data. With the Sentiment Analysis results, the data was classified as "Positive", "Negative" and "Neutral". Then, a "Correlation Analysis" was performed between the total, positive, negative and net positive number of mentions and the stock value of the current and next day. As a result, only 26 of 64 Correlation Analyzes were found to be significant. As a result, it can be seen that social media data can be used to estimate stock values.

Keywords: *Social Media Analytics, Sentiment Analysis, Stock Price Prediction*

Investigation of Criteria for The Best Architecture Selection of Artificial Neural Networks

Cagatay BAL¹, Serdar DEMIR¹

¹ *Mugla Sitki Kocman University, Mugla TURKEY*

Abstract

The architecture selection in artificial neural networks is a process which determines a satisfactory neural network model(s) that will give the most accurate forecasts. The architecture that minimizes the error between the target values of the neural network which are the real observations and the predicted values generated by the model represents the best forecasts namely the most appropriate model. There are many common criteria that are used in the literature to accomplish this process. In addition, there are the modified criteria which are created by combining the common criteria and called as weighted criteria. In this study, the performances of the criteria available in the literature are compared by applying to both simulated and real-world data sets. For this purpose, three different exchange rates time series, four simulated time series with different features, and three well-known real-world data sets are used. The results show that the performances of criteria are varying between different datasets but on the most cases weighted criteria perform similar or better among other successful criteria on each dataset.

Acknowledgement: This study is supported as 16/130 numbered Scientific Research Project by Mugla Sitki Kocman University, TURKEY.

Keywords: *Artificial Neural Networks, Performance Criteria, Best Architecture, Time Series*

An RFM-Based Customer Segmentation Model for Targeted Offers in Mobile Gaming Industry

Inanç KABASAKAL ¹

¹*Ege University, Izmir, TURKEY*

Abstract

Mobile gaming has been a significantly rapid growing industry over the last decade within the video game sector. The pervasion of broadband Internet and the progress in mobile phones have enabled mobile gaming and created a massive market with billions of potential customers. On the other hand, the two major ecosystems Google Play and App Store host a vast number of games developed by various software houses. In this competitive environment, retention of customers is crucial as well as attracting new customers. For such purpose, designating offers for appropriate customer segments is essential to enhance the impact of campaigns. For customer segmentation, various clustering models have been proposed in data mining field. In this study, a behavioral customer segmentation model will be proposed for a business in mobile game development. Customer segmentation will be conducted with the Recency, Frequency, and Monetary (RFM) Model over the transactional sales data obtained from the business. In the results, the use of RFM segments in marketing promotions will be discussed.

Keywords: *Customer Segmentation, Mobile Gaming Industry, Clustering, RFM Model*

Regime Switching Fuzzy Time Series Model

Nevin GÜLER DİNCER¹, Muhammet Oğuzhan YALÇIN¹, Öznur İŞÇİ GÜNERİ¹

¹ Mugla Sitki Kocman University, Faculty of Science, Department of Statistics, Mugla, Turkey

Abstract

Fuzzy time series (FTS) models have been widely used in forecasting of time series because of their advantages that do not require strict statistical assumptions such as stationarity, number of observations and have the ability of modeling time series consisted of linguistic values. Besides, FTS models can directly handle nonlinear time series. However, many FTS models suggested up to now can fail to satisfy in modeling time series depicted abrupt changes in their behavior or had different behaviors at the different time periods since they based on building single fuzzy relation structure. This study proposes a new FTS model called regime switching fuzzy time series (RSFTS) in order to overcome that kind of disadvantage of existing FTS models and thus improve forecasting performance. Proposed method consists of two steps. In the first step, the regimes are determined and in the second step, a fuzzy time series model is built for each regime. In order to show the performance of proposed method, four FTS models based on fuzzy clustering and proposed version of these models are applied to 29 number time series electricity consumption per capita (ECPC) of Asia countries and their performance are compared by using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

Keywords: *Fuzzy Time Series, Nonlinear Time Series Analysis, Switching Time Series, Forecasting*

Investigation of Ethnic and Cultural Differences in International Business in the Context of Generations

Gözde MERT¹

¹Nisantasi University, Istanbul, TURKEY

Abstract

International businesses include people of different ethnic backgrounds and cultures. This situation is important in the activities of businesses in the international market and businesses have to cooperate with individuals of different ethnic origin and culture. Increasing ethnic and cultural differences necessitates their management. Businesses with more ethnic and cultural diversity face some problems in managing employees. In addition to this, the characteristics of the employees change with respect to age, and people belonging to the same age groups also constitute generations that exhibit common values, emotions, behaviors and thinking. The expectations of the X and Y generation in business life differ from each other and affect the management and leadership policies of the enterprises. For this purpose, the purpose of the study; whether the ethnic and cultural differences of those working in international businesses differ in the context of the X and Y generations. For this purpose, a survey was applied to the managers and employees in the X and Y generations in the international enterprises operating in Istanbul. 285 people participated in the study. Parametric techniques were used for data analysis. At the end of the research, suggestions on management and leadership policies were presented to international enterprises in the X and Y generations, who have different ethnic backgrounds and have a multicultural structure.

Keywords: *International Businesses, Cultural difference, Ethnicity, Generation X, Generation Y*

Uluslararası İşletmelerde Etnik ve Kültürel Farklılıkların Kuşaklar Bağlamında İncelenmesi

Özet

Uluslararası işletmelerde farklı etnik kökenlerden ve kültürden insanlar bulunmaktadır. Bu durum işletmelerin uluslararası piyasadaki faaliyetlerinde önem arz etmekte ve işletmeler farklı etnik kökene ve kültüre sahip bireylerle iş birliği yapmak durumundadır. Etnik ve kültürel farklılıkların artması, bunların yönetimini de gerekli kılmaktadır. Etnik ve kültürel çeşitliliği fazla olan işletmeler, çalışanların yönetilmesi konusunda birtakım sorunlarla karşılaşmaktadır. Bu duruma ek olarak çalışanların özellikleri yaş ile bağlantılı olarak da değişmekte, aynı yaş gruplarına mensup kişiler aynı zamanda ortak değer, duygu, davranış biçimi ve düşünce yapısı sergileyen kuşakları oluşturmaktadır. X ve Y kuşağının iş yaşamındaki beklentileri ise birbirinden farklılık göstermekte ve işletmelerin yönetim ve liderlik politikalarını da etkilemektedir. Bu maksatla, çalışmanın amacı; uluslararası işletmelerde iş görenlerin etnik ve kültürel farklılıklarının, X ve Y kuşağı bağlamında farklılık gösterip-göstermediğini ortaya koymaktır. Bu amaçla, İstanbul'da faaliyet gösteren uluslararası işletmelerde, X ve Y kuşağında bulunan yönetici ve çalışan personele anket uygulanmıştır. Araştırmaya 285 kişi katılmıştır. Verilerin analizinde parametrik teknikler uygulanmıştır. Araştırma sonunda X ve Y kuşağında bulunan, farklı etnik kökenli çalışana ve çok kültürlü bir yapıya sahip olan uluslararası işletmelere, yönetim ve liderlik politikaları konularında öneriler sunulmuştur.

Anahtar Kelimeler: *Uluslararası İşletmeler, Kültürel Farklılık, Etnik Köken, X Kuşağı, Y Kuşağı*

Classification of Eye Blinks in EEG Signals

Halil ÖZTÜRK ¹, Ali Osman SELVİ ²

1 Bilecik Seyh Edebali University, Faculty of Engineering, Computer Engineering Department, Bilecik, TURKEY

2 Bilecik Seyh Edebali University, Vocational School, Department of Computer Programming, Bilecik, TURKEY

Abstract

Brain activity is recorded by Electroencephalography (EEG) devices placed on the skull. EEG signals contain a lot of information. Those informations is used in many areas such as Brain Computer Interface (BCI) studies. The purpose of these systems is to provide direct control of computers or other devices via computers, using direct thoughts. EEG signals include muscle signals from body movements or disturbing signals from external factors. In such systems, the EEG signal is removed from the disturbing signals before processing. In this study, EEG signals were recorded with EMOTIV EPOC+ portable EEG device. This device records from 14 channels. In this study, AF3, AF4, F7 and F8 channels were used. This study focuses on the eye muscle signals within the EEG signal. Right-eye, left-eye and two-eye blink classification were made by using deep artificial intelligence algorithms on Matlab platform.

