

57th Hawai'i International Conference on System Sciences
Decision Analytics and Service Sciences Track
Big Data and Analytics
Minitrack Introduction

Stephen Kaisler
SHK & Associates

skaisler1@comcast.net

J. Alberto Espinosa
American University

alberto@american.edu

Derrick Cogburn
American University

dcogburn@american.edu

Frank Armour
American University

fjarmour@gmail.com

Introduction

For HICSS-57, we are pleased to introduce six papers by minitrack authors and two papers by minitrack co-chairs.

Our first paper, entitled *The Effectiveness of Fitness Wearable Technologies: Applying Big-Data Techniques to Analyze User Experience and Perception* by Arpit Sharma, Deepti Koshi, William Money, and Denise McWilliams, focuses on user motivations by analyzing user perceptions across three themes: Accomplishment (fitness goals), Health (chronic), and Social (connections with peers), espoused by wearers. Fitness Wearable Technologies (FWT) have been analyzed as input devices for various user outcomes. User motives behind device usage have not been assessed from a longitudinal or stability perspective, nor have they been compared across different devices. FTWs are "small, lightweight technologies that can be easily placed on human bodies as they move around in time and space". As the authors note, FTWs are popular, but questions about their influence on wearers remain unanswered.

The authors analyzed Twitter theme data drawn in the samples from the spring quarter of 'Pre, During, and Post' "Covid" years (2019–2022). They compared sentiment analysis across three sentiment analyzers and found issues with the output at various places. For many tweets, the sentiment analyzers showed a different outcome, which raises trust issues. They used several algorithms - Sprinklr, TextBlob, and VADER to analyze the data. Overall, they observed an increase in positive sentiment across health themes for the devices that were studied.,

They concluded that for fitness routines, the fitness challenge feature motivates users to continue to use the device while focusing on their health goals. While this helps accomplish user's fitness goals, it does so by connecting them socially. Under the theme accomplish, two prominent features - tracking steps

and fitness challenges - motivated users to continue to use the device. Under the theme health, sleep, heart rate, and interoperability were prominent features motivating users. Under the theme social, network connectivity, sharing fitness goals/ challenges, and tracking lost devices were prominent features motivating individuals to use the device. While sentiment analyzers are useful for assessing overall sentiment scores, multiple sentiment analyzers should be used to support conclusions, because the sentiment analyzers still have room to improve in detecting sarcasm.

Our second paper, entitled *Exploring the Intellectual Composition of Academic Research Conferences: Computational Text Analysis of the HICSS Paper Archive from 2017-2022* by Derrick Coburn (co-Chair), Theodore Ochieng, Sierra Buehlman Barbeau, and Haiman Wong, provides a computational method for assessing academic research conferences by exploring the intellectual composition of the HICSS conference asking: what themes are most prevalent across the conference? With HICSS, the diversity of the tracks and minitracks makes it difficult to determine the track of a paper from its abstract. The rapidly evolving nature of information systems, computer science, and system sciences makes it difficult to determine the dominant research themes, and correlate them with perceived emerging research issues.

The authors analyzed the HICSS conference as the premier conference for state-of-the-art and emerging topics. The authors examined 5,024 papers accepted for HICSS conferences from 2017 to 2022 and 18 metadata variables. Text mining techniques, including TF-IDF word and phrase frequencies and NLP techniques to develop models of themes contained within the papers. They identified the most frequent 100 keywords using stemming followed by TF-IDF analysis to identify important research themes. NLP techniques were used to identify topics and people followed by LDA topic modeling to

identify the top 10 topics, which showed the diversity of research interests across the conference. This paper represents a way to track the evolution of emerging research areas as indicated by the breadth and depth of papers accepted by HICSS.

Our third paper, entitled *A Graph-Theoretic Approach for Examining Team Communities in Sports Transfer Markets*, by Eleuterio Lillo Portero and Hesham Ali, examines the acquisition and sales of players in the top teams in the European League and English Premier League. This market has grown significantly as has the magnitude of salaries and the size of barter where players and draft positions are traded among teams. It addresses the factors that affect trades in the top teams based on performance, monetary power, and trophy winning records, and the correlation between money sent versus team performance.

This paper examines the correlations between factors by constructing graphs of trading activities. Using the Clauset-Newman-Monet greedy modularization maximization algorithm, teams are assigned to communities and then combines pairs of communities to achieve the largest modularity. Several metrics, including Betweenness Centrality, the Small World Property, and the Spearman Rank Correlation, assess the relative ranking of teams.

They conclude that winning and achieving high performance levels is closely related to the level of investments (in players), particularly among international trophy winners. This was not an expected observation, but is related to the “status” of the team, but also establishing and maintaining a status level in the sport that may improve the overall performance of those teams occupying these high status levels. This innovative applies graph theory to interesting economic and sociological questions that had previously been raised, but not adequately explained.

Our fourth paper entitled *Impact of Sustainability Disclosures on Financial Performance: A Natural Language Processing Perspective*, by Cladie Lui, Jeremy Chia, and Connie Shum, examines how sustainability reporting and corporate financial success are related. This research uses Natural Language Processing (NLP) techniques to process such reports and compare the results against corporate performance results to show that certain disclosure have a positive relation with economic success indicators. It examines the change in attitudes of Singapore corporations related to Corporate social Responsibility (CSR) and its effect Environment, social, and Governance (ESG) performance.

