A Data Fusion Approach for Digital Dependency

1st Gustavo Lazarotto Schroeder
dept. name of organization (of Aff.)
Universidade do Vale do Rio dos Sinos (UNISINOS)
São Leopoldo, Brazil
gustavo.schroeder@outlook.com

***Abstract*—**
***Index Terms*—**

I. Introduction

Modern technology has paved the way for the internet and devices like smartwatches and smartphones. Computers are increasingly faster, more portable, and higher-powered than ever before. Thanks to rapid advancements in technology people are more connected than ever, smartphones provide connectivity to limitless information on-demand, enabling people to solve a wide variety of everyday problems [1].

This kind of connection brings a melancholy side, like internet overuse and screen addiction [2], fear of missing out, anxiety and depression [3], mobile game addiction relation to depression, social anxiety, and loneliness [4], and a growing fear of being without a smartphone [5].

Many individuals have become addicted to using technologies and, as a consequence have experience negative mental effects, furthermore the home confinement due to the COVID-19 pandemic may have worsened this situation [6].

At present, there are studies that explore internet addiction disorder (IAD) level based on the log data [7], IAD based on the browser history [10], and Smartphone Addiction based on questionnaires and data collected from the smartphones daily use [7].

In this paper we tackle the problem of turning data acquired from smartphone sensors, ecological momentary assessment (EMA), and surveys into into highlevel descriptions. This approach can be described as data fusion, an integration of data and knowledge from several sources [11], in order to create more consistent, accurate, and useful information about behavioral characteristics, which encompasses several technologies and mental problems. Using data fusion this paper aims to combine context information into a context record to later identify the level of digital dependency, fusing contexts of mental problems caused by the overuse of smartphones and social networks, IAD and video game addiction.

II. Related Work

A. Analysis of behavioral characteristics of smartphone addiction using data mining

This study analised smartphone addiction by considering the differences between smartphone usage patterns as well as cognition. Proposing a method that involves automatically collecting and analyzing data through an app. With the use of the app information about smartphone use and questionnaires was collected, data mining was used to divide user users into groups of high-risk, potential-risk and normal users. The authors identified that the variables "number of screen turns" and "actual use time—perceived use time" are more influential on poisoning than the previous research that used the smartphone as poisoning [7].

B. A Smartphone Addiction and Disease Prevention System Through the Collection and Analysis of Smartphone Usage Patterns

In their study on smartphone usage pattern, Ryu et al. [9] studied a smartphone addiction and disease prevention system through smartphone usage pattern collection and analysis. Smartphone usage pattern is defined as the current state of the smartphone.

The smartphone use pattern data was collected through the orientation sensor and display activation state available in the smartphone. In particular, the objective of their study was to prevent the prevalence of addiction and illness caused by smartphone overuse by educating the users about usage patterns through the collected data. However, although their proposed method can be used to predict smartphone addiction disease according to the behavior type, its simplicity in evaluating for the disease using only the smartphone usage time serves as a limitation.

C. Correlation Analysis of Internet Addiction with Daily Behavior: A Data-Driven Method

This study used the Internet Addiction Test (IAT) and context awareness to identify students who were addicted to the internet using internet data from the Campus, thus creating the level of internet addiction. For the quantification of the level of internet addiction five measures was selected. The measures indicate a grade based on previous psychological research, that is, network traffic, total online time on weekends, total online time on work days, average initial login time and last average logout time [8].

D. Internet Addiction and Mental Health Prediction Using Ensemble Learning Based on Web Browsing History

This study [10] analyzed the web browser history of 30 undergraduate students at Universitas Indonesia (UI) during the course of five weeks. The data was analyzed using the support vector machine (SVM) with radial basis function (RBF) kernel as a machine learning method for prediction. The results were subsequently compared using ensemble learning, such as random forest (RF) and gradient boosting (GB). It was then matched with respondents' responses to an Internet Addiction Test (IAT) questionnaire, which measures IAD levels.

The extracted features became input to classify participants' IAD. The results were compared with their IAD results from the IAT questionnaire. Machine learning was also employed to classify the input into respondents' general health (GH) status, which was matched with their responses to the GHQ-12 questionnaire.

E. Smartphone dependence classification using tensor factorization

This study [14] attempted to classify smartphone dependence with the use of a data-driven prediction algorithm. A mobile application was developed to collect smartphone usage data, resulting in a total of 41,683 logs of 48 users.

Using the Korean Smartphone Addiction Proneness Scale for Adults (S-Scale) and a face-to-face offline interview by a psychiatrist and a clinical psychologists the participants were classified into the control group (SUC) or the addiction group (SUD). This study derived smartphone usage patterns using tensor factorization and found the following six optimal usage patterns: 1) social networking services (SNS) during daytime, 2) web surfing, 3) SNS at night, 4) mobile shopping, 5) entertainment, and 6) gaming at night. For all patterns, the usage times of the SUD were much longer than those of the SUC. Demonstrating that usage patterns and membership vectors are effective tools for the assessment and prediction of smartphone dependence.

