Use Cases in Big Data Software and Analytics

Vol. 1, Fall 2017

Bloomington, Indiana

Tuesday 31st October, 2017, 19:48

Editor:
Gregor von Laszewski
Department of Intelligent Systems
Engeneering
Indiana University
`laszewski@gmail.com`

Contents

Chapter 1

Preface

1.1 List of Papers

Name	HID	Title
hid101	Huiyi Chen	Benchmarking a BigData Docker deployment
hid102	Dianprakasa, Arif	Benchmarking a BigData Docker deployment
hid104	Jones, Gabriel	Big Data = Big Bias? Ethical Challenges of Big Data
hid105	Lipe-Melton, Josh	Benchmarking a BigData Docker deployment
hid106	Qiaoyi Liu	Benchmarking a BigData Docker deployment
hid107	Ni,Juan	Benchmarking a BigData Docker deployment
hid109	Shiqi Shen	Big Data and Application in Amazon
hid111	Lewis, Derek	Benchmarking a BigData Docker deployment
hid201	Arnav, Arnav	None
hid202	Himani Bhatt	This is my paper about the other abc
hid203	error: yaml	This is my paper about the other abc
hid204	Chaturvedi, Dhawal	Big Data and Support Vector Machines
hid205	Chaudhary, Mrunal L	This is my paper about the other abc
hid208	Devineni, Jyothi Pranavi	Algorithms for Big Data Analysis
hid209	Han, Wenxuan	Clustering Algorithms in Big Data Analysis
hid210	error: yaml	Clustering Algorithms in Big Data Analysis
hid211	Ajinkya Khamkar	Machine learning optimizations for big data
hid212	Kumar, Saurabh	Not yet decided
hid213	Liu, Yuchen	Big Data and Face Identification
hid214	Lu, Junjie	This is my paper about the other abc
hid215	Mallala, Bharat	to be decided
hid216	Millard, Mathew	n/a
hid218	Niu, Geng	This is my paper about the other abc
hid219	Syam Sundar Herle Parampali Sreenath	Benchmarking a BigData Docker deployment
hid224	Rawat, Neha	Big Data Applications in the Energy and Utilities Sector
hid225	Schwartzer, Matthew	...
hid228	Swargam, Prashanth	TBD
hid229	ZhiCheng Zhu	TBD
hid230	YuanMing Huang	Big data with natural language processing
hid231	Vegi, Karthik	Using Big Data to Battle Air Pollution
hid232	Rahul Velayutham	This is my paper about the other abc
hid233	Wang, Jiaan	Big Data Applications in Virtual Assistants

hid234	Weixuan Wang	Big Data and Edge Computing in Health Informatics for People with Disabilities.
hid235	Wu, Yujie	Big Data
hid236	Yang Weipeng	Benchmarking a BigData Docker deployment
hid237	Ahmed, Tousif	Benchmarking a BigData Docker deployment
hid301	Arora, Gagan	Prediction of psychological traits based on Big Data classification of associated social media footprints
hid302	Sushant Athaley	Hadoop and MongoDB in support of Big Data Applications and Analytics
hid304	Ricky Carmickle	Big Data and Analytics in Deep Space Telemetry and Navigation
hid305	Andres Castro Benavides	Big Data applied to zoning and city planning.
hid306	Cheruvu, Murali	Why Deep Learning matters in IoT Data Analytics?
hid308	Pravin Deshmukh	Parallel Computing and Big Data
hid309	error: yaml	Parallel Computing and Big Data
hid310	error: yaml	Parallel Computing and Big Data
hid311	Durbin, Matthew	Benchmarking a BigData Docker deployment
hid312	Neil Eliason	To be decided
hid313	Tiffany Fabianac	Big Data Applications in Laboratories
hid314	Fadnavis, Sarang	Benchmarking a BigData Docker deployment
hid315	Garner, Jeffry	Big Data Opportunity and Challenges with Smart Helmets
hid316	Robert Gasiewicz	Big Data on IoT Smart Refrigerators
hid318	Irey, Ryan	Benchmarking a BigData Docker deployment
hid319	Mani Kumar Kagita	Mini Project: ESP8266 and Raspberry PI Robot Car
hid320	Elena Kirzhner	This is my paper about Big Data Analytics and Applications in Sustainable Fish Breeding
hid321	Knapp, William	Benchmarking a BigData Docker deployment
hid323	Uma M Kugan	None
hid324	Ashok Kuppuraj	TBD
hid325	J. Robert Langlois	The importance of data sharing and the replication of the sciences
hid326	Mahendrakar, Mohan	Benchmarking a BigData Docker deployment
hid327	Marks, Paul	The Impact of Self-Driving Cars on the Economy
hid328	Dhanya Mathew	Big data analytics in data center network monitoring
hid329	Ashley Miller	Big Data Analytics and the Impact on Personal Privacy
hid330	Janaki Mudvari Khatiwada	MQTT for Big Data and Edge Computing
hid331	Tyler Peterson	Big Data Applications in Using Neural Networks for Medical Image Analysis
hid332	Judy Phillips	Big Data Analytics in Developing Countries
hid333	Anil Ravi	Nartural language processing (NLP) for speech analysis and voice recognition
hid334	Peter Russell	Advancements in Drone Technology for the US Military
hid335	Sean Shiverick	Big Health Data from Wearable Electronic Sensors (WES) and the Treatment of Opioid Addiction
hid336	Jordan Simmons	Big Data Analysis for Computer Network Defense
hid337	Ashok Reddy Singam	Natural Language Processing (NLP) to analyze human speech data
hid338	Sriramulu, Anand	Benchmarking a BigData Docker deployment
hid339	Hady Sylla	Benchmarking a BigData Docker deployment
hid340	Tim Thompson	Big data on the blockchain? Distributed networks and large-scale analytics
hid341	Tibenkana, Jacob	This is my paper about the other abc
hid342	Udoyen, Nsikan	Still under consideration

hid343	Usifo, Borga	None
hid345	Wood, Ross	Big Data Analytics and influence on althetics.
hid346	Zachary Meier	This is my paper about the other abc
hid347	Jeramy Townsley	Sociological Methods of Big Data
hid348	Budhaditya Roy	Security aspect of NOSQL database in Big Data Applications

Using Big Data to Battle Air Pollution

Karthik Vegi
Indiana University Bloomington
2619 East 2nd Street, Apt 11
Bloomington, IN 47401, USA
kvegi@iu.com

ABSTRACT

We have come a long way from the stone age to build large scale industries, big cities, bullet trains, and a booming automobile industry. Technological and industrial advances are making our cities smarter by the day and yet a nagging side-effect is air pollution. Air pollution is not only creating local health hazards like respiratory and heart problems, but also directly leading to an increase in temperatures and contributing to global warming. We show how the advances in *Big Data*, *Cloud Computing*, and *Internet Of Devices* can be used to combat air pollution.

KEYWORDS

i523, hid231, big data, environment, air pollution, global warming

1 INTRODUCTION

Big Data is playing a crucial role in building a smarter planet. Each and every action that we take leaves a digital footprint. Big Data is lending a great helping hand to crunch this data and make smarter decisions. "*Big Data* is at the heart of the smart revolution. It is already completely transforming the way we live, find love, cure cancer, conduct science, improve performance, run cities, and countries and operate business" [**?**].

REFERENCES

Sociological Methods of Big Data

Jeramy Townsley
IUPUI
425 University Ave
Indianapolis, Indiana 46202
jtownsle@indiana.edu

ABSTRACT

KEYWORDS

i523, hid347, social sciences, big data, methods, data mining, regression

1 INTRODUCTION

2 CONCLUSION

REFERENCES

Big Data Analysis for Computer Network Defense

Jordan Simmons
Indiana University
Smith Research Center
Bloomington, IN 47408, USA
jomsimm@iu.edu

ABSTRACT

Computer security threats and attacks are constantly evolving. Everyday, hackers are creating new techniques to bypass network security for the purpose of malicious attacks. To keep up with the changing intrusion technologies, the technologies that defend these attacks need to constantly evolve also. Modern day technologies use deep learning techniques to monitor network activity, and detect malicious code. We will provide an overview of network security and modern technologies being used to protect computer systems and networks.

KEYWORDS

i523,HID336, Computer Network Security, Big Data Analysis, Deep Learning, Intrusion Detection Systems,

1 INTRODUCTION

Everyday a different computer network is being breached with the intent to cause harm to the system or to steal valuable data. Computer hackers are constantly creating new ways to evade network security and create malicious code that can not be detected by security systems. As malicious technologies continue to advance, the technologies that defend against these technologies need to adapt with these advances. The problem with computer network defence is that the technologies used to breach systems constantly change. Once a solution is created to defend a technology, a new malicious technology could be created the next day. Today many security specialist are using deep learning technologies to monitor network intrusions, and detect malicious code. In order to better understand computer network defense, an overview of modern attacks, network data collection processes, and the technologies used to analyze network data is provided.

2 DATA COLLECTION

2.1 Network Intrusion Data Collection

2.2 Malware Data Collection

3 DEEP LEARNING FOR NETWORK INTRUSIONS

4 DEEP LEARNING ON MALWARE

5 CONCLUSION

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use "quotes" , instead of " "

To emphasize a word, use *emphasize* and not "quote"

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬁgure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

Big Data Applications in Virtual Assistants

Jiaan Wang
Indiana University Bloomington
3209 E 10 St
Bloomington, IN 47408
jervwang@indiana.edu

ABSTRACT

This paper provides

KEYWORDS

i523, HID233, Big data, Virtual Assistants, Artificial intelligence

1 INTRODUCTION

Put here an introduction about your topic. We just need one sample refernce so the paper compiles in LaTeX so we put it here [11] [13] [3] [5] [8] [12] [7] [6] [10] [2] [9] [4] [1].