Keywords: *Electroencephalogram (EEG), Brain computer interface (BCI), Eye-blink*

EEG Sinyalleri İçerisindeki Göz Kırpma İşaretlerinin Sınıflandırılması

Özet

Beyin aktiviteleri kafatası üzerine yerleştirilen Elektroensefalografi (EEG) cihazları ile kaydedilir. EEG sinyalleri bünyesinde birçok bilgi barındırmaktadır. Bu bilgiler Beyin Bilgisayar Arayüzü (BBA) çalışmaları gibi bir çok alanda kullanılmaktadır. Bu sistemlerin amacı doğrudan düşünceleri kullanarak, bilgisayarların veya başka cihazların bilgisayarlar aracılığı ile kontrolünü sağlamaktır. EEG sinyalleri bünyesinde vücut hareketlerinden kaynaklı kas sinyallerini veya dış etkenlerden kaynaklı bozucu sinyalleri de barındırmaktadır. Bu tarz sistemlerde EEG sinyali işlenmeden önce bozucu sinyallerden arındırılmaktadır. Bu çalışmada EEG sinyallerinin kaydı EMOTIV EPOC+ taşınabilir EEG cihazı ile yapılmıştır. Bu cihaz 14 kanaldan kayıt yapmaktadır. Bu çalışmada kafa tasının ön kısmında bulunan AF3, AF4, F7 ve F8 kanalları kullanılmıştır. EEG sinyali bünyesinde bulunan göz kası sinyallerine odaklanılmıştır. Matlab platformu üzerinde derin yapay zeka algoritmaları kullanılarak sağ, sol ve iki göz kırpması sınıflandırılması yapılmıştır.

Anahtar Kelimeler: *Elektroensefalografi (EEG), Beyin Bilgisayar Arayüzü (BBA), Göz Kırpması*

Clustering of Biometric Data

Aykut DURGUT¹, Serdar BIROGLU²

¹ Balıkesir University, Balıkesir, TURKEY

² Duzce University, Duzce, TURKEY

Abstract

In this study, clustering methods have been tried to improve performance in the comparison stage of biometric data. The main purpose of using clustering method is to provide fast results as a result of querying biometric data from database. We tried to achieve the best performance by applying clustering in various ways. Kmeans, means, classification tree, hierarchical clustering, two-step clustering methods were applied with SPSS program. Artificial neural networks have been tried to adapt to the system, but failed. The reason is that the number of points is not constant, so there are problems in the learning phase. Since the number of points cannot be obtained in a fixed number in every biometry and the data increase is observed, the clustering results are affected, but the best results are obtained by K-means clustering method.

Keywords: *Biometric, Data, Clustering*

Biyometrik Verilerin Kümelenmesi

Özet

Bu çalışmada biyometrik verilerin karşılaştırma aşamasında performans arttırımı için kümeleme yöntemleri denenmiştir. Kümeleme yönteminin kullanılmasının temel amacı biyometrik verinin veri tabanından sorgulanması sonucunda hızlı sonuç göstermesini sağlamaktır. Kümelemeyi çeşitli şekillerde uygulayarak en iyi performans elde edilmeye çalışıldı. Kmeans, means, sınıflandırma ağacı, hiyerarşik kümeleme, iki adım kümeleme metotları SPSS programı ile uygulanmıştır. Yapay sinir ağları sisteme adapte edilmeye çalışılmış fakat başarılı olunamamıştır. Nedeni nokta sayılarının sabit olmaması dolayısı ile öğrenme aşamasında sorunlar yaşanılmasıdır. Nokta sayıları her biyometride sabit sayıda elde edilemediğinden ve veri artışı olduğunda kümeleme sonuçlarının etkilendiği fakat buna rağmen en iyi sonuçların K-means kümeleme yöntemi ile elde edildiği belirlenmiştir.

Anahtar Kelimeler: *Biyometrik, Veri, Kümeleme*

A Hybrid Feature Selection Algorithm to Enhance Turkish News Classification

Emin BORANDAG¹, Akın OZCIFT¹, Yeşim KAYGUSUZ¹

¹*Manisa Celal Bayar University Department of Software Engineering*

Abstract

There is a continuous increase in the textual information from various sources such as news, social media, reviews of customers and similar kinds of digital documents. Automatic extraction of categories with the use of Machine Learning algorithms from unstructured text is an important field of text mining. Any text to be processed with classifiers should first be represented in a model such as vector space model. However, this model generates high number of irrelevant features. In order to decrease computational cost and increase classification accuracy some feature selection algorithms are needed to be applied to the documents. This study proposes a hybrid feature selection algorithm based of information gain, relief, symmetrical uncertainty and correlation based feature selection algorithms. The proposed feature selection algorithm is applied to Turkish news dataset classification using J48, SVM and Random Forests algorithms. The experimental results show that the proposed hybrid feature selection algorithm is promising.

Keywords: *Feature Selection Method, Text Classification, Text Mining, Machine Learning*

Estimation Modelling of Ames Housing Price with Machine Learning Algorithms

Abdullah ALTAY¹, Serdar NESLIHANOGLU¹

¹Eskisehir Osmangazi University, Eskisehir, TURKEY

Abstract

In the world, the value of the housing is important for the owners and investors who want to buy or sell housing through the mortgage system. Especially when calculating the housing loan, knowing the price of the housing is of great importance for investors. When calculating the value of the house, the price of the house is calculated using different criteria such as the number of rooms in the house, the number of bathrooms and the location of the house. In this study, it is aimed to estimate the house price with the least error by modeling in the light of multiple variables with machine learning algorithms. The Ames Housing data set, a modern and improved version of the Boston Housing data set, compiled by Dean De Cock for frequent use in data science education, was used as the data set for the study, showing housing data in Ames, Iowa, USA. Lasso, Elastic Net, Ridge, Random Forest, Gradient Boosting, Stacking Regression and Support Vector Machine algorithms compared the best performance model in this data set. As a result, it is clear that the Elastic Net model makes a better estimate of the house price estimate in Ames compared to other models.

Keywords: *Advanced Regression Techniques, House Price Estimation, Machine Learning*

Makine Öğrenmesi Algoritmaları ile Ames Konut Fiyat Tahmin Modellemesi

Özet

Dünya'da, konut finansman sistemi (Mortgage) ile konut almak veya satmak isteyen konut sahipleri ve yatırımcılar için konutun değerinin ne kadar olabileceği önem taşımaktadır. Özellikle de konut kredisi hesaplanırken, konutun fiyatının bilinmesi yatırımcılar için büyük önem teşkil eder. Konutun değeri hesaplanırken evdeki oda sayısı, banyo sayısı ve evin lokasyonu gibi çok farklı kriterler kullanılarak, evin fiyatı hesaplanmaktadır. Konut fiyatının, makine öğrenmesi algoritmaları ile birden çok değişkenin ışığında modellenerek en az hata ile tahmin edilmesi bu araştırmada amaçlanmıştır. Araştırmada veri seti olarak, veri bilimi eğitiminde kullanılmak üzere Dean De Cock tarafından derlenen ve sıklıkla araştırmalarda kullanılan Boston Housing veri seti'nin modern ve geliştirilmiş bir versiyonu olan ve Amerika Birleşik Devletleri'nin Iowa eyaletindeki Ames şehrindeki konut verilerini gösteren Ames Housing veri seti tercih edilmiştir. Bu veri setindeki konut fiyat tahmininin en iyi performans ile modellenmesi için Lasso, Elastik Net, Ridge, Random Forest, Gradient Boosting, Stacking Regression ve Support Vector Machine algoritmaları karşılaştırmıştır. Sonuç olarak, Elastik Net modelinin Ames'deki konut fiyat tahmininde diğer modellere kıyasla daha iyi tahmin yaptığı net olarak görülmüştür.

Anahtar Kelimeler: *İleri Regresyon Teknikleri, Konut Fiyat Tahmini, Makine Öğrenmesi*

Comparison of CPU And GPU Based Acceleration Techniques in Machine Learning

Gökhan Egemen ÖĞÜT¹, Hüseyin GÜRÜLER¹

¹ Information Systems Engineering, Muğla Sıtkı Kocman University

Abstract

Within the scope of the study; parallel computing methods have been used on CPU and GPU cores to increase the speed as after comparing the use and not use of parallel computing methods on machine learning and deep learning algorithms. In this context, Intel's CPU accelerated Python libraries were used. This method speeded up from 1.3 to 30 times in different algorithms. Nvidia's GPU accelerated CUDA platform has increased the speed of convolutional neural networks by 10 times. The results obtained predicts the change in the hardware in which the application changes. As a result, it is thought that this study will contribute by describing the acceleration amounts of parallel computing in machine learning and deep learning processes.