F. A systemic smartphone usage pattern analysis: focusing on smartphone addiction issue

With the use of a comprehensive smartphone usage logging system various statistical analysis obtained from more than 800 man-days usage logs are presented. The data collection was done through SAMS (Smartphone Addiction Management System). On the client side, the SAMS application continuously monitored the applications in use and stored the usage records in its local storage, later passed to a server [15].

The analysis shows 1) significant difference in usage frequency and time statistics among the applications, 2) different application category preferences between addicts and non-addicts, and 3) little usage variation between weekdays and weekends, but higher usage in night time for addicts [15].

III. Materials and Methods

A. Data Fusion

The automatic and unobtrusive identification of user activities from sensor data is one of the most challenging goals of context-aware computing [TODO].

The most accepted definition for data fusion was provided by the Joint Directors of Laboratories (JDL) workshop [12]: "*A multi-level process dealing with the association, correlation and combination of data and information from single and multiple sources to achieve refined position, identify estimates and complete and timely assessments of situations, threats and their significance*."

Briefly, we can define data fusion as a combination of multiple sources to obtain improved information; in this context, improved information means less expensive, higher quality, or more relevant information [11].

1) Classification Based on the Relations between the Data Sources: Based on the classification proposed by Durrant-Whyte [13] this proposed study can be defined as cooperative. When the provided information is combined into new information that is typically more complex than the original information.

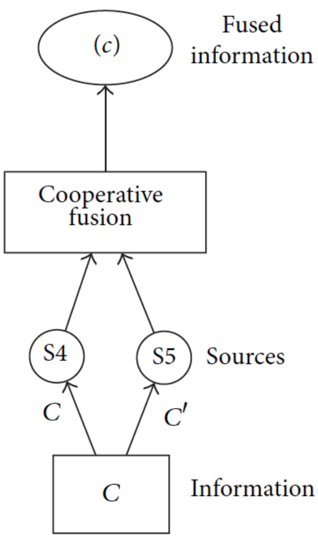

Fig. 1. Whyte's classification based on the relations between the data sources.

2) JDL Data Fusion Framework: This classification is the most popular conceptual model in the data fusion community. It was originally proposed by JDL and the American Department of Defense (DoD) [12]. These organizations classified the data fusion process into five processing levels, an associated database, and an information bus that connects the five components [11]. The framework organization can be seen in Figure 2.

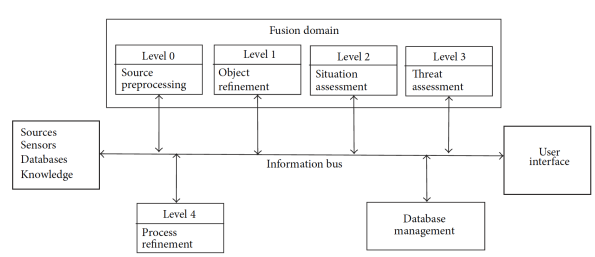

Fig. 2. The JDL data fusion framework.

3) JDL Data Fusion Framework Components:

- Sources are in charge of providing the input data. Here we can use context aware information, questionnaires and sensors.
- Human-computer interaction is an interface that allows inputs to the system from the operators and produces outputs to the operators.
- Database management system stores the provided information and the fused results.

4) JDL Data Fusion Framework Levels:

Level 0 (source preprocessing) is the lowest level of the data fusion process. This level reduces the amount of data and maintains useful information for the high-level processes [11].

Level 1 (object refinement): This level employs the processed data from the previous level. Common procedures of this level include spatio-temporal alignment, association, correlation, clustering or grouping techniques, state estimation, the removal of false positives, identity fusion, and the combining of features that were extracted from images. This stage transforms the input information into consistent data structures [11].

Level 2 (situation assessment): this level focuses on a higher level of inference than level 1. Situation assessment aims to identify the likely situations given the observed events and obtained data. The aim of this level includes performing high-level inferences and identifying significant activities and events (patterns in general).The output is a set of high-level inferences [11].

Level 3 (impact assessment): this level evaluates the impact of the detected activities in level 2 to obtain a proper perspective. This level includes (1) an evaluation of the risk or threat and (2) a prediction of the logical outcome [11].

Level 4 (process refinement): this level improves the process from level 0 to level 3 and provides resource and sensor management. The aim is to achieve efficient resourcem anagement while accounting for task priorities, scheduling, and the control of available resources [11].

B. StudentLife Dataset

Considering this is a preliminary study the StudentLife dataset provided by Wang et al. [17] was used. The StudentLife continuous sensing app assessed the day-today impact of workload on stress, sleep, activity, mood, sociability, mental well-being and academic performance of a single class of 48 students across a 10 week using Android phones [17]. The whole StudentLife dataset presented in one big file, which contains all the sensor data, EMA data, survey responses and educational data. Figure [TODO] presents the strucure of the StudentLife dataset file.