2 FIGURES

3 LONG EXAMPLE

4 CONCLUSION

Put here an conclusion. Conclusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

[1] Ethan Baron. 2017. One bot to rule them all? Not likely, with Apple, Google, Amazon and Microsoft virtual assistants. Web Page. (Feb. 2017). http://www.mercurynews.com/2017/02/06/one-bot-to-rule-them-all-not-likely-with-apple-google-amazon-and-microsoft-virtual-assistants/ HID: 233, Accessed: 2017-10-24.

[2] Clint Boulton. 2016. Slack CEO describes 'Holy Grail' of virtual assistants. Web Page. (Oct. 2016). https://www.cio.com/article/3131536/collaboration/slack-ceo-describes-holy-grail-of-virtual-assistants.html HID: 233, Accessed: 2017-10-24.

[3] Mike Elgan. 2016. These three virtual assistants point the way to the future. Web Page. (June 2016). https://www.computerworld.com/article/3078829/artificial-intelligence/these-three-virtual-assistants-point-the-way-to-the-future.html HID: 233, Accessed: 2017-10-18.

[4] Darrell Etherington. 2014. The Virtual Assistant Could Be The Next Interpreter Of Enterprise Data, Starting With Google Now. Web Page. (Aug. 2014). https://techcrunch.com/2014/08/13/the-virtual-assistant-could-be-the-next-interpreter-of-enterprise-data-starting-with-google-now/ HID: 233, Accessed: 2017-10-24.

[5] Lars Hard. 2014. The Disruptive Potential of Artificial Intelligence Applications. Web Page. (Jan. 2014). http://data-informed.com/disruptive-potential-artificial-intelligence-applications/ HID: 233, Accessed: 2017-10-18.

[6] Eran Kinsbruner. 2017. Building the virtual assistant everyone wants. Web Page. (July 2017). https://www.infoworld.com/article/3210488/machine-learning/building-the-virtual-assistant-everyone-wants.html HID: 233, Accessed: 2017-10-24.

[7] Rob Marvin. 2017. What Are Virtual Assistants and What Can You Do With Them? Web Page. (June 2017). https://www.pcmag.com/article/354371/what-are-virtual-assistants-and-what-can-you-do-with-them HID: 233, Accessed: 2017-10-24.

[8] Susanne Mueller. 2016. Rhiza Launches Rhizabot, First Virtual Assistant for Analytics. Web Page. (Aug. 2016). http://rhiza.com/2016/08/10/rhiza-launches-rhizabot-first-virtual-assistant-analytics/ HID: 233, Accessed: 2017-10-18.

[9] Tom Simonite. 2016. How Alexa, Siri, and Google Assistant Will Make Money Off You. Web Page. (May 2016). https://www.technologyreview.com/s/601583/how-alexa-siri-and-google-assistant-will-make-money-off-you/ HID: 233, Accessed: 2017-10-24.

[10] Anubhav Srivastava. 2016. Why the virtual assistants market is on the upswing? Web Page. (July 2016). http://thinkbigdata.in/virtual-assistants-market-upswing/ HID: 233, Accessed: 2017-10-24.

[11] David Tal. 2015. Forecast — Rise of the big data-powered virtual assistants: Future of the Internet P3. Web page. (Nov. 2015). http://www.quantumrun.com/prediction/rise-big-data-powered-virtual-assistants-future-internet-p3 HID: 233, Accessed: 2017-10-18.

[12] Spotfire Blogging Team. 2012. Meet Your Companyfis New Virtual Assistant fi?! Big Data. Web Page. (May 2012). https://www.tibco.com/blog/2012/05/11/meet-your-companys-new-virtual-assistant-big-data/ HID: 233, Accessed: 2017-10-18.

[13] Richard Waters. 2015. Artificial intelligence: A virtual assistant for life. Web page. (Feb. 2015). https://www.ft.com/content/4f2f97ea-b8ec-11e4-b8e6-00144feab7de?mhq5j=e5 HID: 233, Accessed: 2017-10-18.

We include an appendix with common issues that we see when students submit papers.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To do this change TODO with DONE.

Why Deep Learning matters in IoT Data Analytics?

Murali Cheruvu
Indiana University
3209 E 10th St
Bloomington, Indiana 47408
mcheruvu@iu.edu

ABSTRACT

The Deep Learning is unique in all machine learning algorithms to analyze supervised and unsupervised datasets. Big Data challenges, such as high volumes, multi-dimensionality and feature engineering, are well addressed using Deep Learning algorithms. Deep Leaning, with Edge and distributed Mesh computing, is best suited to handle IoT Analytics from millions of sensors producing petabytes of time-series data.

KEYWORDS

i523, hid306, IoT, Deep Learning, Big Data Analytics

1 INTRODUCTION

Supervised machine learning algorithms: decision trees, linear regression, Support Vector Machines (SVMs), Naive Bayes, neural networks, etc. are popular for classification and regression problems by analyzing labeled training data. K-means clustering algorithms are good for unsupervised datasets to categorize based on the identified patterns in unlabeled data. While there are so many factors - nature of the domain, sample size of the dataset and number of attributes defining characteristics of the data - decide which machine learning algorithm works better, Deep Learning algorithms are, getting greater traction, addressing complex analytics tasks, including high-dimensionality and automatic creation of new features from existing complex hierarchical features, very well.

2 NEURAL NETWORKS

Neural Networks are inspired by human brain, the way they solve complex problems. *Perceptron*, the first generation neural network, created a simple mathematical model, mimicking neuron - the basic unit of the brain, by taking several binary inputs and produced single binary output. *Sigmoid Neuron* improved learning by giving some *weightage* to the input based on importance of the corresponding input to the output so that tiny changes in the output due to the minor adjustments in the input weights (or biases) can be measured effectively. Neural Network is, a *directed graph*, organized by layers and layers are created by number of interconnected nodes (or neurons). Every node in a layer is connected with all the nodes from the previous layer; there will no interaction of nodes within a layer. As shown in Figure (1), a typical Neural Network contains three layers: input (left), hidden (middle) and output (right) [3]. The middle layer is called *hidden* only because the nodes of this layer are neither an input nor an output but the actual processing happen in the hidden layer. As data passes through layer by layer, each node acts as an *activation function* to process the input. The performance of a Neural Network is measured using *cost or error function* and the dependent input *weight* variables. *Forward-propagation* and *back-propagation* are two techniques, neural network uses repeatedly until all the input variables are adjusted or calibrated to predict accurate output. During, forward-propagation, information moves in forward direction and passes through all the layers by applying certain weights to the input parameters. *Back-propagation* method minimizes the error in the *weights* by applying an algorithm called *gradient descent* at each iteration step.

[Figure 1 about here.]

3 DEEP LEARNING

Deep Learning is an advanced neural network, with multiple hidden layers (thousands or even more deep), that can work well with supervised (labeled) and unsupervised (unlabeled) datasets. Applications, such as speech, image and behavior patterns, having complex relationships in large-set of attributes, are best suited for Deep Learning Neural Networks. Deep Learning vectorizes the input and converts it into output vector space by decomposing complex geometric and polynomial equations into a series of simple transformations. These transformations go through neuron activation functions at each layer parameterized by input weights. For it to be effective, the cost function of the neural network must guarantee two mathematical properties: *continuity* and *differentiability*.

[Figure 2 about here.]

3.1 Feature Engineering

The dataset with too many dimensions, also known as attributes or features, create large sparsity and make it difficult to process. *Curse of dimensionality* is a scenario where the value added by the dimensions is much smaller in comparison to the processing cost. However, in certain applications, such as face recognition and patient electronic medical records, the complexity created by multiple dimensions might add value to the context. *Feature Engineering* is an exploratory analysis to identify the features that collectively contribute to better predictive modeling by removing irrelevant features and creating new features, using the training information to identify the patterns, from existing interrelated features [6]. *Principal Component Analysis* (PCA) is a technique to analyze the interdependency among the features and keep only the principal, most relevant, features with minimum loss in the model. With enough training, Deep Learning makes neurons learn new features themselves, in an unsupervised manner, from existing features distributed in several hidden layers. *Stacked Autoencoder* (AE) is, a Deep Belief Network algorithm, to create advanced predictive models for large datasets having thousands or even millions of dimensions, automatically, with complex hierarchical attributes in non-linear fashion for simpler computing. Though AE is sophisticated, it is very difficult to understand the algorithm logic and so unable to reuse the learnings from the modeling to other systems.