Keywords: Machine Learning, Deep Learning, Parallel Computing, Scikit-learn, Cuda, CNN

Makine Öğrenmesinde CPU ve GPU Tabanlı Hızlandırma Tekniklerinin Karşılaştırılması

Özet

Bu çalışmada; CPU ve GPU çekirdekleri ile makine öğrenmesi ve derin öğrenme algoritmaları üzerinde paralel hesaplama yöntemleri kullanılmış olup, paralel hesaplama yöntemleri kullanılmamasına oranla ne kadar hız artışı sağladıkları karşılaştırılmıştır. Bu kapsamda, öncelikle Intel firmasının CPU hızlandırmalı Python kütüphaneleri kullanılmıştır. Bu yöntemin farklı algoritmalarda 1.3 kattan 30 kata kadar farklı hızlar kazandırdığı tespit edilmiştir. Nvidia firmasının GPU hızlandırmalı CUDA platformunun ise evrimsel sinir ağlarında 10 kat hız artışı sağladığı tespit edilmiştir. Çalışmadaki uygulamaların yapıldığı donanım değişikçe alınan sonuçların da değişeceği öngörülmektedir. Sonuç olarak bu tez çalışmasının paralel hesaplama yöntemlerinin makine öğrenmesi ve derin öğrenme işlemlerindeki hızlandırma miktarlarını betimleyerek katkı sağlayacağı düşünülmektedir.

Anahtar Kelimeler: Makine Öğrenmesi, Derin Öğrenme, Paralel Hesaplama, Scikit-learn, Cuda, CNN

ThesisManager: An Editor for Easy Formatting of The Graduate Thesis

Said NADEEM¹, Hüseyin GÜRÜLER¹

¹Information Systems Engineering, Mugla Sıtkı Kocman University

Abstract

This study introduces a web-based software (ThesisManager) has been prepared to help graduate theses in appropriate format. In this way, it allows the graduate students to focus on the main components of the theses such as text, pictures and tables, and to facilitate the writing of the thesis. ThesisManager consists of four parts: 1) Basic inputs; 2) Main menu; 3) Text editor and 4) A section to add references, figures and tables in the main text. The literature review section is linked to Scopus and NCBI databases and references can be added to the library. The table of contents, list of tables and figures are automatically added to the thesis document prepared in the TezManager system. TezManager, which is prepared within the scope of this thesis, enables the parties concerned to focus only on the content of the thesis, will significantly reduce the workload of students, supervisor and institute staff.

Keywords: Text Editor, Thesis Management System, Laravel, Visual Studio Code

TezManager: Lisansüstü Tez Yazım Stilini Kolaylaştırmak için Editör Yazılımı

Özet

Bu çalışmada lisansüstü tezleri uygun formatta hazırlamaya yardım eden Web tabanlı bir yazılım (TezManager) tanıtılmaktadır. Bu sayede, öğrencilerin tezlerin yazım formatı yerine tezlerdeki metin, resim ve çizelge gibi ana bileşenlerine odaklanmasına, yüksek lisans ve doktora tez yazımının kolaylaştırılmasına imkan vermektedir. TezManager temel olarak dört bölümden oluşmaktadır: 1) Temel girdiler 2) Ana menü 3) Metin editörü 4) Ana metine kaynak, şekil ve tablolar ekleme bölümü. Ayrıca, metin içerisinde kaynakların yönetimini için literatür taraması kısmı Scopus ve NCBI veritabanlarına bağlıdır. TezManager sisteminde hazırlanan tez belgesine içindekiler listesi ile şekil ve çizelge dizinleri otomatik olarak eklenir. TezManager sisteminin, tez ile ilgili tarafların sadece tez içeriğine odaklanmalarını sağladığından öğrenciler, ilgili danışmanlar ve enstitü çalışanlarının iş yükünü önemli ölçüde azaltacağı düşüncesindeyiz.

Anahtar Kelimeler: Metin Editörü, Tez Yönetim Sistemi, Laravel, Visual Studio Code

Diabetic Retinopathy Detection Using Computer Vision

Vongani Maluleke¹, Sicelukwanda Zwane²

¹*University of Cape Town, Cape Town, SOUTH AFRICA*

²*University of Witwatersrand, Johannesburg, SOUTH AFRICA*

Abstract

Diabetic Retinopathy (DR) is one of the leading causes of blindness in diabetic patients, especially those with limited access to healthcare facilities. In most cases, early-stage DR can be painless with visible symptoms and partial vision impairment only occurring during advanced stages. As such, early detection is left to the medical specialist. This paper will demonstrate the use of computer vision to detect DR from images of retinas taken using fundus photography. Different models will be compared with each other and the best performing one will be chosen to perform the computer vision task. The images have been labelled by a clinician in accordance to the severity of the DR on a scale of 0 to 4, where 0 represents no DR and 1 to 4 corresponds to the four DR stages, which are namely mild, moderate and severe nonproliferative diabetic retinopathy and proliferative diabetic retinopathy. Over the past few years, computer vision has broadened the landscape of medical screenings and reviews, as automated detection systems, driven by computer vision, provides nurses and doctors with expertise that only specialised individuals have capabilities to perform. An automated DR detection system will be very beneficial in rural areas where there are no or few doctors that have the ability to review eye screenings and diagnose the presence and severity of DR.

Keywords: *Diabetic Retinopathy, Computer Vision, Machine Learning, Medical Screenings*

Triggering Diversity of Artificial Intelligence Based Art Research by using Turkish-Ottoman Art Genre

Aysin TASDELEN¹, Eren BOZARIK², Can U MAY³, Yavuz KOM ECOGLU⁴, Burak KOZAN⁵, Zumrut MUFTUOGLU⁶, Sebnem OZDEMIR⁵

¹*Bilgi University, Istanbul, TURKEY*

²*Karadeniz Technical University, Trabzon, TURKEY*

³*Ankara Yildirim Beyazit University, Ankara, TURKEY*

⁴*Kodiks, Istanbul, TURKEY*

^{5,7}*Istinye University, Istanbul, TURKEY*

⁶*Yildiz Technical University, Istanbul, TURKEY*

Abstract

Along with the acceleration of artificial intelligence(AI) studies around the world, innovative developments not only in engineering but also in different disciplines such as art have been recorded. Related to those developments, artificial intelligence has become a strategic and cultural competition among the countries. This study has been prepared for contributing to that cultural competition by emphasizing one of the Turkish-Ottoman art genres, called Ebru. In order to fulfill that contribution, an unique data set (Deep Novice Ebru-DNE) with 5 classes have been prepared and style transfer algorithm has been applied on it. Thus a different art genre was presented to triggering diversity of AI-based art research. GitHub: <https://github.com/DeepScienceLab/Ebru-Dataset>

Keywords: *Artificial Intelligence, Deep Learning, Art, Ebru*

VGG16-CNN Ethnic Origin Classification with Deep Learning Model

Furkan AKIN¹, Erkan DUMAN¹

¹Firat University, Elazig, TURKEY

Abstract

In recent years, one of the most frequently discussed subjects in the field of computer vision is face analysis. It is widely used in identity verification and recognition studies because it contains very specific demographic characteristics. In cases where individuals should be classified according to their characteristics such as age and gender, facial analysis may be insufficient due to various masks. Conversely, ethnicity, which is difficult to change by using masks, will play a more effective role in the classification. This classification process can be realized with deep learning methods.

Keywords: Deep Learning, Keras, VGG16, UTKFace

VGG16-CNN Derin Öğrenme Modeli ile Etnik Köken Sınıflandırması

Özet

Son yıllarda bilgisayarla görme alanında yapılan çalışmalarda, sıklıkla ele alınan konulardan biri de yüz analizidir. Kişiyeye özgü çok belirgin demografik özellikler barındırması nedeni ile kimlik doğrulama ve tanıma çalışmalarında yaygın olarak kullanılmaktadır. Bireylerin yaş, cinsiyet gibi özelliklere göre sınıflandırılması gerektiği durumlarda, yüz analizi çeşitli maskelemeler nedeni ile yetersiz kalabilir. Bunun aksine maskelemeler kullanılarak zor değiştirilebilecek etnik köken özelliği sınıflandırmada daha etkin rol oynayacaktır. Bu sınıflandırma işlemleri derin öğrenme yöntemleri ile gerçekleştirilebilmektedir.