IV. Information Fusion Using Manhattan Distance

One of the simplest classification algorithms, k -NN measures the distance between the unlabelled observations and the training samples to infer which class they belong to [18]. The unlabelled observation is assigned the label of its nearest neighbours where k is the number of training observations to be taken into account. Distance measures include the Euclidean and Manhattan distance [18]. According to Aggarwal and Hinneburg [19] Manhattan distance (L1 norm) may be preferable to Euclidean distance (L2 norm) for the case of high dimensional data, which is the case with the StudentLife dataset [17].

It is called Manhattan distance because it represents the total distance of streets that we would have to travel between two points, in this case the user type persona and the technology user in the real world.

References

[1] KUSHLEV, Kostadin; PROULX, Jason DE; DUNN, Elizabeth W. Digitally connected, socially disconnected: The effects of relying on technology rather than other people. Computers in Human Behavior, v. 76, p. 68-74, 2017.

[2] KHALILI-MAHANI, Najmeh; SMYRNOVA, Anna; KAKINAMI, Lisa. To each stress its own screen: a cross-sectional survey of the patterns of stress and various screen uses in relation to self-admitted screen addiction. Journal of medical Internet research, v. 21, n. 4, p. e11485, 2019.

[3] ELHAI, Jon D. et al. Depression, anxiety and fear of missing out as correlates of social, non-social and problematic smartphone use. Addictive behaviors, v. 105, p. 106335, 2020.

[4] WANG, Jin-Liang; SHENG, Jia-Rong; WANG, Hai-Zhen. The association between mobile game addiction and depression, social anxiety, and loneliness. Frontiers in public health, v. 7, p. 247, 2019.

[5] RODRÍGUEZ-GARCÍA, Antonio-Manuel; MORENO-GUERRERO, Antonio-José; LOPEZ BELMONTE, Jesus. Nomophobia: An individual's growing fear of being without a smartphone—a systematic literature review. International journal of environmental research and public health, v. 17, n. 2, p. 580, 2020.

[6] ZARA, M. C.; MONTEIRO, L. H. A. The negative impact of technological advancements on mental health: An epidemiological approach. Applied Mathematics and Computation, v. 396, p. 125905, 2021.

[7] LEE, MyungSuk; HAN, MuMoungCho; PAK, JuGeon. Analysis of behavioral characteristics of smartphone addiction using data mining. Applied Sciences, v. 8, n. 7, p. 1191, 2018.

[8] ZHANG, Xingxu et al. Correlation Analysis of Internet Addiction with Daily Behavior: A Data-Driven Method. In: Proceedings of the 2020 3rd International Conference on Big Data Technologies. 2020. p. 41-46.

[9] RYU, Myeong-Un et al. A Smartphone Addiction and Disease Prevention System Through the Collection and Analysis of Smartphone Usage Patterns. Journal of Internet Computing and Services, v. 16, n. 3, p. 95-104, 2015.

[10] PURWANDARI, Betty et al. Internet Addiction and Mental Health Prediction Using Ensemble Learning Based on Web Browsing History. In: Proceedings of the 3rd International Conference on Software Engineering and Information Management. 2020. p. 155-159.

[11] CASTANEDO, Federico. A review of data fusion techniques. The scientific world journal, v. 2013, 2013.

[12] WHITE, F. E. JDL, Data Fusion Lexicon. Technical Panel for C3. FE White, San Diego, Calif, USA, Code, v. 420, 1991.

[13] DURRANT, W. Sensor models and multisensor intergration. The International Journal of Robotics Reseach, v. 7, n. 6, p. 97-113, 1988.

[14] CHOI, Jingyun et al. Smartphone dependence classification using tensor factorization. PloS one, v. 12, n. 6, p. e0177629, 2017.

[15] AHN, Heejune; WIJAYA, Muhammad Eka; ESMERO, Bianca Camille. A systemic smartphone usage pattern analysis: focusing on smartphone addiction issue. Int J Multimed Ubiquitous Eng, v. 9, n. 6, p. 9-14, 2014.

[16] SHIFFMAN, Saul; STONE, Arthur A.; HUFFORD, Michael R. Ecological momentary assessment. Annu. Rev. Clin. Psychol., v. 4, p. 1-32, 2008.

[17] WANG, Rui et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. 2014. p. 3-14.

[18] KING, Rachel C. et al. Application of data fusion techniques and technologies for wearable health monitoring. Medical engineering & physics, v. 42, p. 1-12, 2017.

[19] AGGARWAL, Charu C.; HINNEBURG, Alexander; KEIM, Daniel A. On the surprising behavior of distance metrics in high dimensional space. In: International conference on database theory. Springer, Berlin, Heidelberg, 2001. p. 420-434.