3.2 Deep Neural Networks

Convolutional Neural Network (CNN) is a deep feedforward network, also called multilayer perceptron (MLP), consists of (1) convolutional layers - to identify the features using weights and biases, followed by (2) fully connected layers - where each neuron is connected from all the neurons of previous layers - to provide non-linearity, sub-sampling or max-pooling, performance and control data overfitting [2]. CNN is used in image and voice recognition applications by effectively using multiples copies of same neuron and reusing group of neurons in several places to make them *modular*. CNNs are constrained by *fixed-size* vectorized inputs and outputs. *Recursive Neural Network* (RNN) is, another type of Deep Learning, that uses same shared feature weights recursively for processing sequential data, emitted by sensors or the way spoken words are processed in natural language processing (NLP), to produce arbitrary size input and output vectors. RNN uses a technique called *loop*, where multiple copies of the same chunk of network (module), each passing a message to the next, to persist the information. Long Short Term Memory (LSTM) is an advanced RNN to learn and remember *longer* sequences by composing series of repeated modules of neural network and a concept called *cell state*, a memory unit, to memorize the learning by adding and removing information using *input*, *output* and *forget* gates, in a regularized fashion while data flows through the layers [9]. The Convolutional and Recursive Neural Networks can complement each other to produce better and effective models where problem space has both - hierarchical features and temporal data. Deep Learning can also work well with related *Reinforcement Learning* algorithms where the focus is on how to maximize the learning based on rewards and punishments.

[Figure 3 about here.]

[Figure 4 about here.]

4 IOT DATA ANALYTICS

Internet of Things (IoT) is getting lots of traction, due to the massive volumes of data, making it *Big Data*; however, business needs to convert this data into *information* whether to monitor and control the devices or to analyze the sensor data for betterment. Time series data has non-stationary time aspects collected at certain intervals over a short period of time and correlate this sequence of data with past or future sequences. Stock prices and IoT sensor data are examples of time series data. *InfluxDB*, an open source time series database, is offering high write performance, data compaction through down-sampling and automatic deletion of expired old time series data, to address IoT data storage challenges [5].

4.1 Complexity

Unique traits of IoT data, such as noise, high dimensionality and high streaming of time-series data in real-time, make it challenging to process using traditional machine learning algorithms [10]. Autoregressive Moving Average Model (ARIMA), converts time-series from non-stationary into stationary, but only for short-time predictions. Deep Learning, using LSTM, can detect anomalies in the IoT Data and train time series data very well. Deep Learning algorithms involve complex mathematics - geometry, matrix algebra, differential calculus, statistics and probability, and intensive distributed computing to train the massive amounts of sensor data.

4.2 Scalability

Deep Learning, by design, allows parallel programming, as each module - with all the dependencies among neurons - can run independently and parallelly from other modules within the network. Using Graphics Process Unit (GPU), module networks can achieve parallel programming without needing much of Central Processing Unit (CPU) allocation. Though GPU is intended for graphical processing, it works efficiently to run thousands of small mathematical functions, such as matrix multiplications, in parallel. Cloud computing and Edge Analytics offer flexible scale out options, using virtualization and containerization, for distributed processing. Sophisticated algorithms and distributed computing make Deep Learning scale and perform well to process huge datasets.

4.3 Case Study

Hewlett Packard (HP) Labs has given a presentation of their experiments to check how effectively Deep Learning algorithms can be applied for IoT Sensor Data Analytics. Sample data - vision, speech, text and sensor data such as signals, have been collected from scripted video and accelerometer from 52 subjects on average 20 minutes of activity recognition per subject - 12,000 measurements per minute per person with 16 classifications, such as walk to bed, enter bed, lie down, roll left, roll right and speak. They have analyzed and trained the sample data using various machine learning algorithms including Support Vector Machines (SVMs), Decision Trees and traditional Neural Networks; compared the results with Recurrent, Deep Learning, Neural Network. Deep Learning showed 95% or more accuracy in various scenarios, performed much better than all the other algorithms, without sophisticated feature engineering. However, Deep Learning algorithms were slow and expensive for results to converge as the sample dataset is huge with lots of instances (10^6-10^9) and very large number of features (>10^6). They have concluded the presentation with scale-out hardware options using CPU/GPU clusters and futuristic Edge Analytics and distributed Mesh Computing alternatives for better scalability and performance [11].

5 CONCLUSION

In contrast to traditional machine learning solutions, Deep Learning not only scales well with high volumes of input data but also facilitates in automatic decomposition of complex data representations of unsupervised and uncategorized data. Automatic discovery of new features, from convolutional or recurrent neural networks, makes Deep Learning predominant among all machine learning algorithms. It is very difficult to understand fuzzy and complex logic of Deep Learning, perhaps, more adoption helps getting better handle at them. Deep Learning algorithms need deep research in validating the process of advanced Big Data Analytics tasks, such as IoT sensor time-series data, semantic learning, scalability, data tagging and reliability of the predictive models without extreme generalization.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski and the Teaching Assistants for their support and valuable suggestions.

REFERENCES

[1] Mark Chang. 2016. Applied Deep Learning 11/03 Convolutional Neural Networks. (Oct. 2016). https://www.slideshare.net/ckmarkohchang/applied-deep-learning-1103-convolutional-neural-networks

[2] Christopher Olah. 2014. Conv Nets: A Modular Perspective. (July 2014). http://colah.github.io/posts/2014-07-Conv-Nets-Modular/

[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org

[4] Vikas Gupta. 2017. Understanding Feedforward Neural Networks. (Oct. 2017). https://www.learnopencv.com/understanding-feedforward-neural-networks/

[5] Influx. [n. d.]. *InfluxDB is the Time Series Database in the TICK stack.* Technical Report. Influx. https://www.influxdata.com/time-series-platform/influxdb/

[6] Jason Brownlee. 2014. Discover Feature Engineering, How to Engineer Features and How to Get Good at It. (Sept. 2014). https://machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. (May 2015). http://www.nature.com/nature/journal/v521/n7553/fig_tab/nature14539_F5.html

[8] Nicholas Leonard. 2016. Language modeling a billion words. (July 2016). http://torch.ch/blog/2016/07/25/nce.html

[9] Christopher Olah. 2015. Understanding LSTM Networks. (Aug. 2015). http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[10] Rajesh Sampathkumar. 2016. Time Series Analysis of Sensor Data. (Aug. 2016). http://www.thedatateam.in/time-series-analysis-of-sensor-data/

[11] Natalia Vassilieva. 2016. *Sense Making in an IOT World: Sensor Data Analysis with Deep Learning*. Technical Report. Hewlett Packard Labs. http://on-demand.gputechconf.com/gtc/2016/presentation/s6773-natalia-vassilieva-sensor-data-analysis.pdf

List of Figures

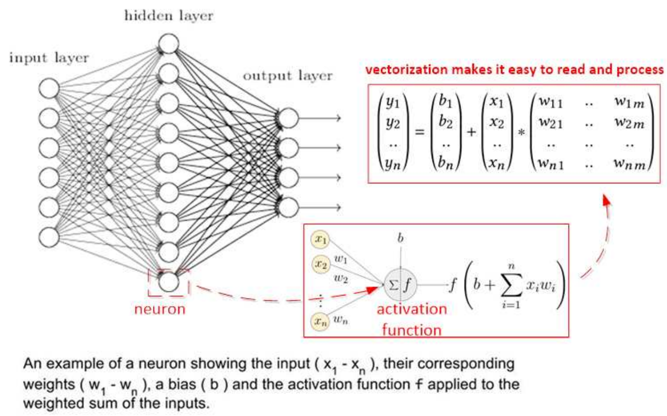


Figure 1: Simple Neural Network [3, 4].

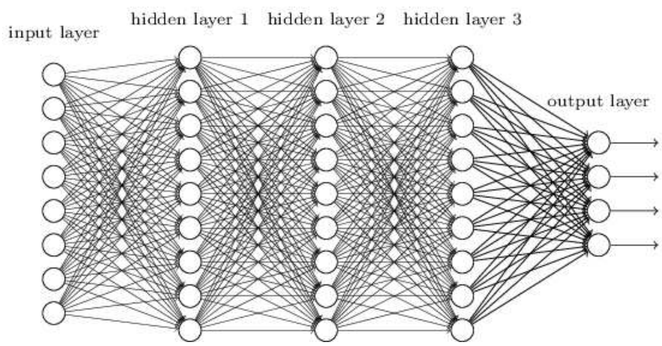


Figure 2: Deep Neural Network with three hidden layers [3].

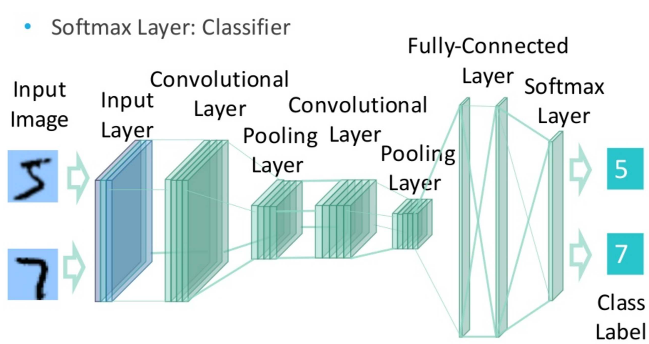


Figure 3: Sample Convolutional Neural Network [1].

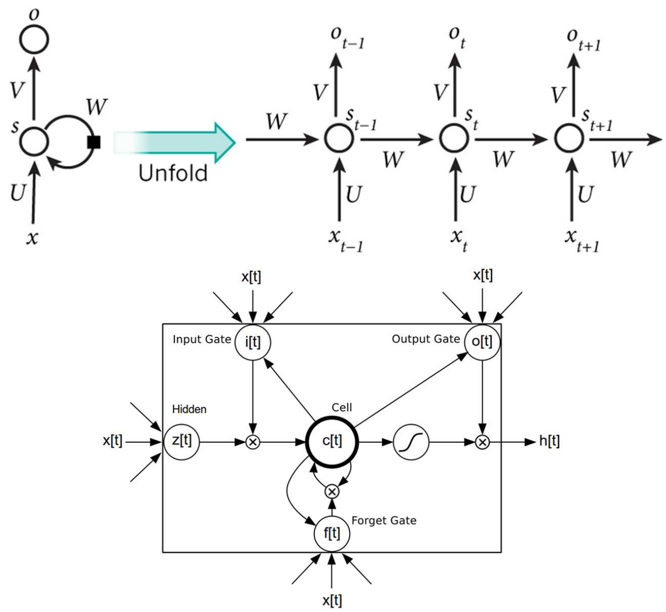


Figure 4: Recursive Neural Network Loop and LSTM Cell State [7, 8].