Anahtar Kelimeler: Derin Öğrenme, Keras, VGG16, UTKFace

Turkey Printing Industry of Several Variables in Terms of Future Projections

Orhan SEVİNDİK¹, Burcu BEKTAŞ², Rüya ŞAMLI³

¹*Cerrahpaşa Istanbul University, Department of Design, Istanbul, TURKEY*

²*National Defense University, the Naval Academy, Computer Engineering, Istanbul, TURKEY*

³*Istanbul-Cerrahpaşa University, Faculty of Engineering, Computer Engineering, Istanbul, TURKEY*

Abstract

Printing House, B.C. Although it has been a technology that has existed since the 3rd century, it was only possible in the 13th century that Guttenberg developed the necessary technology to make the printing system suitable for mass production in order to achieve a structure in the present sense. The printing industry, which is known as an industry in today's world, is the fifth largest branch of business in England and the third largest branch of business in America. The size of the sector's turnover in Turkey at the level of 13 billion Turkish Lira is the turnover of the industry over the past decade has exceeded twice. The study, several variables of the position of the printing industry in the coming years and the situation in Turkey (turnover, investments, employment and consumption) to determine what terms focused on the main problem. In this context, the printing industry has been analyzed by machine learning methods in terms of various variables. In addition, the data obtained were compared with the labor market needs analyzes performed before and the qualified personnel needs of the sector were examined.

Keywords: *Printing Industry, Artificial Intelligence, Machine Learning, Statistics*

Türkiye Basım Endüstrisinin Çeşitli Değişkenler Açısından Gelecek Öngörüsü

Özet

Matbaa, M. Ö. 3. yüzyıldan beri var olan bir teknoloji olmasına rağmen, bugünkü anlamda bir yapıya kavuşabilmesi için 13. yüzyılda, Guttenberg'in baskı sisteminin seri üretime uygun hale gelmesi için gerekli teknolojiyi geliştirmesiyle mümkün olmuştur. Günümüz dünyasında bir endüstri olarak anılan basım sektörü, İngiltere beşinci, Amerika'da ise üçüncü büyük faaliyet koludur. Sektörün Türkiye'deki ciro büyüklüğü ise 13 milyar Türk Lirası düzeyinde olup son on yıllık süreç içinde sektörün cirosu iki katını aşmıştır. Çalışma, Türkiye'deki basım endüstrisinin gelecek yıllarda konumunun ve durumunun çeşitli değişkenler (ciro, yatırım, istihdam, tüketim) açısından ne olduğunu tespit etme ana problemine odaklanmıştır. Bu bağlamda basım endüstrisi çeşitli değişkenler açısından makine öğrenmesi yöntemleri ile analiz edilmiştir. Ayrıca elde edilen veriler daha önce gerçekleştirilen iş piyasası ihtiyaç analizleri ile karşılaştırılarak sektörün ihtiyaç duyacağı kalifiye eleman durumu da incelenmiştir.

Anahtar Kelimeler: *Basım Endüstrisi, Yapay Zeka, Makine Öğrenmesi, İstatistik*

Real Time Estimation of Data from Color Sensor

Kazım BAKAY¹, Ali Osman SELVİ¹

¹*Bilecik Seyh Edebali University, Bilecik, TURKEY*

Abstract

It is an important factor for the development of technology that analogue data available in daily life is displayed as close to reality as possible in the digital environment. The multiplicity and diversity of these data caused the need to be classified. For this reason, tcs3200 color sensor was used to transfer the colors to the digital environment in order to classify the colors that may be needed in many areas and in line with the classification, the classification of colors was realized by using KNN (k-nearest neighbor) algorithm.

Keywords: *KNN, Classification of Colors*

Renk Sensöründen Alınan Verilerin Gerçek Zamanlı Tahmini

Özet

Günlük hayatta var olan analog verilerin dijital ortamda gerçeğe olabildiğince yakın gösterilmesi, teknolojinin gelişmesi için önemli bir etkidir. Aktarılan bu verilerin çokluğu ve çeşitliliği sınıflandırılma ihtiyacının doğmasına neden olmuştur. Bu nedenle birçok alanda ihtiyaç duyulabilecek renklerin sınıflandırılması amacı ile çalışmada renklerin, dijital ortama aktarılması için tcs3200 renk sensörü kullanılmış olup sınıflandırılması doğrultusunda k-en yakın komşu (KNN) algoritmasından faydalanılarak renklerin sınıflandırılması gerçekleştirilmiştir.

Anahtar Kelimeler: *K-En Yakın Komşu, Renklerin Sınıflandırılması*

Feature Extraction of P300 Signals Using Spectral Entropy Method

Ali Osman SELVİ¹, Abdullah FERİKOGLU², Derya GÜZEL³

¹ *Bilecik Seyh Edebali University, Bilecik, TURKEY*

² *Sakarya University of Applied Sciences, Sakarya, TURKEY*

³ *Sakarya University, Sakarya, TURKEY*

Abstract

EEG signals are signals of different amplitude and frequency, which are received via sensors via the skull. P300 signals are positive deflection signals that occur approximately 300ms after the stimulus shown to the participant. P300-based hyphenation systems are usually created by displaying flashing characters in different sequence on the computer screen. In the animation process, green frame was displayed around the desired symbol so that the participant did not focus. Spectral entropy measures how straight the spectrum of a signal is. In this study, spectrogram calculation is made by using short time fourier transform using 255 time windows. The feature extraction vectors were compared between the p300-containing and non-p300-based signals.

Keywords: P300, EEG, Spektral Entropi

P300 Sinyallerinin Spektral Entropi Yöntemi ile Özniteliklerinin Çıkartılması

Özet

EEG işaretleri kafatası üzerinden sensörler aracılığı ile alınan farklı genlik ve frekansa sahip sinyallerdir. P300 sinyalleri katılımcıya gösterilen uyarandan yaklaşık 300ms sonra oluşan pozitif defleksiyon sinyalleridir. P300 tabanlı heceleme sistemleri genellikle bilgisayar ekranında farklı sıra ile yanıp sönen karakterlerin gösterilmesi ile oluşturulmaktadır. Analiz için alınan verilerde 4 uyarıcı katılımcılara sıra ile gösterilmiştir. Animasyon sürecinde katılımcının odaklanmasını istenilen sembol etrafında yeşil renk çerçeve gösterim yapılmıştır. Spektral entropi, bir sinyalin spektrumunun ne kadar düz olduğunu ölçer. Bu çalışmada 255 zaman penceresi kullanarak kısa süreli fourier dönüşümü ile spektrogram hesabı yapılmıştır. Yapılan öznitelik çıkartım sonuçları p300 içeren ve p300 içermeyen sinyaller arasında karşılaştırarak sınıflandırma öncesi öznitelik vektörleri oluşturulmuştur.

Anahtar Kelimeler: P300, EEG, Spektral Entropi

An Expert System of Price Forecasting for Second Hand Cars Using Regression Models

Kübra AKGÖZLÜOĞLU¹, Deniz KILINÇ¹, Fatma BOZYİĞİT¹

¹Manisa Celal Bayar University, Manisa, TURKEY

Abstract

From early human history, transportation has been one of the most important needs of daily life. As urbanization increased, the need to own a car raised. In addition to the new-model cars, people tend towards second-hand vehicles with respect to their financial situation. Consequently, buying and selling second-hand vehicle have a significant rise in world-wide. With the speedy development in digital technology, online platforms enable people to find more alternatives in the second-hand car market. Thanks to these platforms, the buyer and the seller can easily interact and share information with each other. Users can easily get information about the car planned to be bought such as brand, model, color, mileage, etc. They also demand tram records, expert reports, and accident photos (if the car has) from the seller through these platforms. Online platforms provide users to filter the results by selecting the desired features. However, finding optimal alternative answering to user's requests and having the best price can be time-consuming process Price of the second-hand cars having same properties may be different from the each other because parameters affecting the selling value can vary from person to person. Of course, users do not want to give more money for a car while the cheaper alternative is available. Also, if the sellers do not specify their cars' value in a more realistic manner, their advertisements may not arouse the interests of the customers. Considering these problems in the secondary car market, we conduct a study to determine the approximate price of the second-hand car by using frequently used machine learning methods from literature. Our data set was created by selecting the 18 vehicle makes having the highest number of advertisements on "sahibinden.com". We collected 75 advertisements were selected from each vehicle makes and excluded advertisements no, sellers' information (name, contact information and plate number of his/her car, etc.) considering data security policies. As a result, the data set consists of 1350 instances and 15 feature columns which defines the properties of cars. In the proposed methodology, first, we implement feature selection strategy (SelectKBest, Recursive Feature Elimination, Principle Component Analysis, and Linear Discriminant Analysis) on our dataset to obtain the best subset of features increasing system performance. Then, we experiment with different regression models such as Linear, Decision Tree, AdaBoost, Bagging, and Random Forest (RF) and provided the corresponding results in related sections. The optimized version of the RF algorithm on top of reduced feature dimension has produced an F-measure value of 0.90 which is the most successful result obtained. We believe that the results of this study are motivating enough for future works.