The authors used Gibbs Sampling for Latent Dirichlet Allocation to identify an optimal number of topics for Topic Modelling. They determined a minimum number of topics is seven. Emphasis on governance topics is directly related to the perception of a firm’s identity which can mitigate negative perceptions of a firm. Using several lexicons, they performed a sentiment analysis. Prior studies found that over time, firms are consistent in their use of positive and negative words. But, for specific events that could lead to negative perceptions, they tend to obfuscate the bad news. The Flesch Reading Ease score was used to assess readability. Low readability was correlated with perceived obfuscation.

They conclude from the three models they used to assess Return on Assets (ROA) that corporate support for community initiatives, compliance with sustainability activities, and substantive disclosure of climate-related activities lead to financial viability for a firm. This paper shows that disciplined processing of corporate reports can reveal information that allows the assessment of corporate activities in sustainability that can directly affect the perception of a firm and its financial viability.

Our fifth paper, entitled *Data Analytics Capability Maturity Models for Small and Medium Enterprises – A Systematic Literature Review*, by Robert Marohn and Yan Li, performed a systematic literature review to assess 18 models to determine if they could assist subject matter experts (SMEs) in developing a data analytics capability. Many Data Analytics Capability Maturity Models (DACMMs) have been developed to assess data analytics capabilities. This literature review sought to determine what models exist, what their characteristics are, and if any have a prescriptive focus. A total of 18 models were selected based on inclusion and exclusion criteria. These models had different levels with different names which made comparison somewhat difficult. They identified five common areas: Organization, Infrastructure, Analytic Operations, Data Management, and Data Governance.

Several models used existing theories, but no new theories were developed in any model. Nine of the 18 models used Design Science Research to design the maturity model. The result of this review is that all the models had some limitations and none seemed to be the best fit for SMEs. The author’s recommendation was that a DACCM needs to address essential characteristics and assessment components. The authors propose a DACMM of eight elements: Dimensions and Catalog, Stages of Maturity Diagram,

Maturity Assessment Process, Maturity Output Form, Prescriptive Process, Prescriptive Level Output Form, and Roadmap, which they believe would better serve the needs of SMEs.

Our sixth paper, entitled *Classification of Experience for Proactive In-Car Function Recommendations Based on Customer Usage Data*, by Christian Micus, Daniel Homola, Markus Böhm, and Helmut Krcmar, examines why used data collected from cars indicates why driver assistance functions a low usage rate. They classified customer experience based on the driving context and developed a machine learning program to make recommendations based on customer usage. They evaluated their approach using over 716,000 function activations by users from an automobile manufacturer.

The collected data reflected the actual use of driver assistance functions. Their ML program generated recommendations when all condition variables were satisfied as well as for different driving contexts. The data set was analyzed and 21 features related to the function activation context were extracted. These were organized in three categories: (1) the car context, (2) the weather context, and (3) the environment. Two labels – positive and negative – were defined, such that positive led to supportive experience for the driver, while negative led to a discouraging experience for the driver. Class imbalance of 74% negative versus 26% positive was observed. After correcting for the class imbalance, they used a 6-feature model for both fixed conditions and the ML program and a 21-feature ML program.

The ML program demonstrated 73% accuracy in recommending positive activations for the driver versus 56% for the fixed conditions model. While the authors note some key limitations, this paper demonstrates that ML programs under varying conditions can provide support to possibly warn drivers of function activation that will benefit their driving experience.

Our seventh paper, entitled *Analytcs Morphology and Transformation*, by Stephen Kaisler, William Money, J. Alberto Espinosa, and Frank Armour, begins a research effort to classify the structure of analytics and explore transformations of analytics from one structure to another given certain characteristics of an analytic, such as performance. Classes of an analytic taxonomy from previous research had structures identified based on the literature review.

The authors define the term ‘analytic’ and identify two types: algorithms and heuristics. Through

a literature survey, a taxonomy of structure was developed for representing analytics, including strings, matrices, equations, rules, etc. They also identified functional and non-functional attributes that could characterize a particular representation. Transformers were defined as an algorithm/heuristic which could transform an analytic from one representation another. Two types of transformers were identified – structural and performance. Two examples of a representation scheme were presented for graphs and equations to demonstrate that these schemes often have common elements.

The authors note that handling dynamic analytics and sets, such as a set of equations representing an analytic, posed additional problems to be further investigated. Further research is proposed for an analytics catalog describing classes of analytics and their features as well as a catalog of transformations.

Our eighth paper, entitled *Interoperability for Autonomy*, by Trevor Bihl, Kara Combs, and James Pennington, noted that autonomous systems try to augment human capabilities with machine-based decision-making in the absence of a user. Supporting legacy systems with this capability requires consideration of interoperability with a decision-making system. The authors develop a taxonomy of interoperability, extend it to include autonomy, and develop a unified autonomy stack embracing hardware and software as a mechanism supporting such systems. Autonomy involves a complex interaction between systems and subsystems. These interactions often involve (may, require) data conversion to effect communication. Technology stacks, such as the OSI Model, can mediate some of the problems associated with implementing such systems.

The authors examine the features of such stacks to determine metrics of interoperability which can enable evaluation of such stacks in a standard way. They identify both current metrics and suggest several new concepts that can further the development of assessment and evaluation projects. A brief case study of four autonomy architectures with respect to their proposed unified autonomy stack was performed which identified weaknesses in each architecture with respect to interoperability with a paired architecture, but highlighted the need for a more rigorous methodology for designing autonomous systems or enhancing existing architectures that communicate with each other to improve their interoperability.