Big Data for Edge Computing

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Gregor von Laszewski
Indiana University
Smith Research Center
Bloomington, IN 47408, USA
laszewski@gmail.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

Big Data, Edge Computing i523

1 INTRODUCTION

Put here an introduction about your topic. We just need one sample refernce so the paper compiles in LaTeX so we put it here [**?**].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 CONCLUSION

Put here an conclusion. Conlcusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below.* Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show.* It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬁgure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named ”images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

2

List of Figures

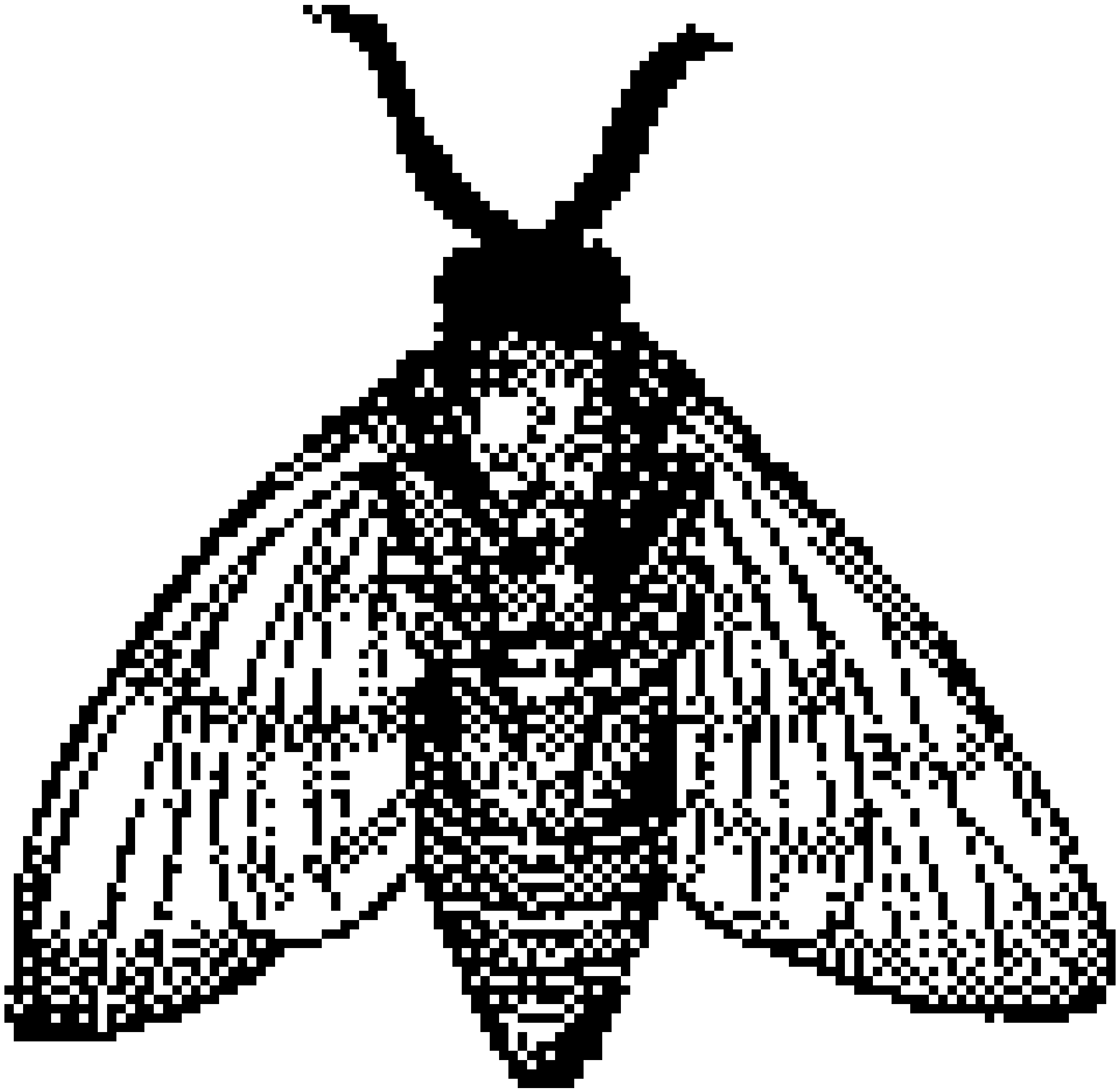


Figure 1: Example caption

4

List of Tables

Table 1: My caption

1	2	3
4	5	6
7	8	9

Big Data for Edge Computing

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Gregor von Laszewski
Indiana University
Smith Research Center
Bloomington, IN 47408, USA
laszewski@gmail.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

Big Data, Edge Computing i523

1 INTRODUCTION

Put here an introduction about your topic. We just need one sample refernce so the paper compiles in LaTeX so we put it here [?].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 CONCLUSION

Put here an conclusion. Conlcusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬁgure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named ”images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

List of Figures

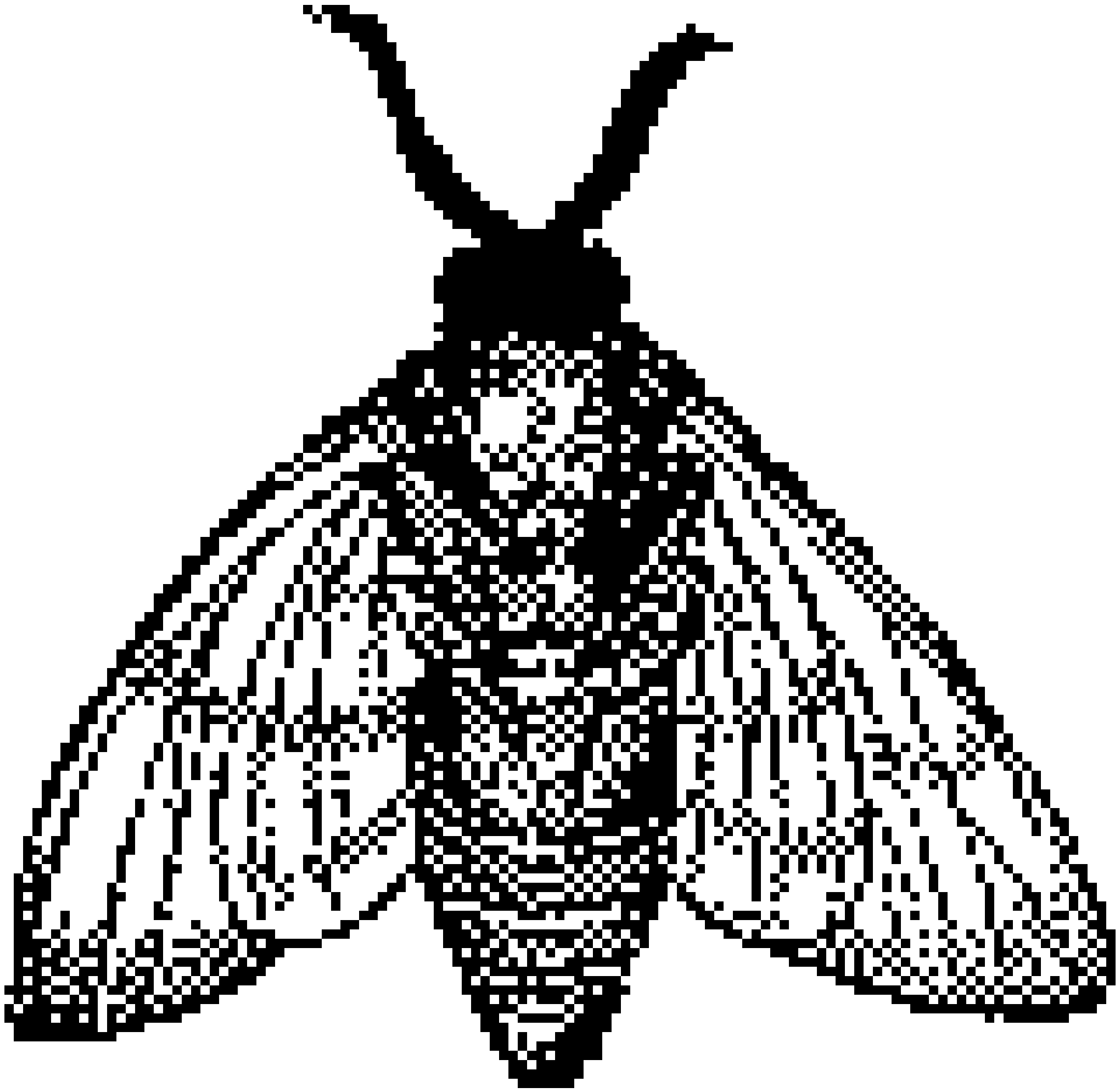


Figure 1: Example caption

List of Tables

Table 1: My caption

1	2	3
4	5	6
7	8	9

Big Data for Edge Computing

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Gregor von Laszewski
Indiana University
Smith Research Center
Bloomington, IN 47408, USA
laszewski@gmail.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