Keywords: *Second-Hand Vehicle, Price Forecasting, Data Science, Machine Learning Algorithms, Regression Models*

Classification of Eye Blinks in EEG Signals Using Neural Networks

Emre DANDIL¹, Ayşegül SELVİ¹

¹Bilecik Seyh Edebali University, Bilecik, TURKEY

Abstract

The electroencephalography (EEG) device records the activation signals generated by the brain taken over the skull. The structure of the EEG signals comprises a plurality of signals. These signals now have a wide range of fields of operation. Some of these studies are called Brain Computer Interface (BBA). With these systems, different motor and emotion states of the stimulus are given or the stimulus is analyzed from the EEG signal and different devices and applications are operated and controlled. In such systems, distorting signals generated by the human body or caused by external influences must be filtered before data can be processed. In this study, the classification of the blink in the EEG data in the EEG signals recorded with the EPOC + device of Emotiv was performed. In the developed system, clipping of EEG data in Artificial Neural Networks is provided with the application prepared.

Keywords: EEG, Blinking, Classification, Artificial Neural Networks

EEG Sinyallerindeki Göz Kırpmalarının Yapay Sinir Ağları ile Sınıflandırılması

Özet

Elektroensefalografi(EEG) cihazı ile kafatası üzerinden alınan beyin tarafından üretilen aktivasyon sinyallerini kayıt altına alınmaktadır. EEG sinyallerinin yapısı itibari ile birçok sinyali içerisinde barındırmaktadır. Bu sinyaller günümüzde geniş bir yelpazede çalışma alanına sahiptir. Bu çalışmaların bir kısmı Beyin bilgisayar Arayüzü (BBA) isimlendirilen çalışmalardan oluşmaktadır. Bu sistemler ile uyarı verilen veya uyarının farklı motor ve duyu durumları EEG sinyali içerisinde analiz edilerek farklı cihaz ve uygulamaların çalıştırılması ve kontrolü sağlanmaktadır. Bu tarz sistemlerde veriler işlenmeden önce insan vücudu tarafından oluşturulan veya dış etkilerden kaynaklı bozucu sinyallerin filtrelenmesi gerekmektedir. Bu çalışmada, Emotiv firmasının EPOC+ cihazı ile kayıt altına alınan EEG sinyallerinde göz kırpmalarının EEG verisi içerisinde sınıflandırılması gerçekleştirilmiştir. Geliştirilen sistemde, hazırlanan uygulama ile EEG verileri içerisindeki kırpmalarının Yapay Sinir Ağları ile sınıflandırılması sağlanmıştır.

Anahtar Kelimeler: EEG, Göz Kırpmaları, Sınıflandırma, Yapay Sinir Ağları

The Estimation of Regression Coefficients in Composite Regression Models Using Heuristic Optimization Algorithms

Hande KONŞUK ÜNLÜ¹, Ayten YİĞİTER¹, Yasemin GENÇTÜRK¹

¹*Hacettepe University, Ankara, TURKEY*

Abstract

Composite models are the special case of finite mixture models and frequently used when the data has fat-tailed behavior. These models are formed by combining two or more distributions at specified threshold value(s) by weighting them. In addition to fat-tail behavior of data, there may be heterogeneity between individuals. Therefore, composite regression models are required to build a predictive model capturing the peculiar characteristics of data and individuals' heterogeneity. However, these regression models are quite complex and it is difficult to obtain parameters' estimates as compared to traditional regression models. Instead of traditional optimization methods, using heuristic optimization methods in estimation process could make it easier to estimate the composite regression models' parameters'. The aim of this study is to model Singapore automobile insurance dataset by using Exponential-Pareto, Weibull-Pareto and Lognormal-Pareto composite regression models and to obtain their parameters' estimates via Genetic algorithm and Particle Swarm optimization methods. In addition, a simulation study is conducted to compare the performance of these algorithms.

Keywords: *Composite Regression Models, Genetic Algorithm, Particle Swarm Optimization, Fat-tailed Distributions, Singapore Automobile Insurance*

The Investigation of Factors Affecting Level of Happiness Using Data Mining Methods

Hande KONŞUK ÜNLÜ¹, Gamze AKTUNA¹, Serpil AKTAŞ ALTUNAY¹

¹*Hacettepe University, Ankara, TURKEY*

Abstract

Happiness, which is an abstract concept, is associated with many situations such as labor force participation, social cohesion, personal and health care, and the concept of happiness is also investigated in life satisfaction researches. Determining which social, health and social factors are more effective on happiness may also be guiding for improving the quality of life. In this study, 'Life Satisfaction Survey – 2017' data conducted by Turkish Statistical Institute were used. The level of happiness of the individuals was taken as dependent variable and the effect of independent variables such as sociodemographic, work life, education, income status, living environment and health status were investigated with data mining methods which were Logistic Regression, Artificial Neural Networks and Naive-Bayes. The performance of the models were evaluated with the classification accuracy and confusion matrix. Since the data set had been obtained by two-stage cluster sampling method, the data were weighted in all analyses.

Keywords: *Happiness, Logistic Regression, Artificial Neural Networks, Naive-Bayes Classifier, Classification*

Very High Resolution Remote Sensing for The Diachronic Study of Beaches in The Central Algerian Region

Karima REMMACHE¹, Nour El Islam BACHARI², Fouzia HOUMA¹

¹*National School of Marine Science and Coastal Planning "ENSSMAL", Algiers, ALGERIA*

²*Houari Boumediene University of Science and Technology "USTHB", Algiers, ALGERIA*

Abstract

Shoreline change and beach evolution are governed by several factors, mainly climate change, coastal hydrodynamics and anthropogenic phenomena. The main objective of this work is to set up a geographical database (GIS) for the monitoring and qualification of beaches on the Algerian metropolitan coast over a time series between 2002 and 2017. The methodology developed to improve the use of Google Earth satellite images compared to download methods is based on three methods either directly via Google Earth, or by other software including Offline Map Maker and Smart GIS. The GIS techniques used to process the downloaded images allowed us to rectify the Google earth images by eliminating the gap between its spacious backgrounds caused by the superposition of the two old and recent dates in order to make a diachronic study of the beaches and a diagnosis on the historical variations of the shoreline. Subsequently, the digitization of the geometry of the beaches allowed us to identify the beaches with large/small reception areas. This work is based on the calculation of the distances of linear evolutions of the coastline and surface budgets, to do this the digital shoreline tool DSAS system analyses integrated under arc gis was used. The results obtained show very remarkable evolutions (Erosion/Accretion) of the shoreline of several beaches on a considerable linear scale and significant anthropogenic pressure by the recharging operations (spurs, breakwaters, ports and protective walls) which are held responsible for the modification of the coastal space and the creation of artificial accretion zones. However, our results clearly show a correspondence between erosion and accretion phenomena on the one hand and coastal hydrodynamic phenomena on the other hand, which are represented by two parameters: the speed and significant wave height for the studied area. In the end, the maps drawn up, each representing a municipality on the Algerian coast, will constitute key information for monitoring, enhancing and qualifying the beaches. Planning this vulnerable area requires integrated management of this coastal area that is difficult to manage.

Keywords: *Coastline, Beaches, Surveillance, GIS, Database, Central Algerian Region*

A Change Detection Framework: Satellite Image Application

Hayet SI SALAH¹, Nour El Islam BACHARI²

¹ *Boumerdes Algeria University "UMBB", Boumerdes, ALGERIA*

² *Houari Boumediene University of Science and Technology "USTHB", Algiers, ALGERIA*

Abstract

Very high resolution satellite images can be transformed to urban plan. If we observe the same area at different times, it is possible to detect changes in space occupation. Many change detection methods have been described in the literature, ranging from simple differencing to machine learning techniques. Faced with this large number of suggested techniques, researchers have tried to classify them into different categories in an attempt to simplify the topic. A number of different categorization schemes have been proposed, based on various dimensions, aspects or features of the techniques considered. Nevertheless, using these schemes to understand the different techniques and to select suitable techniques for a specific remote sensing change detection project remains a hard task for practitioners. In this context, we provide a critical study of proposed categorization schemes in order to extract the main dimensions used for their construction. We proposed any rules algorithms transformed RGB image in any objects(roads, line cost, vegetation, buildings, etc.) these rules applied to RGB Quickbird images taken from two different dates in the area of Boumerdes before and after the Earthquake of Boumerdes in 2003. Experiments have shown very good results in terms of detecting changes (damages roads, buildings and line cost changes).

Keywords: *Change Detection, Framework, Rule-Based Algorithms*

Contribution of Synthetic Cloudiness Images in The Classification of Meteorological Phenomena Using Msg Images

M'Hammed BOUSSIR¹, Nour El Islam BACHARI², Abdelhak RAZAGUI³

¹*Science and Technology Mohamed Boudiaf of Oran University (USTO-MB), Bir El Djir, ALGERIA*

²*Houari Boumediene University of Science and Technology "USTHB", Algiers, ALGERIA*

³*Renewable Energy Development Center Bouzareah, Algiers, ALGERIA*

Abstract

The second-generation Meteosat satellite can observe the Euro-African disk every $\frac{1}{4}$ hours. MSG images are used to observe weather phenomena such as hurricanes, storms and to monitor the evolution of the planet's climate. The objective of our study is the development of a classification method that allows the transformation of raw images into class images of weather phenomena. In the first place we transform the raw images into synthetic cloudiness images that present the pixel rate in clear sky condition. Second, we used the RGB presentation mode by combining the 12 channels to establish rules for the choice of kernels associated with different weather phenomena. Radiometric kernels are used to transform images into classes of phenomena based on minimum radiometric distance decision rules. The rules decision is composed of the radiometric of the images of the channel visible $0,6 \mu\text{m}$ and images IRT $10,8 \mu\text{m}$ respectively, cloudiness, clear sky image, raw image, variance. For each phenomenon a decision rule is established. Combining the different rules an image of the different phenomena is established. The classification realized is compared to other classifications based on RGB images. We correlated the classified images against the reference images and the lowest correlation rate is 90%. In order to achieve our objectives, we have developed an application witch can do different treatments, namely the segmentation by the k-means method, the calculation of the cloudiness, the calculation of the clear and cloudy sky image and finally the classification by the method developed in this study.

Keywords: *Meteosat Second Generation, Digital Images, Cloudiness, Color Composition, Classification*

A Modified Type-2 Fuzzy Clustering Algorithm Using Atanassov's Intuitionistic Fuzzy Sets

M.Bahar BASKIR¹

*¹Department of Mathematics, Faculty of Science,
Bartın University, Bartın, TURKEY*

Abstract

Fuzzy clustering algorithm is one of the major approaches in machine learning. Fuzzy c-means is the well-known and remarkable fuzzy clustering technique. In the literature, from type-1 to type-2, there are several versions of fuzzy c-means algorithm. In this study, a modified fuzzy clustering algorithm is presented. The modified algorithm is based on the bases of the combination of two applicable methods: i) Full type-2 fuzzy system modeling approach and ii) Atanassov's intuitionistic fuzzy sets. The Atanassov's intuitionistic fuzzy sets demonstrate more realistic results owing to its "hesitation degree" description. Besides, this algorithm reduces the computational complexity by means of Full type-2 fuzzy approach. Some numerical examples are given to show the achievement of this modified algorithm.

Keywords: *Full Type-2, FCM, Intuitionistic Fuzzy Sets, Machine Learning*

Sismik Dalganın Yer İindeki Yayılımının Zaman ve Frekans Ortamında Modellenmesi-Modeling Seismic Wave Propagation in Time and Frequency Domain

Bedri SEVİNC¹

¹*Mugla Sıtkı Kocman University, Mugla, TURKEY*

Abstract

In seismic research, mathematical modeling of the propagation of seismic wave according to the layer properties; called synthetic seismogram. The synthetic seismogram is an estimate of what the seismic record will look like for the defined the underground structure. Establishes a relationship between seismic recording characteristics and layer parameters. The accuracy of the synthetic seismogram is related to how successful it is to represent the Earth's interior. In order for the synthetic seismogram to be realistic, complex models including density, lithology, porosity, liquid and gas saturation, intrinsic attenuation, scattering, dispersion, pressure, temperature, permeability, etc., must be calculated. In this study, synthetic seismograms were calculated with MATLAB program in time and frequency domain. Today, the effects of environmental properties on the measured seismic wave are used in underground mining research (especially oil and natural gas) and reserve determination. The mathematical model was used to investigate the changes in seismic wave parameters of porous media with liquid and gas and non-elastic environments.

Keywords: *Synthetic Seismogram, Finite Difference Method, Convolution Method, Wave, Propagation, Anelastic Attenuation, Dispersion*

Fake Account Detection on Twitter Using Machine Learning Algorithms

İrfan KÖSESOY¹

¹*Yalova University, Department of Computer Engineering, Yalova, TURKEY*

Abstract

With the widespread use of internet and mobile technology, the number of users of social media sites such as twitter, Facebook, and Instagram has increased and reached billion-levels. The number of speculative and malicious accounts, automatically created and managed through a software, among social media users are quite high. Twitter bot accounts are created to advertise event and bias within the public opinion by sending malicious message, manipulating hashtags, sponsor public characters etc. In this study, we predict whether an account is fake or not based on the data of twitter user accounts . To do this, supervised machine learning algorithms are used. A data set consisting of 3,474 genuine and 3,351 fake accounts is used in the experiments.

Keywords: *Cyber Security, Fake Accounts, Social Spambots, Machine Learning*

An Investigation Into the Noise Filtering Performance of Quadratic Image Filters*

Süleyman UZUN¹, Devrim AKGÜN²

¹*Bilecik Seyh Edebali University*

²*Sakarya University*

Abstract

Nonlinear filters are of importance in image processing due to nonlinear characteristics of images. One of the alternative of linear filters is Volterra filters which contains polynomial nonlinear terms. Due to its computational load, Volterra filters are usually truncated to second degree terms and this type of filters are usually called as quadratic image filters. Quadratic image filters use all possible second degree of multiplications of a group of pixels which were selected according to window size. In this study, noise filtering characteristics of quadratic image filter are investigated using images with Gaussian noise. The weights of the quadratic filter are trained using optimization algorithm using Gaussian noise added reference images. Experimental results were obtained for various noise levels and are compared with Gaussian filter which is a kind of linear filter. According to the results, quadratic image filter preserves the edges better than Gaussian filter.

Keywords: *Quadratic Image Filter, Noise Removal, Optimization*

* The extended version of this study was published in the journal of Data Science and Applications.

R Programming Language Process Mining Package: BUPAR*

Ufuk ÇELİK¹

¹*Bandirma Onyedi Eylul University, Balikesir, TURKEY*

Abstract

Process mining, which is a sub-branch of business intelligence, aims to reach important information about a process through analyzes performed on event logs. In these analyzes, processes are modeled according to event, time and responsible information. The bottlenecks in the processes are determined according to the variants and frequencies of the events. The more suitable models are determined on the obtained models. Event logs, consisting of processes that occur over a short period of time, generate large data in total. Specially developed software is needed to analyze this data. Developed in the R programming language with its increasing popularity in data mining, the process mining package BUPAR has important features that make a difference in this area. Thanks to the flexible and convenient development of R programming language, the ability to develop code that is not available in other ready-made programs and use different data mining methods makes this package superior to other process mining software. In this study, a process mining analysis using BUPAR package is explained. In the analysis, data preparation, process flows, bottlenecks and process model are exemplified.

Keywords: *Data Mining, Business Intelligence, Process Mining, R Programming*

* The extended version of this study was published in the journal of Data Science and Applications.

A Fashion Image Product Recommendation System using Deep Learning*

Fatih Mehmet YILDIRIM¹, Abdullah KAYA¹, Deniz KILINC²

¹*Morhipo, Istanbul, TURKEY*

²Bakircay University, Izmir, TURKEY

Abstract

This study aims to benchmark and propose a data science application about the real-world applications of machine learning and deep learning techniques (such as Faster R-CNN Resnet 50, Faster R-CNN Resnet 101) by implementing various object detection algorithms and evaluating them in terms of accuracy and performance. We created an experiment and developed a system capable of executing different object detection algorithms and generated models trained on a custom dataset generated by using real product catalog data collected from morhipo.com, an online fashion commerce platform. Extracted data from product images are used in cross-sell product recommendations and the obtained results are promising.

Keywords: Machine Learning, Deep Learning, Object Detection, Recommender Systems

* The extended version of this study was published in the journal of Data Science and Applications.