Big Data, Edge Computing i523

1 INTRODUCTION

Put here an introduction about your topic. We just need one sample refernce so the paper compiles in LaTeX so we put it here [?].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table??. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 CONCLUSION

Put here an conclusion. Conlcusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬀigure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inﬂate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named ”images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

List of Figures

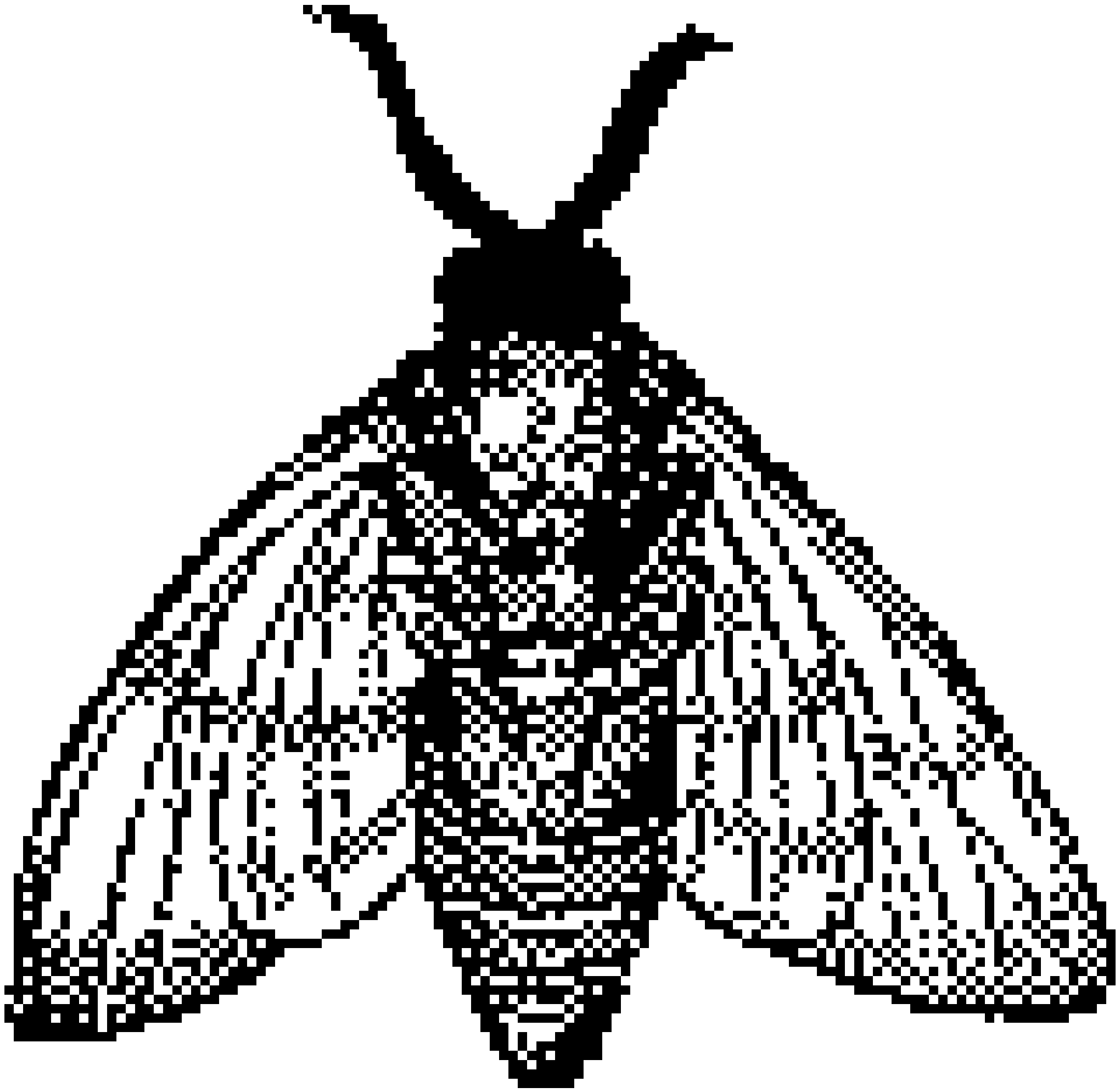


Figure 1: Example caption

List of Tables

Table 1: My caption

1	2	3
4	5	6
7	8	9

Big Data Applications in Aviation Industry

Swargam, Prashanth
Indiana University Bloomington
107 S Indiana Ave
Bloomington, Indiana 47408
pswargam@iu.edu

ABSTRACT

Data generated by aviation industry is being increased enormously. The data generated by all the components of aviation industry can be analysed for reducing the operational costs, predict customer behaviour, analyse customer satisfaction. These applications of big data in aviation industry makes it a prominent player. Hence, collecting this data, storing and processing them for desired results can help the aviation industry in boosting their profits and improve customer satisfaction.Various applications of Big data, their challenges and models are discussed here.

KEYWORDS

HID228, I523, Big Data, Aviation Industry,Analytics ,

1 INTRODUCTION

1.1 4 V's in Big Data in Aviation Industry

2 DATA SOURCES

3 DATA INTEGRATION

3.1 Service Oriented Model

4 DATA STORAGE AND PROCESSING

4.1 Apache Spark for Realtime data

5 CHALLENGES IN IMPLEMENTING BIG DATA

5.1 Information Security

6 CONCLUSION

Big Health Data from Wearable Electronic Sensors (WES) and the Treatment of Opioid Addiction

Sean M. Shiverick
Indiana University Bloomington
smshiver@indiana.edu

ABSTRACT

Wearable electronic sensors (WES) generate to collect vital health data in the treatment of opioid addiction.

KEYWORDS

Big Data Applications, Health Analytics, Wearable Sensors, i535, HID335

1 INTRODUCTION

In the increasingly connected digital age, personal electronic devices are generating large volumes of data with important applications for health analytics. Wearable electronic sensors (i.e., *wearables*) and fitness monitors (e.g, FitBit, iWatch) can record our movements and vital physiological measures such as heart rate, temperature, and blood pressure [5]. Consumers are using wearables to self-monitor stress and hypertension, and wearable sensors can be used to help track recovery following medical procedures such as surgery [1]. The development of personalized health care models are also enabling individuals to self-monitor and manage their own health in partnership with care providers. This paper explores approaches to using personal electronic devices and wearable sensors for the treatment of opioid addiction and prevention of drug overdose. Past research has shown that *Mobile Health* platforms have been used to address prescription medication abuse in several ways: (a) monitor patient health conditions at any time and remotely, (b) monitor medication consumption, and (c) connect patients with health care providers and treatment interventions [?]. A review of the literature shows that emerging digital technologies, such as tatoo biosensors and long range (lora) backscatter, can provide health data in real time to assist patients in addiction recovery to resist physical cravings and prevent relapse. Mobile applications can play an important role in supplementing traditional treatment approaches to drug addiction and recovery.

1.1 Medication Abuse and Opioid Addiction

The abuse of prescription opioid medication in the U.S. has become a major health crisis that the Department of Health and Human Services (HHS) has described as an epidemic [14]. Approximately 2 million Americans were dependent on or abused prescription opioids (e.g., oxycodone, hydrocodone) in 2014 [6]. Overdose deaths from prescription opioids has quadrupled since 1999, resulting in more than 180,000 deaths between 1999 to 2015. Public health agencies are implementing comprehensive efforts to address four major risk areas of prescription opioid abuse, overdoses, and deaths: (i) Increasing knowledge of opioid abuse and improving decisions among medication prescribers, (ii) Reducing inappropriate access to opioids, (iii) Increasing effective overdose treatment, (iv) Providing substance-abuse treatment to persons addicted to opioids. In addition, researchers have developed Mobile Health applications to monitor prescribed medication consumption for potential abuse [?]. Figure 1 shows steps involved in the process and decision support structure for a medication abuse monitoring system. There are several problems faced in implementing a monitoring systems for medication abuse; foremost is the difficulty in getting reliable data on medication consumption from potentially addicted individuals, based on self-reports. Concstructing an accurate prediction model for complex behavior such as addiction is also a challenge. Other issues arise related to patient privacy, confidentiality, and regulation of controlled opioid medications.

[Figure 1 about here.]

1.1.1 Drug Addiction and Treatment. For millions of people struggling with substance abuse and dependency in the U.S., addiction and relapse are chronic health conditions [2]. Drug addiction has many similar characteristics to other chronic medical illnesses; however, there are unique challenges to the treatment of addiction illnesses. For example, drug addicted patients undergo intense detoxification in rehabilitation treatment programs, but then are released back into the same environment associated with their drug use. The lack of continuity in the treatment of addiction disorders, leaves addicts in recovery at high risk for relapse into substance use and abuse. Second, individuals with addiction disorders present for care to emergency rooms after acute intoxication, often following law enforcement interventions. Emergency personal and very capable at crisis intervention for drug overdose, but lack resources to evaluate severe addiction disorders or provide follow-up. Furthermore, addicted individuals seeking treatment often relapse at night or on weekends when treatment centers are not open. Various theories of addiction and relapse have been proposed. According to the classical conditioning model, situational cues or events can elicit a motivational state underlying relapse to drug use. A slightly more complex model suggests that addictive behavior can be reinstated after extinction of dependency by exposure to drugs, drug-related cues, or environmental stressors [12]. Understanding that a user‘s affective response to cues in the environmental can lead to relapse and drug use are key to developing strategies for prevention and treatment.

1.2 Technology-Based Interventions for Addiction Treatment

Technology-based interventions have been used for drug addiction assessment, treatment, prevention and recovery [9]. In terms of assessment, data about individuals substance use can be obtained from mobile cell phone reporting outside of treatment settings.

Web-based approaches to treatment have been implemented online to improve behavioral and psychosocial functioning for addicted individuals in recovery [10]. For example, the Therapeutic Education System (TES) is a self-directed, web-based interactive treatment program consisted of 65 training modules that focuses cognitive-behavioral skills, psychosocial functioning (family/social relations). This online approach helped to increase access to treatment for individuals in rural areas, and included an optional contingency management module. A computer based Training in Cognitive Behavioral Therapy (CBT) program was found to enhance treatment outcomes when provided in conjunction with traditional substance abuse treatment, and helped improve coping skills and decision-making skills [4]. In evaluating the effectiveness of mobile applications for addiction treatment, several questions remain to be answered: First, if mobile applications primarily are regarded primarily as supplements to traditional therapeutic treatment, can their effectiveness be measured independently from the approach used in treatment? Second, over what time period period can the benefits of mobile applications be observed? Research evidence suggests that the benefits of mobile interventions may be limited to 12 or 15 weeks [13]. However, it is unclear whether individuals struggling from addiction would continue to use mobile treatment applications in the long terms, beyond a limited course of treatment.

1.2.1 Mobile-Based Applications. Mobile based applications have been used for monitoring and treatment of substance abuse and addiction disorders for several decades [2]. Early applications included the use of electronic pagers (i.e., beepers) for experience sampling with paper-based assessments that generated data about daily life behavior and experiences [13]. In the 1990s, programmable personal digital assistants (e.g., palm-pilot) enabled collection of data electronically, and subsequent mobile research tools facilitated the collection of information about psychological factors (e.g., daily stressors, emotional states, thoughts) and other variables related to addiction (e.g., craving, contextual cues, actual substance use). Assessments performed several times throughout the day (commonly, every 2 to 4 hours) allowed for analysis of the daily fluctuations of these symptoms and features. Historically, addiction research has faced some unique challenges that the use of mobile technologies may help to overcome. Methodological aspects of traditional research using retrospective, cross-sectional, or longitudinal assessments (over periods of weeks, months, or years) have been problematic for investigating risk factors including behaviors and symptoms (severe physiological cravings, withdrawal, and substance use) that can span a relatively short time. An additional factor is the comorbidity, or co-occurence, of substance use disorders (SUDs) with other psychological disorders, such as anxiety and mood disorders. For example, the “self-medicationfifi model has commonly been used to explain the association between alcohol abuse is used as an effort by an individual to reduce or cope with a high degree of anxiety (or depression). It has also been challenging for researchers to capture the role of environmental or contextual cues (e.g., people, places, things) associated with substance abuse and addiction, which can trigger relapse for individuals in recovery.

Smartphone Applications. Continued care is an important ingredient for recovery from addiction that involves monitoring, outreach, planning, case management, and social support [8]. Smartphone applications can help individuals in recovery to monitor cravings at critical points in daily life, track contextual cues associated with substance use, and provide outreach to support services. A team of researchers at the University of Wisconsin evaluated the effectiveness of a smartphone application called Addiction Comprehensive Health Enhancement Support System (A-CHESS), designed to provide recovery support patients leaving residential alcohol treatment center [7]. A-CHESS provided anytime, anywhere access to support services in audio-visual format, GPS monitoring and warnings for risky locations, and communication with counselors. Over an 8-month period and 4 month follow-up, patients who used the A-CHESS intervention reported fewer risky drinking days, on average, per month than patients in a comparable control group. The findings provide evidence that the smartphone intervention was effective at treating a critical behavioral measure for treatment of alcohol use disorder (AUD).

1.3 Medication Abuse Monitoring System [?]

[Figure 2 about here.]

1.4 Wearable Sensors

Many smartphones have built-in sensors (accelerometer, odometer, GPS) that can track movement and activity. Although not specifically designed to collect clinical data, smartphone sensors can be repurposed help addicted individuals to track signs of potential relapse. Additional sensors could be added to a smartphone to monitor heart rate, respiration, and body temperature and communicate physiological data to care providers or treatment specialists to provide support or initiate an intervention if necessary [8]. GPS coordinates can be used to monitor proximity to locations or persons associated with substance use

Real-Time Mobile Detection of Drug Use with Wearable Biosensors: A Pilot Study [3]

1.5 Emerging Technologies

1.5.1 Tattoo sensors. If you like to see a more elaborate example, please look at report-long.tex.

1.5.2 LoRa Backscatter. If you like to see a more elaborate example, please look at report-long.tex.

1.6 Effectiveness of Technology-Based Interventions Addiction Treatment and Recovery

Figure 1 shows that dramatic increase in overdose deaths in the U.S. between 2000 and 2016 are from synthetic opioids (other than methadone), natural and semi-synthetic opioids, and heroin [11]. Of the estimated 64,000 drug overdose deaths in 2015, over 20,000 overdose deaths were from fentanyl and other synthetic opioid analogs.

[Figure 3 about here.]

2 FIGURES

just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as google for them). They even allow you to manage tables as CSV.

Note that the caption is before the tabular environment.

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 CONCLUSION

Put here an conclusion. Conlcusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

[1] Louis Atallah, Gareth G. Jones, Raza Ali, Julian J. H. Leong, Benny Lo, and Guang-Zhong Yang. 2011. Observing Recovery from Knee-Replacement Surgery by Using Wearable Sensors. In *Proceedings of the 2011 International Conference on Body Sensor Networks (BSN '11)*. IEEE Computer Society, Washington, DC, USA, 29–34. https://doi.org/10.1109/BSN.2011.10

[2] E.W. Boyer, D. Smelson, R. Fletcher, D Ziedonis, and Picard R. W. 2010. Wireless Technologies, Ubiquitous Computing and Mobile Health: Application to Drug Abuse Treatment and Compliance with HIV Therapies. *Journal of Medical Toxicology* 6, 2 (2010), 212–216. https://doi.org/doi:10.1007/s13181-010-0080-z

[3] Stephanie Carreiro, David Smelson, Megan Ranney, Keith J. Horvath, R. W. Picard, Edwin D. Boudreaux, Rashelle Hayes, and Edward W. Boyer. 2015. Real-Time Mobile Detection of Drug Use with Wearable Biosensors: A Pilot Study. *Journal of Medical Toxicology.* 11, 1 (Oct. 2015), 73–79. https://doi.org/10.1007/s13181-014-0439-7.

[4] K.M. Carroll, S.A. Ball, S. Martino, and et al. 2008. Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. *Am J Psychiatry* 165, 7 (2008), 881fi?!8. https://doi.org/10.1176/appi.ajp.2008.07111835

[5] Melinda Gomez Michael Schwartz David Metcalf, Sharlin T.J. Milliard. 2016. Wearables and the Internet of Things for Health. *IEEE Pulse* (Oct. 2016). https://pulse.embs.org/september-2016/wearables-internet-of-things-iot-health/

[6] Centers for Disease Control and Prevention. 2017. Prescription Opioid Overdose Data. online. (Oct. 2017). https://www.cdc.gov/drugoverdose/data/overdose.html

[7] D.H. Gustafson, F.M. McTavish, M.-Y. Chih, A.K. Atwood, R.G. Johnson, M. Boyle, and M. ... Shah. 2014. A smartphone application to support recovery from alcoholism: A randomized controlled trial. *JAMA psychiatry.* 71, 5 (May 2014), 566–572. https://doi.org/10.1001/jamapsychiatry.2013.4642

[8] K. Johnson, A. Isham, D.V. Shah, and D.H. Gustafson. 2011. Potential Roles for New Communication Technologies in Treatment of Addiction. *Current psychiatry reports.* (2011). https://doi.org/10.1007/s11920-011-0218-y

[9] Lisa A. Marsch. 2012. Leveraging teachnology to enhance addiction treatment and recovery. *Journal of Addictive Diseases* 31, 3 (2012), 313–318. https://doi.org/10.1080/10550887.2012.694606

[10] Dallery J. Marsch LA. 2012. Advances in the Psychosocial Treatment of Addiction: The Role of Technology in the Delivery of Evidence-Based Psychosocial Treatment. *The Psychiatric Clinics of North America* ;35(2):. doi:, 2 (2012), 481–493. https://doi.org/10.1016/j.psc.2012.03.009

[11] National Institute on Drug Abuse (NIDA). 2017. *Overdose Death Rates.* Summary. National Institutes of Health (NIH), Washington D.C. https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates

[12] Yavin Shaham, Uri Shalev, Lin Lu, Harriet de Wit, and Jane Stewart. 2003. The reinstatement model of drug relapse: history, methodology and major findings. *Psychopharmacology* 168, 1 (01 Jul 2003), 3–20. https://doi.org/10.1007/s00213-002-1224-x

[13] J. Swendsen. 2016. Contributions of mobile technologies to addiction research. *Dialogues Clinical Neuroscience* (2016). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4969708/

[14] Nora D. Volkow, Thomas R. Frieden, Pamela S. Hyde, and Stephen S. Cha. 2014. Medication-Assisted Therapies: Tackling the Opioid-Overdose Epidemic. *New England Journal of Medicine* 370, 22 (2014), 2063–2066. https://doi.org/10.1056/NEJMp1402780 arXiv:http://dx.doi.org/10.1056/NEJMp1402780 PMID: 24758595.