Estimation of Censored Regression in case of Skew and Bimodal Errors*

Ismail YENILMEZ¹, Yeliz Mert KANTAR¹

¹*Eskisehir Technical University, Eskisehir, TURKEY*

Abstract

The left censored data has become quite common in econometrics literature. The reference value is a specific censorship point. Values above this point can be observed (Left-Censored data). However, values below this point cannot be observed. For each censored (non-observable) values, the censorship point is added in a data set. According to the percentage of censorship, there have been agglomeration at this point. In such data sets, the censorship point is often zero. At this case, the ordinary least squares (OLS) estimates are biased and inconsistent. To solve this problem, Tobin's censored normal regression estimator was proposed. However, several potential misspecifications cause inconsistency for the Tobit. One of these misspecifications is non-normality. In particular, it is the subject of this study that the errors are skewed and bimodal. To cope with skew and bimodal errors for the censored regression, partially adaptive estimators can be used. In this study, we consider the estimator based on the alpha skew logistic (ASLG(α)) distribution for censored data. The considered estimator for the censored regression is evaluated by means of a simulation study designed in different combination of various skew and bimodal errors and sample sizes. The new estimator has small mean square error (MSE) and bias values relative to Tobit and OLS.

Keywords: *Censored Data, Partially Adaptive Estimator, Alpha Skew Logistic Distribution*

* The extended version of this study was published in the journal of Data Science and Applications.

An Optimization Energy Consumption of Sensor Nodes*

Maad M. MIJWIL¹

¹*Baghdad College of Economics Sciences University, Baghdad, IRAQ*

Abstract

Wireless sensor networks (WSNs) can be deployed in a variety of environmental monitoring and information gathering applications. This paper, clarify a significant Structure of parking management using WSN. The energy resource is very important for the sensor nodes and has a direct effectiveness on the sensor nodes life and thus entire WSN. We offer a hybrid communication mode to optimize energy consumption. We compare the energy consumption in the proposed parking management architecture with the three modes of communication; a multi-hop, single hop, and hybrid.

Keywords: *Energy Consumption, Wireless Sensor Networks, Sensor Nodes*

* The extended version of this study was published in the journal of Data Science and Applications.

Using Machine Learning Techniques to Predict the Type of Asthmatic Allergy*

Sevinç ILHAN OMURCA¹, Ekin EKINCI¹, Bengisu ÇAKMAK¹, Selin Gizem ÖZKAN¹

¹*Kocaeli University, Kocaeli, TURKEY*

Abstract

To identify the relationship between the characteristics of patients and allergic diseases such as asthma, allergic rhinitis, atopic dermatitis, and eczema is an important goal in allergy researches. In this study, we aim to develop an intelligent diagnostic assistant to predict the type of an allergic disease automatically by using well-known machine learning algorithms such as Decision Tree, Logistic Regression, Support Vector Machines (SVM), K Nearest Neighbor (kNN) and ensemble classifiers by using these algorithms. An allergic diseases dataset, which is taken from Kocaeli University Research and Application Hospital, is used in experiments. The maximum accuracy rate of the established classifiers in detecting 18 different allergy diagnoses is 77%.

Keywords: *Allergy, Classification, Ensemble Classifiers, Machine Learning*

* The extended version of this study was published in the journal of Data Science and Applications.

Generalized Fuzzy Rough Approximation Spaces Formed by a New Fuzzy Set Valued Function on Canonical Hypergroups*

Canan AKIN¹

¹Giresun University, Giresun, TURKEY

Abstract

In this paper, a generalized product of two fuzzy subsets is given. Then some special set valued mappings on canonical hypergroups are defined and some of their features are examined. Generalized fuzzy rough approximations are formed by these new and special set valued mappings on canonical hypergroups and some properties of these generalized fuzzy rough approximations are investigated.

Keywords: Fuzzy Set Valued Mappings, Generalized Fuzzy Rough Sets, Canonical Hypergroups

* The extended version of this study was published in the journal of Data Science and Applications.

A Method Based on Natural Language Processing for Identify and Correct Errors in Text Data*

Hüseyin POLAT¹, Hayri SEVER², Saadin OYUCU¹, Sükran TEKBAS³

Gazi University, Ankara, TURKEY

Cankaya University, Ankara, TURKEY

EMFA Software Consultancy, Ankara, TURKEY

Abstract

Acoustic model used in automatic speech recognition systems is supported by the language model. Plain text data is used to create language model. Therefore, the accuracy of the plain text data forming the language model is also very important for the training of speech recognition models. In this study, the detection and correction of the spelling errors in the plain texts used to create the Turkish language model were carried out using the natural language processing technique. Firstly, misspelled words have been identified. Lack of characters in a word, excess character, displacement of characters and incorrect words due to error typing of the characters are considered incorrect. The words that were misspelled in the text were determined by morphological analysis. Wrong words are replaced with the correct words. As a result of the tests, it has been determined that the method developed in the process of identifying the wrong words and replacing them with the right words was successful in 93%.

Keywords: *Speech Recognition, Natural Language Processing, Corpus Error, Grammatical Word Error*

* The extended version of this study was published in the journal of Veri Bilimi.

DCC-GARCH-Copula Approach in Modelling Dependency between Stock and Government Bonds*

Emre YILDIRIM¹, Mehmet Ali CENGİZ¹

¹ Ondokuz Mayıs University, Samsun, TURKEY

Abstract

The dependence structure between financial assets constitutes the main source of uncertainties in the economics and financial area. In this study, the dependency structure between Turkey 10-year government bonds and BIST 100 index is modelled via the copula-DCC-GARCH approach. This method gives rather effective results in modeling the features of financial assets such as asymmetry and heavy tail. It also allows for modelling many dependence structures such as tail dependency and non-linear dependencies. In this study first, standardized residuals are obtained by applying the DCC-GARCH method to returns of the government bonds and BIST 100. Then, the dependence structure between returns of related financial variables is modelled by means of static and dynamic copulas. Finally, it is found that static Student t copula models the best dependency between Turkey 10-year government bonds and BIST 100 and Turkey 10-year government bonds prices tend to comovement during the extreme events in the financial markets.

Keywords: *Dependence Modelling, Copula Approach, Stock, Government Bond*

* The extended version of this study was published in the journal of Veri Bilimi.

Classification of Human Activities Using Deep Learning*

İbrahim Ali METİN¹, Bahadır KARASULU¹

¹*Canakkale Onsekiz Mart University, Canakkale, TURKEY*

Abstract

The perception of human activity is to try to understand the action of the human by evaluating the data obtained from the environment in which human beings interact. Human movements are basically; short events, basic activities and complex activities are divided into three main groups. In the study, there are movements from the basic activities group which can be dynamic or static. These motion data to be classified were retrieved from a data set warehouse which was obtained from the accelerometer and gyroscope data of wearable devices. Distinguishing features in classical machine learning are usually determined manually by the user. The more successful this subject is, the greater the overall performance in such an experimental environment with sufficiently large data. Therefore, the improvement of performance by using deep architecture is targeted in various studies. In this study, the data obtained with the sensors were used and the classification was made by means of Long-ShortTerm Memory (UKSB-LSTM), a deep learning architecture model with pre-labeled basic movement and motion transitions. As a result of this classification, high accuracy test results were obtained. Based on the results of this test, the systematic evaluation was made by considering the various performance criteria with the proposed systematic approach. In addition, discussion on the results of the study was also included.

Keywords: Human Activities, Deep Learning, Long-Short Term Memory

* The extended version of this study was published in the journal of Veri Bilimi.

Application of Fiber Bragg Grating Sensors in Helicopter Flight Test Instrumentation*

Ayşenur HATİPOĞLU¹, Ahmet Can GUNAYDIN¹, Kemal FIDANBOYLU¹

¹TUSAŞ, Uludag University R&D Center, Bursa, TURKEY

Abstract

Flight Test Instrumentation (FTI) is responsible for monitoring and recording the performance of aircraft, which is equipped with sensors, under various flight test conditions for the design verification and certification process. In FTI, various test data such as strain, load, pressure, acceleration, vibration and temperature are collected for performance, structural dynamics, usage, reliability and maintenance measurements using many different sensors on the experimental aircraft. In this process, which is also called sensor fusion, data are obtained from different parts of the aircraft such as engine, landing gear and blades. These data are processed and filtered using various algorithms. Electrical and mechanical sensors are generally preferred for data collection in these applications. These conventional sensors require extensive utilization of cables in the systems. The cables cause extra weight on the aircraft and increase cost due to the long installation phase. Besides, temperature, humidity and corrosion have negative effects on traditional sensors. The accuracy and life time of these sensors can be negatively affected by the vibrations of the rotating structures of the aircrafts such as helicopter. The use of fiber optic sensors has recently become widespread as an alternative to electrical and mechanical sensors. Fiber optic sensors are very light, robust, high-precision sensors, they are not affected by electromagnetic interference and do not conduct electric current. They are resistant to harsh environments and capable of multiplexing for remote sensing and sensor fusion. In addition to these advantages, they can detect different parameters such as strain, pressure, temperature, corrosion and vibration on a single cable. A specific type of fiber optic sensors, which are called Fiber Bragg Gratings (FBG), are very thin and have cylindrical structure. Due to their cylindrical geometry, they can be easily integrated into different structures such as composite materials. FBG sensors have multiplexing and dynamic measurement capabilities, and they are suitable for remote sensing. Harness complexity which is caused by electrical sensors does not exist in FBG sensor systems. Because of all these advantages, FBG sensors are preferred for both FTI and Health and Use Monitoring Systems (HUMS). HUMS is a system, which is developed to demonstrate the airworthiness of helicopters, minimizes periodic maintenance, provides real-time monitoring and determinates safety risks in advance. This paper provides an overview of the recent advances, research and applications of FBG sensors and big data processing in FTI and HUMS.

Keywords: *Data Gathering, Fiber Bragg Grating (FBG) and Fiber Optic Sensors, Flight Test Instrumentation (FTI), Health and Usage Monitoring Systems (HUMS), Helicopter, Sensor Fusion*

* The extended version of this study was published in the journal of Veri Bilimi.

Comparison of Methods Cox Regression and Random Forests in Survival Analysis*

Pelin AKIN¹, Yüksel TERZİ¹

¹*Ondokuz Mayıs University, Samsun, TURKEY*

Abstract

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. Survival data are usually censored. Observations are called censored when the information about their survival time is incomplete. Cox regression, Kaplan-Meier techniques are generally applied to model this type of data. An alternative way to these methods is to use machine learning techniques. These methods are becoming increasingly popular largely due to their ability to handle nonlinear relationships, higher-order interactions, and high-dimensional covariates. In this study, patients who have been in Chest diseases service in the Hospital of Ondokuz Mayıs University. In the data set, the smote sampling method was used because the categories of the output variable were unbalanced. In this study, Random Forest and Cox Regression were used to determine the risk of each patient in leukemia. These two methods are compared to the C-index. According to the result, it was found that random forest's C-index higher than Cox regression so could be an alternative method for survival analysis.

Keywords: *Random Forest, Smote Resampling Method, Survival Analysis*

* The extended version of this study was published in the journal of Veri Bilimi.

Supplier Selection Using Fuzzy Analytic Hierarchy Process Method: Application of an Automotive Supply Industry Company*

Çağın KARABIÇAK¹, Burcu ÖZCAN¹, Mehlika KOCABAS AKAY¹

¹*Kocaeli University, Kocaeli, TURKEY*

Abstract

In a globally competitive business environment, decision-making is vital for companies because of the rapid development of technology and the rapid change of society's lifestyle. For this reason, companies need to use scientific methods when making a future decision in order to maintain their competitiveness. The methods used to decide the problems in which there are multiple criteria and where there is not a single optimum solution are generally referred to as "Multi Criteria Decision Making Methods (MCDM). When the methods used in complex decision making processes are examined; It is seen that many of the developed methods are considered to be MCDM. When these methods are compared, each has several advantages and weaknesses. Analytical Hierarchy Process (AHP) method is the most common method in multi-criteria decision making because it is the most common method in literature. In the classical sense, the criteria are expressed by using precise expressions. However, real-life problems may not be expressed in absolute values and contain difficult uncertainty or subjective judgments. This situation may occur especially in qualitative information, incomplete information and information that cannot be obtained. Language variables are used to express uncertainties and subjectivities in decision-making preferences, and fuzzy logic / numbers are used to model these variables. The FAHP method, which is an improved form of classical AHP and uses fuzzy numbers, enables human opinions and convictions to be more effective in the decision-making process. In this paper, the supplier selection process in a company operating in automotive industry has been examined and FAHP method is used when selecting from three candidate mold companies. The supplier selection has a complex and fuzzy structure that contains many criteria in the selection process. When the literature was examined, among the many FAHP methods encountered, Chang's Extended Analysis Method was preferred to be used in the related problem. The criteria to be used in the evaluation of the mold supplier candidates were determined by taking into consideration the purchasing policies of the enterprise and the general supplier selection criteria. The company, which has the best value in the evaluation made according to the four main and fourteen sub-criteria determined as quality, price / cost, delivery and flexibility, was found suitable for the production, modification and transfer of the molds.

Keywords: *Automotive Sub Industry, Multi-criteria Decision Making, Fuzzy AHP, Supplier Selection*

* The extended version of this study was published in the journal of Veri Bilimi.

Investigation of the Efficiency of Solid Waste Management in Level 1 Regional Municipalities with Data Envelopment Analysis*

Ebru ŞAŞMAZ¹, Selen AVCI¹, Zerrin ALADAG¹

¹*Kocaeli University, Kocaeli, TURKEY*

Abstract

This Solid wastes have become a problem that cannot be ignored with the development of industry, advancement of technology and changing consumption habits. Municipalities have the responsibility of collecting, storing and disposing of solid wastes. In this study, the solid waste management efficiencies of 12 regions in Turkey (Istanbul, Western Marmara, Aegean, Eastern Marmara, Western Anatolia, Mediterranean, Central Anatolia, Western Black Sea, Eastern Black Sea, Northeast Anatolia, Middle East Anadolu and Southeastern Anatolia) were examined by using Data Envelopment Analysis (DEA) with two inputs and three outputs. The inputs used in the study are the average amount of waste per capita, the number of municipalities in the region and the municipal environmental expenditures. Outputs were selected as municipal environmental revenues and the amount of disposed waste in landfills. According to the Charnes-Cooper-Rhodes (CCR) model of DEA, Istanbul and Southeast Anatolia regions; according to the Banker-Charnes-Cooper (BCC) model, Mediterranean, Istanbul, Southeast Anatolia, Western Anatolia, Eastern Black Sea, Central Anatolia and Northeast Anatolia regions were found to be efficient. Then, both the efficient and inefficient units were ranked with the Super Efficiency (SE) model, another model of DEA, and the results were compared with the results obtained from TOPSIS (Technique for Order Preference by Similarity Ideal Solution) method. Finally, input-output changes for inefficient regions were analyzed and potential improvement suggestions were presented.

Keywords: *Data Envelopment Analysis, Solid Waste Management, Super Efficiency, TOPSIS*

* The extended version of this study was published in the journal of Veri Bilimi.

Development of Image Filter Based on Quadratic Filter for Enhancement of Brain MR Scans*

Süleyman UZUN¹, Emre DANDIL¹, Esin KARAGÖZ²

¹ *Bilecik Seyh Edebali University, Faculty of Engineering, Computer Engineering, Istanbul, TURKEY*

² *Bilecik Seyh Edebali University, Institute of Science, Computer Engineering, Istanbul, TURKEY*

Abstract

Quadratic-based image filters are determined from the second order of Volterra filters, and such filters are the best known of the nonlinear filters class. They produce better results than linear filters in the removal of noise with non-linear characteristics. In this study, a new design based on Quadratic-based image filter has been realized in order to improve brain MR images. The mask weights of the quadratic-based image filters were calculated by genetic algorithms for each brain MRI image used in the study. The brain MR images were filtered using the computed filter weights and the improved images were obtained. The same brain MR images were filtered using a median filter and the obtained image quality was compared with quadratic-based image filters. According to the results obtained in the experimental studies, it was observed that the quadratic-based image filter detected the tumor regions on the brain MR images and the edges of these regions were better protected.

Keywords: *Quadratic Image Filter, MR Images, Remove Noise, Brain Tumors, Genetic Algorithms, Image Improvement*

* The extended version of this study was published in the journal of Veri Bilimi.

Computer Aided Analysis of Brain MRI Images*

Emre DANDIL¹, Gökhan UÇAR¹, Ali AYDOĞDUOĞLU¹

Bilecik Seyh Edebali University, Bilecik, TURKEY

Abstract

Nowadays, with the increase of urbanization, nutritional habits, stress etc. different diseases start to emerge and parallel to this, diseases developing in the brain are increasing in recent years. Magnetic resonance imaging is the first and most commonly used method for the diagnosis of brain diseases. In cases where the diagnosis cannot be made by magnetic resonance imaging, other laboratory tests are performed, but physicians may have difficulty in diagnosing the disease. In this study, a computer aided analysis system software has been developed to help physicians facilitate diagnosis of brain diseases. After pre-filtering the MR images of the patients and separating the brain skull, the region of interest (ROI) was removed and the areas that may be pathological phenomena were marked and the dimensions of these areas were calculated. Finally, the similarities were calculated by comparing the fields marked with the computer aided analysis system and the fields marked by hand.

Keywords: Magnetic Resonance Imaging, Computer Aided Analysis System, Pathology, Region of Interest

* The extended version of this study was published in the journal of Veri Bilimi.