REFERENCES

[1] Louis Atallah, Gareth G. Jones, Raza Ali, Julian J. H. Leong, Benny Lo, and Guang-Zhong Yang. 2011. Observing Recovery from Knee-Replacement Surgery by Using Wearable Sensors. In *Proceedings of the 2011 International Conference on Body Sensor Networks (BSN '11)*. IEEE Computer Society, Washington, DC, USA, 29–34. https://doi.org/10.1109/BSN.2011.10

[2] E.W. Boyer, D. Smelson, R. Fletcher, D Ziedonis, and Picard R. W. 2010. Wireless Technologies, Ubiquitous Computing and Mobile Health: Application to Drug Abuse Treatment and Compliance with HIV Therapies. *Journal of Medical Toxicology* 6, 2 (2010), 212–216. https://doi.org/doi:10.1007/s13181-010-0080-z

[3] Stephanie Carreiro, David Smelson, Megan Ranney, Keith J. Horvath, R. W. Picard, Edwin D. Boudreaux, Rashelle Hayes, and Edward W. Boyer. 2015. Real-Time Mobile Detection of Drug Use with Wearable Biosensors: A Pilot Study. *Journal of Medical Toxicology.* 11, 1 (Oct. 2015), 73–79. https://doi.org/10.1007/s13181-014-0439-7.

[4] K.M. Carroll, S.A. Ball, S. Martino, and et al. 2008. Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. *Am J Psychiatry* 165, 7 (2008), 881fi?!8. https://doi.org/10.1176/appi.ajp.2008.07111835

[5] Melinda Gomez Michael Schwartz David Metcalf, Sharlin T.J. Milliard. 2016. Wearables and the Internet of Things for Health. *IEEE Pulse* (Oct. 2016). https://pulse.embs.org/september-2016/wearables-internet-of-things-iot-health/

[6] Centers for Disease Control and Prevention. 2017. Prescription Opioid Overdose Data. online. (Oct. 2017). https://www.cdc.gov/drugoverdose/data/overdose.html

[7] D.H. Gustafson, F.M. McTavish, M.-Y. Chih, A.K. Atwood, R.G. Johnson, M. Boyle, and M. ... Shah. 2014. A smartphone application to support recovery from alcoholism: A randomized controlled trial. *JAMA psychiatry.* 71, 5 (May 2014), 566–572. https://doi.org/10.1001/jamapsychiatry.2013.4642

[8] K. Johnson, A. Isham, D.V. Shah, and D.H. Gustafson. 2011. Potential Roles for New Communication Technologies in Treatment of Addiction. *Current psychiatry reports.* (2011). https://doi.org/10.1007/s11920-011-0218-y

[9] Lisa A. Marsch. 2012. Leveraging teachnology to enhance addiction treatment and recovery. *Journal of Addictive Diseases* 31, 3 (2012), 313–318. https://doi.org/10.1080/10550887.2012.694606

[10] Dallery J. Marsch LA. 2012. Advances in the Psychosocial Treatment of Addiction: The Role of Technology in the Delivery of Evidence-Based Psychosocial Treatment. *The Psychiatric Clinics of North America* ;35(2):. doi:, 2 (2012), 481–493. https://doi.org/10.1016/j.psc.2012.03.009

[11] National Institute on Drug Abuse (NIDA). 2017. *Overdose Death Rates.* Summary. National Institutes of Health (NIH), Washington D.C. https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates

[12] Yavin Shaham, Uri Shalev, Lin Lu, Harriet de Wit, and Jane Stewart. 2003. The reinstatement model of drug relapse: history, methodology and major findings. *Psychopharmacology* 168, 1 (01 Jul 2003), 3–20. https://doi.org/10.1007/s00213-002-1224-x

[13] J. Swendsen. 2016. Contributions of mobile technologies to addiction research. *Dialogues Clinical Neuroscience* (2016). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4969708/

[14] Nora D. Volkow, Thomas R. Frieden, Pamela S. Hyde, and Stephen S. Cha. 2014. Medication-Assisted Therapies: Tackling the Opioid-Overdose Epidemic. *New England Journal of Medicine* 370, 22 (2014), 2063–2066. https://doi.org/10.1056/NEJMp1402780 arXiv:http://dx.doi.org/10.1056/NEJMp1402780 PMID: 24758595.

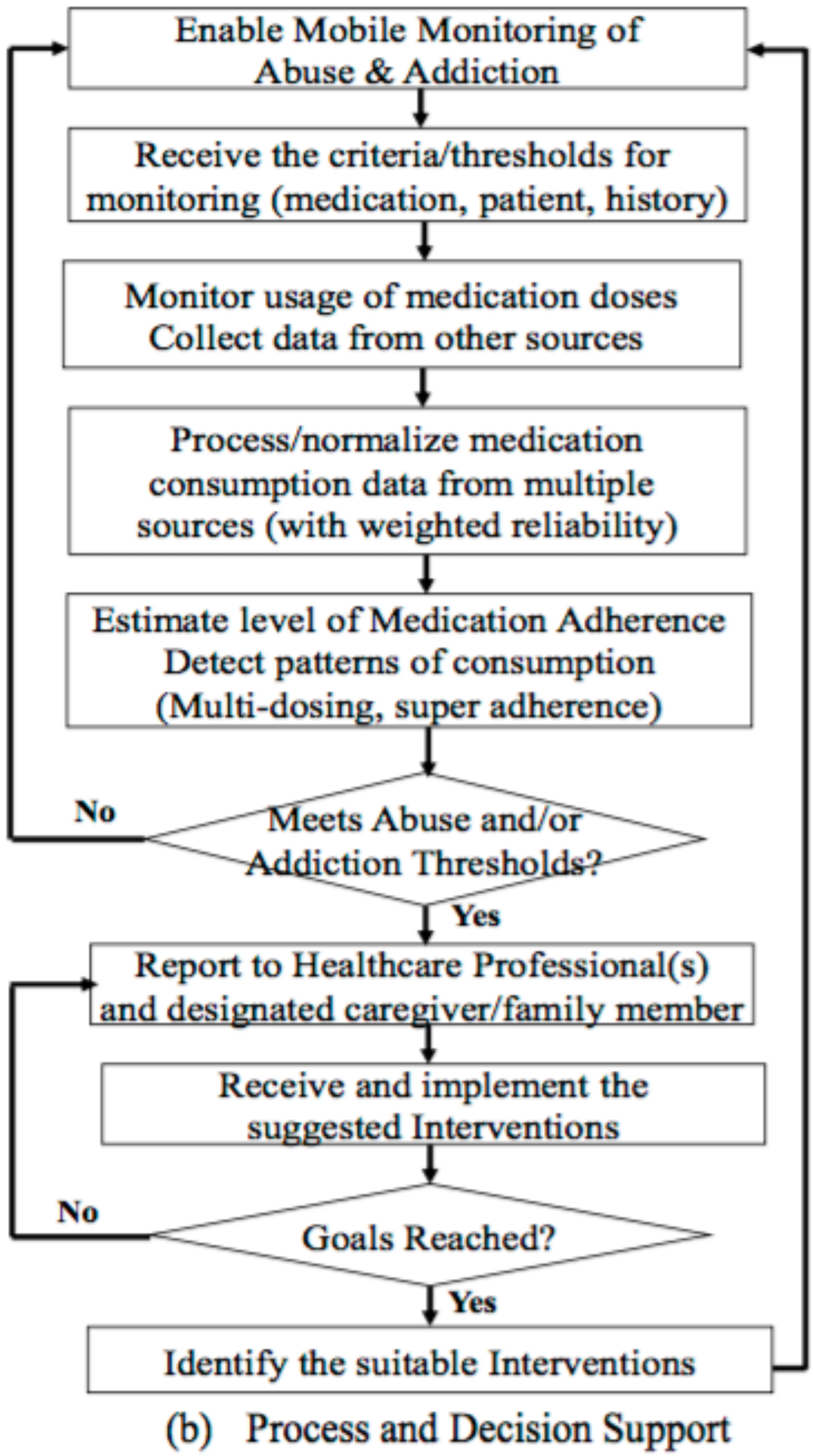
We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

4

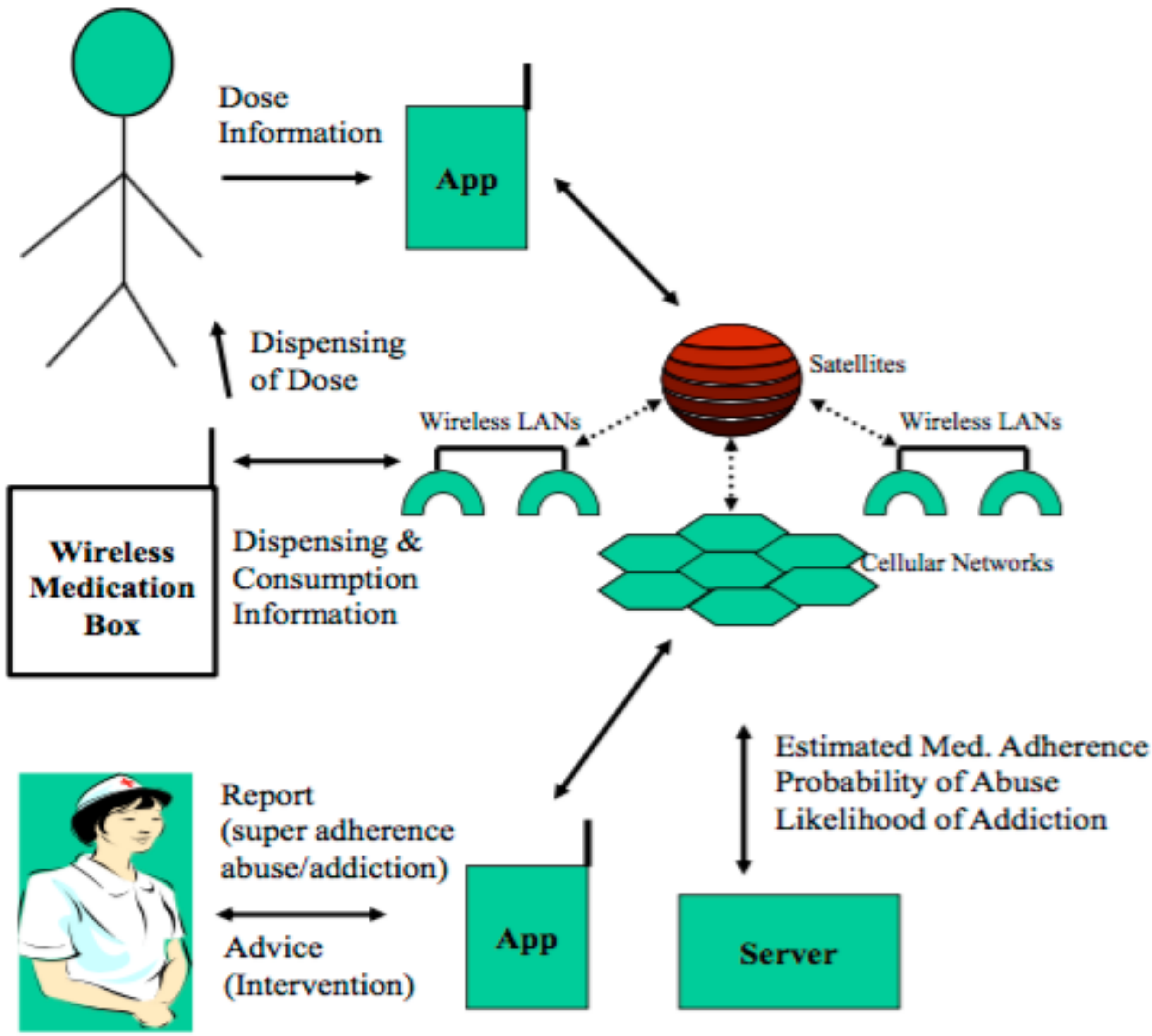
List of Figures

6



7

Figure 1: Process and Decision Support for Abuse Monitoring System [?]



(a) Architecture of the Abuse Monitoring System

8

Figure 2: Architecture of Abuse Monitoring System [?]

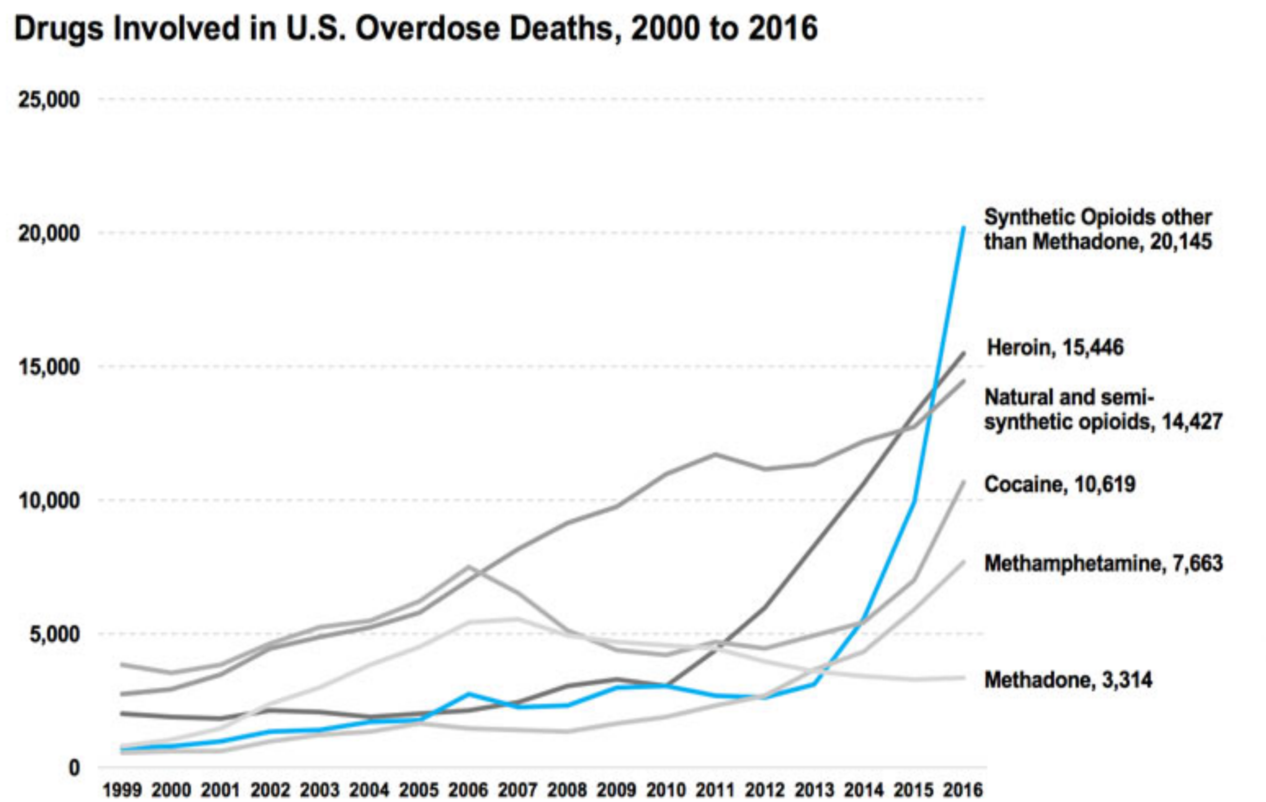


Figure 3: Drugs Involved in U.S. Overdose Deaths from 2000 to 2016, National Institute on Drug Addiction (NIDS) [11]

9

Natural Language Processing (NLP) to analyze human speech data

Ashok Reddy Singam
Indiana University
711 N Park Ave
Bloomington, Indiana 47408
asingam@iu.edu

Anil Ravi
Indiana University
711 N Park Ave
Bloomington, Indiana 47408
anilravi@iu.edu

ABSTRACT

Extracting meaningful information from large volumes of unstructured human language is a challenging big data problem. Automatic speech recognition (ASR) and natural language processing (NLP) based intelligent system can be used in several human machine interface applications both in consumer and industrial sector. Here describing the architecture, building blocks, performance and applications for such system that would use pre-developed ASR and NLP APIs.

KEYWORDS

i523, HID333, HID337, Natural Language Processing

1 INTRODUCTION

As voice becoming a common user interface, the need for accurate and intelligent speech recognition technologies is growing. In speech processing technology there are two main subtasks

- Speaker Recognition
- Speech Recognition

Although the performance of current speaker and speech recognition systems is far from perfect, these systems have already proven their usefulness for certain applications.

2 SPEAKER RECOGNITION

Speaker identification is one of the important task in speech processing. Each person has a voice that is different from everyone elsefis.Speaker recognition is the process of identifying who is speaking by using acoustic features of speech.Speaker recognition has been applied mostly in security applications to control access. Current speaker recognition systems are not very accurate for large speaker populations.

3 NLP FOR SPEECH RECOGNITION

Speech recognition is the ability to identify spoken words. It is the process of converting speech into text. This process prepares the input data (speech) to be appropriate for Natural Language Processing(NLP). NLP is the processing of the text to understand the meaning of the text. It comes as the next step of speech recognition.

Analyzing language for its meaning is a complex task. Modern speech recognition research began in the late 1950s with the advent of the digital computer. The 1960s saw advances in the automatic segmentation of speech into units of linguistic relevance like phonemes, syllables. And now with advancements in the field of Artificial Intelligence, neural networks have been used in many aspects of speech recognition such as phoneme classification,isolated word recognition,audiovisual speech recognition, audiovisual speaker recognition and speaker adaptation. In the context of Speech Recognition, NLP involves 4 basic steps

- Morphological Analysis
- Syntactic Analysis
- Semantic Analysis
- Pragmatic Analysis

3.1 Morphological Analysis

Morphological analysis is the identification, analysis, and description of the structure of a given languagefis root words, word boundaries, affixes, parts of speech, etc. There are two typical problems in this area, which includes word segmentation and part-of-speech (POS) tagging. Word segmentation is the problem of finding word boundaries in a corpus.

3.2 Syntactic Analysis

Syntactic analysis or parsing is the process of analyzing a string of symbols, either in natural language or in computer languages, conforming to the rules of a formal grammar.

3.3 Semantic Analysis

Semantic analysis is the process of relating syntactic structures, from the levels of phrases, clauses, sentences and paragraphs to the level of the writing as a whole, to their language-independent meanings.

3.4 Pragmatic Analysis

Pragmatic Analysis is how sentences are used in different situations and how use affects the interpretation of the sentence. Means what was said is reinterpreted as what it actually means.

4 CONCLUSION

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES