big data and cognitive computing

MDPI

Article

A Dynamic Intelligent Policies Analysis Mechanism for Personal Data Processing in the IoT Ecosystem

Konstantinos Demertzis [1,*], Konstantinos Rantos [1] and George Drosatos [2]

[1] Department of Computer Science, International Hellenic University, 65404 Kavala, Greece; krantos@cs.ihu.gr
[2] Institute for Language and Speech Processing, Athena Research Centre, 67100 Xanthi, Greece; gdrosato@athenarc.gr
* Correspondence: kdemertzis@teiemt.gr; Tel.: +30-694-824-1881

Received: 21 March 2020; Accepted: 20 April 2020; Published: 27 April 2020

Abstract: The evolution of the Internet of Things is significantly affected by legal restrictions imposed for personal data handling, such as the European General Data Protection Regulation (GDPR). The main purpose of this regulation is to provide people in the digital age greater control over their personal data, with their freely given, specific, informed and unambiguous consent to collect and process the data concerning them. ADVOCATE is an advanced framework that fully complies with the requirements of GDPR, which, with the extensive use of blockchain and artificial intelligence technologies, aims to provide an environment that will support users in maintaining control of their personal data in the IoT ecosystem. This paper proposes and presents the Intelligent Policies Analysis Mechanism (IPAM) of the ADVOCATE framework, which, in an intelligent and fully automated manner, can identify conflicting rules or consents of the user, which may lead to the collection of personal data that can be used for profiling. In order to clearly identify and implement IPAM, the problem of recording user data from smart entertainment devices using Fuzzy Cognitive Maps (FCMs) was simulated. FCMs are an intelligent decision-making system that simulates the processes of a complex system, modeling the correlation base, knowing the behavioral and balance specialists of the system. Respectively, identifying conflicting rules that can lead to a profile, training is done using Extreme Learning Machines (ELMs), which are highly efficient neural systems of small and flexible architecture that can work optimally in complex environments.

Keywords: Internet of Things; privacy; GDPR; digital consents management; fuzzy cognitive map; extreme learning machine; feed-forward neural network

1. Introduction

Over the last fifteen years, the emergence of the Internet has led to changes in all areas of everyday life, affecting the lives of ordinary citizens, and within the next decade, the Internet of Things (IoT) revolution will affect the energy, agriculture and transport sectors, as well as the more traditional sectors of the economy and society. In this spirit, we are faced with a phenomenon of great economic and social potential that presents great opportunities but also important challenges associated with unimportant risks, of multidimensional and horizontal nature, which affect businesses and consumers, administrations and citizens alike [1]. For example, the rapid increase in the number of devices capable of collecting personal data from the user's environment is one of the most basic and most serious modern forms of threat to the users' privacy. Participating devices in the IoT ecosystem share information that can be used to track activities and permanently record the status of users and their characteristics or activities. The data in question can be used to make user profiles, with or without their agreement, and enable automated decision-making for third parties [2].

Following the enforcement of General Data Protection Regulation (GDPR), consumers have enhanced their ability to control their personal data and personal preferences, as they can control how the data produced by their IoT devices are used, and to whom and why access to these data is provided [3]. Thus, the legal framework safeguards the absolute right of users to privacy and the protection of their personal data, while providing rights to information access and to be forgotten. However, there are no avoidable breaches related to profiling, with a view to fully automated decision-making for data subjects.

The GDPR on Article 4(4) says that profiling is [4]: "*Any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements*". Specifically, organizations gain personal information about entities from a variety of different sources, such as internet searches, buying behaviors, lifestyle, and activities. This data can be gathered from mobile phones, social networks, video surveillance systems, and the IoT applications and can be analyzed to classify persons into diverse groups that links different characteristics to create profiles for individuals.

Although the majority of Internet users are particularly worried about the exposure of their personal data, since they consider their privacy to be an important issue and want to have control over their personal data, most face threats related to them passively, perhaps even sluggishly. The main reason is their inability to understand the means and techniques used for collecting their personal data, how these data are being used, and most importantly, they are unaware of ways to take active measures to effectively protect their privacy.

In the recent past, several researchers pointed out that users are usually unaware of profiling and tracking activities, which may be carried out by malicious stakeholders even when the user relies on (seemingly) anonymous procedures [5]. For example, profiling issues can be handled well based on privacy-related methods, such as k-anonymity [6] or pseudo-anonymity [7] in the privacy preservation of published data.

Another key aspect involving privacy is that, often, data features multiple individuals or, more specifically, data that seems to involve only certain individuals, in fact, reveal information about others [8]. For instance, in [9], the authors introduced the problem of collaborative privacy on the social networks and provide photo-sharing as a case study. Also, Thomas et al. [10], to handle the problem of multi-party privacy and the conflicting settings between online friends, proposed an approach that takes into consideration the strength of the social ties between online users.

Accordingly, there is a serious gap in applied research and, more generally, in implementations or applications that could help the user obtain meaningful information on how to deal with personal data leakage incidents and how optimal decisions are made with regards to the protection of their personal data, as well as applications that will undertake a thorough analysis of the user's consent, in order to identify possible actions aimed at profiling [11].

This paper presents the Intelligent Policies Analysis Mechanism (IPAM) of the ADVOCATE framework [12–14], which, in an intelligent and fully automated manner, can identify conflicting rules of the user's consents, which may lead to the collection of personal data and, consequently, be used for profiling. The main goal of the proposed approach is the design of an intelligent decision-making system for protecting the privacy of average users. This is achieved by simulating the processes of smart entertainment devices that can collect personal data from the user's environment. The IPAM is an innovative, reliable, low-demand and highly effective system based on sophisticated computational intelligent methods that deliver high-precision results, capable of responding to the problem, as well as in cases of similar complex situations.

The rest of this paper is structured as follows: Section 2 presents the related work in the field. Section 3 introduces the ADVOCATE framework and its main components. Section 4 describes the scenarios and data used in our analysis. Section 5 defines the methodology of Intelligent Policies Analysis Mechanism (IPAM) based on Fuzzy Cognitive Maps (FCMs) and Extreme Learning Machines

(ELMs) methods, while Section 6 presents the results of our approach. Finally, Section 7 concludes the paper and outlines future research objectives.

2. Related Work

An intrinsic characteristic of the IoT is persistent gathering and linkage of user data to provide adapted capabilities. The need to offer users the capability to control the private data produced by IoT devices, is widely recognized [11]. These aspects of the IoT can pose risks to user privacy. Theoretically, aggressive inferences can be drawn from related datasets, comprising data produced through usage of connected devices and services. The GDPR encloses plentiful provisions relevant to the hazards posed by smart identification technologies [15]. Nevertheless, the strict legal requirements well-defined in the GDPR may be inadequate to certify a fair balance among user's interests in privacy and the interests of IoT developers and data controllers [3]. Several research efforts have been made in the direction of providing appropriate solutions for security and privacy in the IoT ecosystem since this is an area that currently attracts many research initiatives [16,17]. For example, in [18], the authors present a JAVA-based security model that handles some of the data security and privacy issues of an open, secure and decentralized system for managing resource sharing among related procedures in the fog-to-cloud (F2C) environment and GDPR, which is called mF2C. The model employs a PKI-based trust model to enable authentication and authorization. It uses policy to certify data privacy and cryptography to provide data confidentiality, integrity, and non-repudiation. However, the proposed framework neither embraces data protection functionalities from the security perspective, nor controls blockchain technology to augment mF2C security and data protection capabilities. In contrast, the ADVOCATE framework proposed by the authors of this paper [12–14], lays the ground for the formation of trust relationships among data subjects and controllers to the GDPR-compliant IoT ecosystem, utilizing blockchain technology to support the integrity, the non-repudiation and the versioning of consents in a publicly verifiable manner.

Moreover, in [19], Subahi and Theodorakopoulos proposed a novel method for ensuring compliance of IoT data disclosure to the agreeing privacy policy. The method was used to analyze the network traffic between the IoT ecosystem and its applications with their distinct Privacy Policy Agreements (PPA). The authors suggest eight norms and compare them with the actual PPA carried out by each IoT subsystem.

Also, there are machine learning based methods for IoT devices with better privacy and security [20–23]. For example, Guntamukkala et al. [24] suggested a machine-learning based method for evaluating the completeness of online privacy policies utilizing an automated approach. The term completeness refers to the presence of eight sections in an online privacy policy that have been recognized as helpful in ensuring the transparency of a privacy policy. The proposed system employs a machine-learning framework to predict a completeness score for the privacy policy and the risk to their privacy. However, the completeness by itself will not warrant transparency, and it may still be uncertain, resulting in an inferior degree of transparency. Therefore, the proposed method needs to be enhanced with the ability to automatically calculate other components that contribute to privacy policy transparency.

In addition, Fuzzy Cognitive Maps (FCMs) are used in the literature to calculate cyber and privacy risks. For example, the aim of the study of the Schläger and Pernul [25] is to show aptitudes for e-commerce organizations or groups using an attribute-based authentication and authorization method to use client data for the derivation of metric trust and reputation values. Using FCMs and trust metrics, the method integrates an organization's user data within an easy to use service provider interface for reputation management. To assure user privacy, the data is categorized and stored in a distributed manner.

ADVOCATE [12–14] addresses the challenges related to privacy protection in the IoT ecosystem, mainly with respect to the supervision of consents as GDPR requires and tries to close a significant gap in this area. It allows users to organize their consents and express their personal data disposal

policies considering the corresponding system recommendations. Similarly, data controllers can utilize ADVOCATE to ensure GDPR-compliance in managing personal data.

3. The ADVOCATE Framework

The ADVOCATE framework [12–14] aims to meet the main requirements of the GDPR with regards to obtaining and managing data subjects' consents. According to the GDPR, consents are data subject's wishes that must be freely given, specific and unambiguous. Moreover, prior to obtaining data subjects' consents, data controllers must appropriately inform them in a transparent manner about (a) personal data they are willing to manage and the sources thereof, (b) the purposes, time periods and legal basis of the processing, (c) the entities that will process it, and (d) recipients or categories of data recipients. ADVOCATE aims to provide users with the means to manage their consents and shape their personal data disposal rules/policies.

An ADVOCATE cloud service approach capable of handling personal data collected by IoT sensors deployed in smart cities and e-health environments is illustrated in Figure 1.

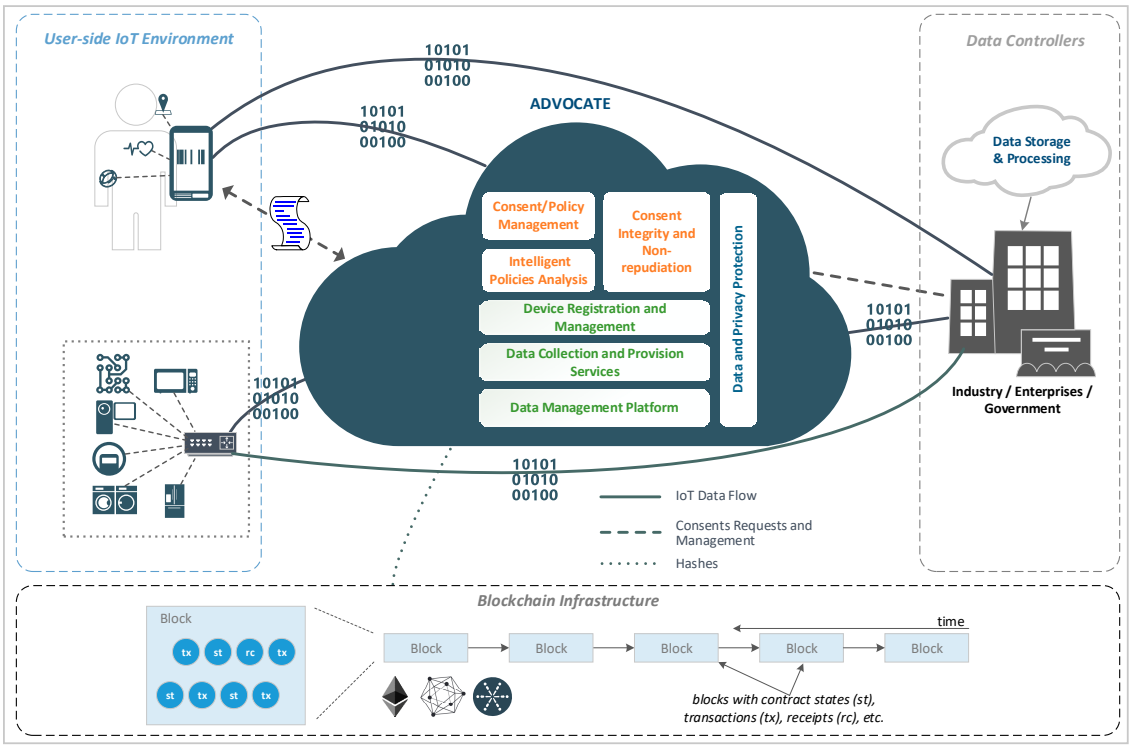

Figure 1. ADVOCATE framework conceptual architecture.

In order to provide immutable versioning control, identifying the latest consents as well as periods that specific consents were valid in an undisputable manner, the ADVOCATE platform uses blockchain technology with smart contracts. This procedure is illustrated in Figure 3.

An interaction of core ADVOCATE components during the introduction of new consents, is presented in Figure 2.

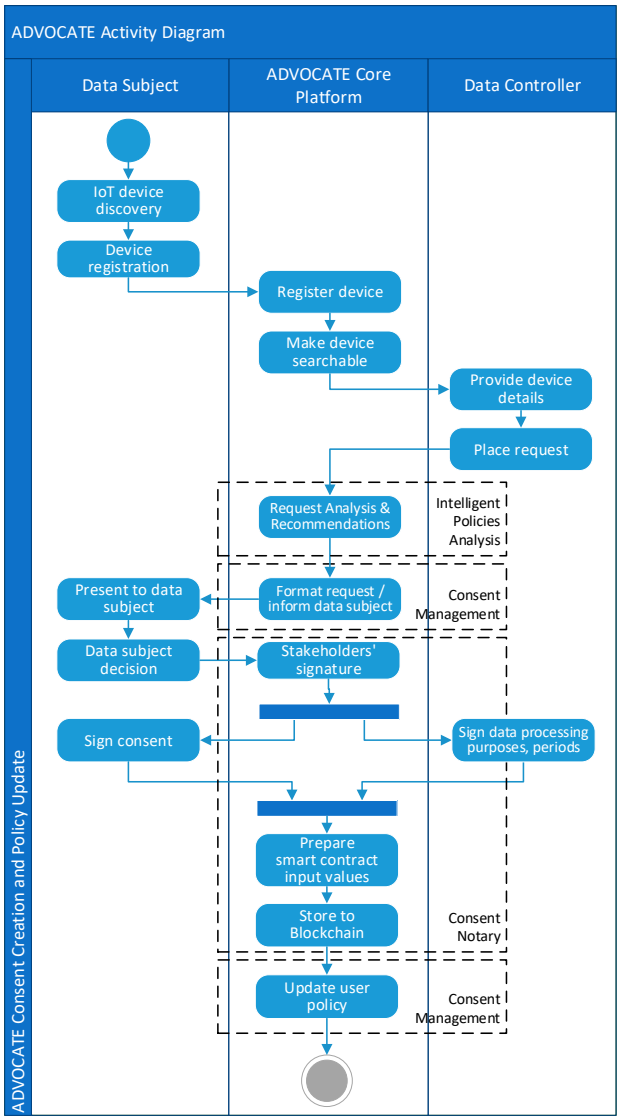

Figure 2. Interaction of ADVOCATE components during the introduction of new consents.

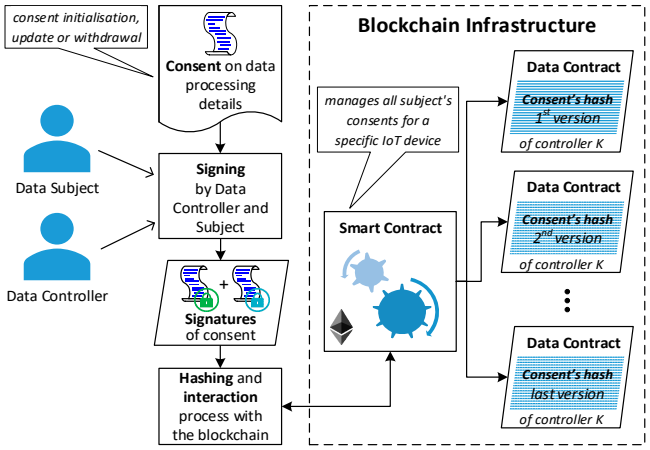

Figure 3. The steps followed by the consent notary component to ensure integrity, non-repudiation, and versioning of given contents.

Finally, the intelligence policies analysis component of the ADVOCATE framework is responsible for performing the necessary analysis for the user's data disposal policy, via an Intelligent Policies Analysis Mechanism (IPAM), to detect conflicting or contradictory rules/policies of a user's consents.

The implementation of IPAM, that is presented in this work, is based on an advanced simulation of the problem of personal data leakage from smart entertainment devices. The problem is modeled using Fuzzy Cognitive Maps (FCMs), which is a technique that belongs to the neuro-fuzzy systems and can be used to solve decision-making problems and model and simulate complex systems [26]. FCMs provide causal acquisition and representation of knowledge, and they can support the causal knowledge reasoning process. The identification process of conflicting rules is accomplished using Extreme Learning Machines (ELMs), which is a very fast and efficient type of Single-Hidden Layer Feed-Forward Neural Network (SHLFFNN) where the parameters of hidden nodes can be randomly assigned and never updated or can be inherited from their ancestors without being changed.

4. Scenarios and Data

Smart entertainment devices greatly upgrade users' home entertainment experience by giving them full control over all the devices of a home ecosystem through voice, applications, and remote control. At the same time, privacy concerns are constantly increasing, arising from the fully accessible and interconnected home environment of smart devices, which produce or record an increasing number of data. These concerns should also be recorded in cases of data collection without the user's full awareness or consent but also by the ignorance of the consequences of their sharing.

In order to clearly identify and implement the proposed IPAM, the problem of recording user data from smart entertainment devices was simulated. After a thorough study of the operation of these devices, bibliographic review and analysis of their manuals, the following variables, which can be collected by smart entertainment devices, were chosen to model the problem of profiling:

User Identity Data

1. Username. The username of the user.
2. Sex. The sex of the user.
3. Age. The age of the user.

Location Data

4. Location_History. If the machine caches the location history of its usage.

Status Data

5. Device_Number. The unique address of the device (mac address).
6. Device_Type. Device type, for example, TV.
7. IP_Address. The IP address of the device.
8. Operating System. The device's OS.

Actionable Data

9. Automation. If the device starts/stops automatically.
10. Average Time. The average time that the device is being used during a day.
11. Average Duration. The average duration of using an application on the device, e.g., Youtube or Netflix.
12. Interests. If the user registers his/her interest on a service, e.g., action movies on Netflix.

Class

13. Yes or No. Automated individual decision-making, including profiling.

All of the above variables are binary data, i.e., either collected (1) or not (0). Table 1 shows some cases of the parameters that make up the problem.

Table 1. Parameters cases.

Username	Sex	Age	Location _History	Device _Number	Device _Type	IP	OS	Automation	Average _Time	Average _Duration	Interests
1	1	1	1	0	1	0	1	1	1	0	0
1	1	1	0	1	1	0	1	1	0	0	0
1	1	0	1	0	1	0	1	1	0	0	0
1	0	0	1	1	1	1	0	0	0	0	0

In fact, the above table depicts the symbolic description and representation of the problem configuration described, which considers the combination of the cases that may arise from the recording or not of the corresponding characteristics. Specifically, and given that our problem contains 12 parameters (username, sex, age, location_history, device_type, device_type, IP, OS, automation, average_time, average_duration, interests) that can receive two possible states, the total dataset includes 2^{12} possible situations, i.e., 4096 cases.

Before proceeding to the problem modeling methodology described in Section 5, it should be noted that the implementation of the scenarios was based solely on a heuristic, approximate method of analysis.

The scenario of the 12 specific parameters with two possible states implements a model that is an abstract representation of the real system. This denotes that it implements only certain properties and characteristics of the real system, without taking into account all of them. The need for selection arises from the fact that real scenarios are extremely complex and cannot be fully represented.

Therefore, the model simplifies reality by choosing a part of it which is suitable for:

1) Giving a general rather than a generalized view of the problem.
2) Focusing on the parameters of how smart entertainment devices work, which is one of the objectives of this research.
3) Creating a technically robust dataset, which can, without serious disadvantages, properly train learning algorithms, while correspondingly being characterized as a relatively high complexity set.
4) Capturing a clear, relatively detailed, and practically possible picture of an unconventional security problem for which no technical details, methodology references, and experimental data are available.

It should be noted that there are exponentially multiple and different cases of variables that can be plotted as system parameters and give a different view of the problem.

The heuristic methodology presented below is indicative of a way of modeling the problem in question. In general, a heuristic technique is any approach to problem-solving that utilizes a practical method without guaranteed to be optimal but is adequate for the immediate goals. Our methodology aims to identify conflicting rules or consents of the user, which may lead to the collection of personal data that can be using the above research design parameters in a certain way to serve this purpose. Specifically, it suggests some basic rules, which are specified as a method of modeling complex systems capable of describing the causal relationships among major concepts that determine the dynamic behavior of the smart entertainment systems. The proposed rules (as an alternative feature selection process) describe the causal relationships among the parameters of the problem we are examining. These causal relationships, which have arisen from existing knowledge and experience, considering the certainty, uncertainty, and risk of each scenario while focusing on its simplicity, illustrates the different aspects of the operation of smart entertainment devices. In order to capture this knowledge, a directed graph is created, with the nodes representing the variables of the problem. An integral part of this methodology is the process of verifying it with tests of validity, reliability, and a range of findings. The main purpose of this is to predict the behavior of the system under the given recorded conditions, or not, of the system parameters.

5. Methodology of IPAM Based on FCM and ELM

The problem of recording user data from smart entertainment devices was modeled using FCMs. FCMs are an alternative method of modeling complex systems capable of describing the causal relationships among major concepts that determine the dynamic behavior of a system. In particular, they form a symbolic description and representation of the formation of a dynamic system. The model concepts that illustrate the different aspects of a system's behavior, as well as how these concepts feedback, either in interaction with each other or in the general dynamics of the system. Human experience and the knowledge of experts who know the functioning of the system and its behavior in different situations are simulated by the use of unclear rules where each rule represents an optimal state or feature of the system.

An FCM [27] consists of a set of neural processing entities called concepts (neurons), C_i, $i = 1, 2, 3, \ldots, N$, where N is the total number of concepts, which are properties of the system to be modeled. The concepts are joined together by links with specific weights, which indicate the effect of the concepts on each other. There are three possible types of causal relationships between two concepts C_i and C_j:

- Positive. It states that an increase or decrease in the value of a *cause* causes the *effect* to move in the same direction and is described with a positive weight W_{ij}.
- Negative. It states that changes in *cause* and *effect* concepts occur in opposite directions, with weight W_{ij} having a negative sign.
- Non-existent. It denotes a zero-weight interface. The weight value, e.g., W_{ij}, describes whether the concept C_i affects the meaning of C_j and is defined in the interval [−1, 1].

At any point in time, the value of each concept A_i is calculated from the sum of the effects of all other concepts and the limitation of the total effect, by using a blocking function f according to the following rule:

$$A_i^{t+1} = f\left(A_i^t + \sum_{i=1,\ i\neq j} W_{ji}A_j^t\right) \tag{1}$$

where A_i^{t+1} and A_i^t are the values of concept C_i at times $t+1$ and t, respectively, A_j^t is the value of concept C_j at time t, W_{ji} is the weight arc from concept C_j to concept C_i and f is a threshold function used to limit the value of the concept to a certain range, typically in the interval [0, 1].

In each step, a new state of the concepts arises and after a certain number of iterations, the FCM may end up either at a particular point of equilibrium or in a limited circle or in chaotic behavior. When the FCM reaches a certain point of equilibrium, it is concluded that the map has converged, and the final state corresponds to the actual state of the system to which it changes when the initial weight values are applied to the map.

The design of FCMs relies heavily on the experience and expertise of specialists who have sufficient knowledge to model a system and provide initial weights for interrelationships among concepts. In more flexible FCMs, these weights are calculated through a training process similar to neural network training methods and algorithms. If the FCM reaches a fixed-point attractor, it is converged. Otherwise, the updating process is terminated after reaching a maximum number of iterations. The FCMs' training algorithm is based on the Hebb Active Learning rule [28].

The Hebb algorithm considers that each node is activated asynchronously. This means that the balance of the map will be achieved by activating different nodes at different times. Therefore, based on this algorithm, the nodes of an FCM are distinguished in nodes that are activated and in nodes being activated. In this case, the node values update rule is modified based on the following function:

$$A_i^{t+1} = f\left(A_i^t + \sum_{i=1,\ i\neq j}^{N} W_{ji}A_j^{act(t)}\right) \tag{2}$$

where the *act* index indicates the enabled node. The rule for weight updates, based on this algorithm, is in the form:

$$w_{ij}^{(t+1)} = \left(1 - \gamma^{(t)}\right) w_{ij}^{(t)} + \eta^{(t)} A_i^{(t)} \left(A_j^{(t)} - w_{ij}^{(t)} A_i^{act(t)}\right) \quad (3)$$

where the learning rate η and the weight-reducing factor in the iteration t are calculated based on the following equations:

$$\eta^{(t)} = b_1 e^{(-\lambda_1 t)} \quad (4)$$

$$\gamma^{(t)} = b_2 e^{(-\lambda_2 t)} \quad (5)$$

where $0.01 < b_1 < 0.09$, $0.1 < \lambda_1 < 1$, while b_2, λ_2 are positive constants selected by test and observation.

Based on the above theoretical background, an FCM has been created that describes the causal relationships among the parameters of the problem we are examining. The causal relationships, which have arisen from existing knowledge and experience, illustrate the different aspects of the functioning of smart entertainment devices and the way in which these concepts are feedback, either in their interactions or in the overall dynamics of the system. To capture this knowledge, a directed graph was created, with the nodes representing the variables of the problem. The main goal of this is to predict the behavior of the system under the given recording conditions or not of the system parameters. Learning has been gained through processes that are the result of continuous interventions, including the modeled and specialists' views until the system is balanced.

In simple terms, the written relationships of the map in Figure 4 outline the human experience and knowledge to solve the problem as a result of continuous observation and intervention [29,30].

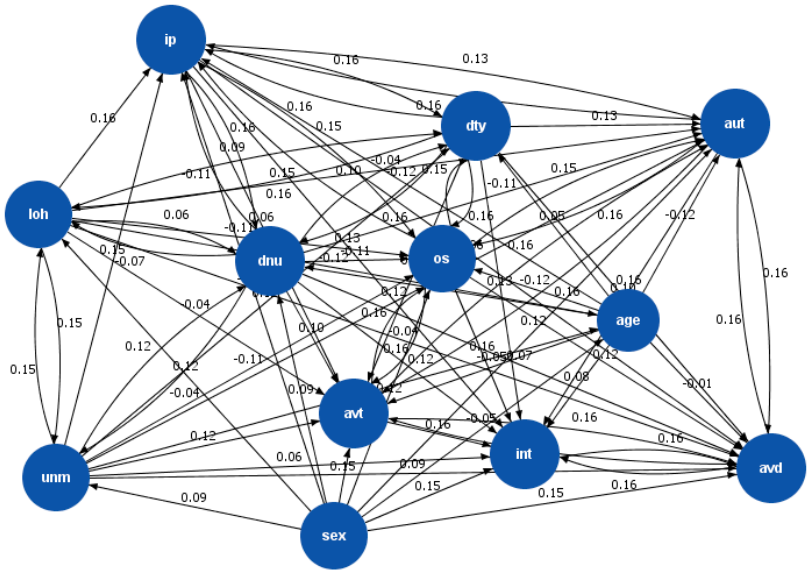

Figure 4. A depiction of the proposed FCM to ADVOCATE framework.

Specifically, FCM concepts like state, feature, concept, neuron, etc., represents knowledge and related variables, states, events, inputs, and outputs in a way that is commensurate with that of human beings. This methodology could support us to develop sophisticated systems, as it is commonly accepted that the more fuzzy and symbolic representation is employed to model a system, the more sophisticated the system is. The FCM concepts it emerged after exhaustive testing (trial and error), taking into account the certainty, uncertainty, and risk of each scenario, while focusing on its simplicity. Trial and error is characterized by repeated, varied attempts which are continued until success. In order to find the best solution by the proposed method, we evaluate each trial model based on the predefined set of criteria, the existence of which is a condition for the possibility of finding an optimal solution.

The above procedure gave us the true inter-relationship among the problem's parameters. Specifically, after a thorough and detailed continuous study through trial and error, and after analogous

thresholds have been set for the expert opinion, it has been realized that in each use scenario, if at least 8 of the parameters to be tested are recorded, we have profiling. Also, there is an increased chance of profiling, more than 95% of cases, when we record at least 3 parameters out of the following: Username, Location_History, IP, Device_Number, and Interests.

This particular revelation of hidden knowledge about this problem has been transformed into the following two rules for creating the classes of the problem, namely:

- The class is Yes (profiling), if there are at least 8 parameters recorded, otherwise it is No (no-profiling).
- The class is Yes (profiling), if at least 3 parameters are recorded from the Username, Location_History, IP, Device_Number and Interests, otherwise they are No (no-profiling).

The combination of the above 2 rules after being applied to the total of 4096 cases attributed 2134 classes Yes and 1962 No. These classes illustrate the problem of user profiling by recording data from smart entertainment devices and can be used for intelligent training standards or algorithms that will allow the immediate information of the user and optimal decision-making on the security of their personal data and their protection against profiling.

Similarly, as stated above, in theory there are exponential multiple and different cases that can be mapped as rules and that can give a different view of the problem. The methodology followed in this paper is indicative and focuses on the simplicity of implementing a solid and valuable synthetic dataset appropriate to train the learning algorithms used in this research. Table 2 presents the Table 1 after applying the above rules.

Table 2. Table 1 after applying the rules.

Username	Sex	Age	Location _History	Device _Number	Device _Type	IP	OS	Automation	Average _Time	Average _Duration	Interests	Class
1	1	1	1	0	1	0	1	1	1	0	0	Yes
1	1	1	0	1	1	0	1	1	0	0	0	No
1	1	0	1	0	1	0	1	1	0	0	0	No
1	0	0	1	1	1	1	0	0	0	0	0	Yes

Specifically, in the first case, we have the recording (1), i.e., 8 different parameters, so we automatically have the case of profiling. In the second and third cases we have, respectively, 7 and 6 records (1), which do not fall under the second rule, i.e., 3 of the Username, Location_History Device_Number, IP and Interests are not recorded. Finally, in the fourth case, although we have the recording (1) of only 5 cases, recording of these four parameters (Username, Location_History, Device_Number and IP) allows user profiling.

Once the problem classes have been created in the dataset modeling our problem, the identification of conflicting rules that can lead to profiling is done using Extreme Learning Machines (ELMs) [31]. ELMs are Single-Hidden Layer Feed-Forward Neural Networks (SHLFFNNs), with N neurons at hidden layer, randomly selected input weights, and random constant polarization values in the hidden layer neurons, while weights at its output are calculated by a single multiplication of tables. SHLFFNNs are used in ELMs because they can handle any continuous function and classify any non-continuous areas, with the ability to accurately read K samples, while their learning speed can be even thousands of times greater than the speed of Conventional Neural Networks (CNNs) which are trained using the Back-Propagation method. This characteristic is due to the general view, which is demonstrated in the theoretical background of ELMs, that SHLFFNN's hidden layer (feature mapping) is not required to work in a coordinated fashion, hidden neurons may have been created randomly and all network parameters are independent of activation functions and training data. It is also noteworthy that ELMs can handle non-differential activation equations and do not address known neural network problems such as the definition of the appropriate stopping criterion, the learning rate, and learning epochs.

In ELMs, the weights of input level W and biases β are randomly set and not adjusted. Because the input weights are fixed, the output weights b are independent of them, as opposed to the

Back-Propagation method, so that to produce an immediate solution without repetition. Essentially, random weights at the input level improve the generalization properties of a linear system because they produce weakly related properties at the hidden layer. Given that the output of a linear system is always correlated with the input data, if the weight range of the solution is limited, the rectangular inputs provide a wider range of solutions than those supported by the weights. Also, small variations in weights allow the system to become more stable and noise-resistant as input errors will not be amplified at the output of a linear system with little correlation between input and output weights. Thus, the random ranking of weights, which produces loosely coupled characteristics at the hidden layer, allows for a satisfactory solution and good generalization performance to be achieved. Corresponding to a linear output level, the Generalized Least Squares Approximation [32] approach is used, where such a solution is also linear and very fast to calculate.

Thus, the random weighting, which produces weakly correlated features at the hidden layer, allows for a satisfactory solution and a good generalization performance. Respectively, for a linear output layer, the Generalized Least Squares Approximation approach is used, where such a solution is also linear and very fast to calculate.

Specifically, in an ELM using SHLFFNN and random representation of the hidden layer neurons, the input data are mapped to a random L-dimensional space with a discrete set of training N, where (x_i, t_i), $i \in [\![1, N]\!]$ με $x_i \in R^d$ *and* $t_i \in R^c$. The output of the network can be represented as follows [29]:

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \; i \in [\![1, N]\!] \tag{6}$$

where $\beta = [\beta_1, \ldots, \beta_L]^T$ is the output of the weights matrix between the hidden nodes and the output nodes, $h(x) = [g_1(x), \ldots, g_L(x)]$ are the exits of hidden nodes (random hidden attributes) for input x, and $g_1(x)$ is the output of the hidden node i. Based on dataset $\mathrm{N}\{(x_i, t_i)\}_{i=1}^{N}$, an ELM can solve the learning problem $H\beta = T$, where $T = [t_1, \ldots, t_N]^T$ are the target tags and H the output panel of the hidden layer shown below:

$$H\left(\omega_j, b_j, x_i\right) = \begin{bmatrix} g(\omega_1 x_1 + b_1) & \cdots & g(\omega_l x_1 + b_l) \\ \vdots & \ddots & \vdots \\ g(\omega_1 x_N + b_1) & \cdots & g(\omega_l x_N + b_l) \end{bmatrix}_{N \times l} \tag{7}$$

Prior to the training, the input weights table ω and the biases β are randomly generated in the space [−1, 1], with $\omega_j = \left[\omega_{j1}, \omega_{j2}, \ldots, \omega_{jm}\right]^T$ and $\beta_j = \left[\beta_{j1}, \beta_{j2}, \ldots, \beta_{jm}\right]^T$. The output weights table of the hidden level H is calculated by the activation function and the training data based on the following function:

$$H = g(\omega x + b) \tag{8}$$

The output weights B can be calculated from the relation:

$$\beta = \left(\frac{}{C} + H^T H\right)^{-1} H^T X \tag{9}$$

where $H = [h_1, \ldots, h_N]$ are the outputs of the hidden layer and $X = [x_1, \ldots, x_N]$ are the input data. B can also be calculated from the general relation:

$$\beta = H^+ T \tag{10}$$

where H^+ is the Moore–Penrose generalized inverse matrix for matrix H.

The hidden layer is responsible for transforming the input data into a different representation. The transformation is a two-stage process:

1. Input data enters the hidden layer, using weights and corresponding biases of the input layer.
2. The data is transformed based on a non-linear transformation function.

The hidden layer, and in particular, the second stage of input data transformation, is the only point where a non-linear process is used throughout the method, which significantly increases the learning abilities of ELMs. After the transformation, the data in the hidden layer are used to find the weights of the output layer. The construction of hidden-layer neurons in a random manner is an ingenious architectural technique for fast learning, which essentially eliminates the problem of overfitting.

At the hidden layer there is no explicit constraint and various transformation functions can be used, such as sigmoidal, tangential, and threshold. Nonetheless, there are some linear neurons that do not need to use any transformation function as they are capable of learning correlations between the input data and the target data directly, without being approximated by a nonlinear function. Usually in the case of linear neurons, these neurons are equal the number of input data, at a 1 to 1 match.

The exceptional characteristics of ELMs [30], such as efficiency, speed, and optimal generalization capabilities, have been shown to have a wide range of problems from different disciplines, with often comparable or even better results than those of deep learning algorithms. Another important fact that enhances the use of ELMs is that they work best when the input patterns are from the same boundary distribution or follow a common cluster structure, as in the case we are examining.

6. Results

In the case of multi-class classifiers, the full probability density of both classes should be known for estimating the actual error during training. The calculation of the classification performance is calculated by constructing a Confusion Matrix (CM), where the numbers of misclassifications are related to the False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN). The Total Accuracy (TA) is defined by using the equation [33]:

$$\mathrm{TA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{N}} \tag{11}$$

Equations (12), (13), and (14) below, define the Precision (PRE), the Recall (REC), and the F-Score indices, respectively:

$$\mathrm{PRE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{12}$$

$$\mathrm{REC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{13}$$

$$\mathrm{F-Score} = 2 \cdot \frac{\mathrm{PRE} \cdot \mathrm{REC}}{\mathrm{PRE} + \mathrm{REC}} \tag{14}$$

The Root Mean Square Error (RMSE) is defined by using the equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(P_{(ij)} - T_j\right)^2} \tag{15}$$

Also, a Receiver Operating Characteristic (ROC) area is a performance metric of all classification thresholds.

Table 3 presents the results of the method proposed in this study, as well as the results of the corresponding methods tested in the dataset under consideration.

Table 3. Algorithm comparison providing classification accuracy and performance metrics.

Classifier	TA	RMSE	PRE	REC	F-Score	ROC Area	Training Time	Validation
ELM	99.87%	0.0353	0.999	0.999	0.999	0.999	1.1 s	10-FCV
SVM	97.92%	0.1441	0.980	0.979	0.979	0.980	14.9 s	10-FCV
k-NN	97.53%	0.2306	0.976	0.975	0.975	0.994	15.3 s	10-FCV
Random Forest	98.17%	0.1424	0.982	0.982	0.982	0.996	15.6 s	10-FCV

As can be seen, the use of ELM creates a robust learning system that effectively solves a realistic, multifactorial, and highly complex problem. The choice of the proposed model is significantly superior to the algorithms compared, both in categorization metrics and in training time, where ELM is almost 15 times faster than the other models. This feature is a clear demonstration of the potential of the proposed method, taking into account the difficulties of the research environment.

In conclusion, it should be emphasized that ELM implements, in the most realistic way, the process of identifying conflicting rules that can lead to profiling, as the training of the algorithm was based on a highly specialized initial set of training, but in which the proposed system managed to generalize and respond optimally to unspecified situations.

It is also important to highlight that the stability of the ELM and the clear identification it offers in the modeling of complex systems offered by FCMs advocate the adoption of a method that delivers high-precision results, capable of responding to the problem, as well as in cases of similar complex situations.

7. Conclusions

An innovative, reliable, low-demand, and highly effective system for detecting conflicting consents and user consent policies regarding personal data, based on sophisticated computational methods, was presented in this work. The Intelligent Policies Analysis Mechanism (IPAM) implements sophisticated conflict recognition rules to assist the average user in making an optimal decision and protecting them from impending profiling, with computational intelligence methods. Specifically, using Fuzzy Cognitive Maps (FCMs), which is an advanced form of neuro-fuzzy decision support system, the protection of user privacy in the IoT ecosystem is modeled in the most efficient and intelligent way. Respectively, with the use of Extreme Learning Machines (ELMs), which are highly efficient neural systems of small and flexible architecture, which can operate optimally in complex environments, it is possible to identify the conflicting rules that may result in profiling.

This proposed mechanism, which is an integral part of the ADVOCATE framework [12–14], greatly enhances its security mechanisms and is a promising intelligent mechanism for protecting the privacy of users whose personal data are the main target of modern cyberattack methods.

It is important to note that the application of artificial intelligence to methods of controlling and protecting users' personal data significantly enhances the active security mechanisms of these methods and creates new perspectives on how to deal with cybercrime. It is also important to emphasize that the complexity of the IoT ecosystem, the uncertainty it brings, as well as the instability of other learning algorithms in such a dynamic environment, favor the adoption of a method that normalizes the noisy field and brings consistent results capable of modeling serious, multi-dimensional problems.

On the other hand, literature points out how the side-channel information can be used to extract some of the sensitive data that the GDPR tries to protect [34]. For example, Palmieri et al. [35] present a private routing protocol for anonymous communication between different networks (e.g., wireless sensor networks, etc.) using technologies such as Spatial Bloom Filters (SBF), tunneling, and homomorphic encryption. The proposed routing protocol preserves context privacy and prevents adversaries from discovering the network topology and structure, as routing information is encrypted and computed by performing calculations in the actual encrypted data. Consequently, the achieved privacy is vital in preventing adversaries from obtaining valuable network information from a successful attack on a single node of the network and reduces the likelihood of an escalation attack.

Future study could include additional behavior analysis of the smart entertainment devices, in order to be improved by additional adjusting of the parameters of the suggested framework, so that an even more effective, precise, and faster classification process could be reached. Multi-format exemplifications may support a reporting system as part of a general decision framework. In addition, more sophisticated methods could be used for precise identification of contradictory and conflicting policies. Also, it would be essential to study the expansion of this method by applying the same architecture in a big data architecture framework like Hadoop [36]. In conclusion, an additional component that could be considered in the way of future development concerns the process of the proposed framework with methods of self-adaptive improvement in order to fully automate the IPAM in contrast to a personal data breach.

Author Contributions: Conceptualization, K.D., K.R., G.D.; investigation, K.D., K.R., G.D.; writing—original draft preparation, K.D., K.R.; writing—review and editing, K.D., K.R., G.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We would like to show our gratitude to Gonzalo Napoles Ruiz, at Tilburg University for sharing his FCM Expert software and for his comments on an early draft of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shahid, N.; Aneja, S. Internet of Things: Vision, application areas and research challenges. In Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 10–11 February 2017; pp. 583–587.
2. AboBakr, A.; Azer, M.A. IoT Ethics Challenges and Legal Issues. In Proceedings of the 2017 12th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 19–20 December 2017; pp. 233–237.
3. Vegh, L. A Survey of Privacy and Security Issues for the Internet of Things in the GDPR Era. In Proceedings of the 2018 International Conference on Communications (COMM), Bucharest, Romania, 14–16 June 2018; pp. 453–458.
4. General Data Protection Regulation (GDPR)—Official Legal Text. Available online: https://gdpr-info.eu (accessed on 11 April 2020).
5. Avoine, G.; Calderoni, L.; Delvaux, J.; Maio, D.; Palmieri, P. Passengers information in public transport and privacy: Can anonymous tickets prevent tracking? *Int. J. Inf. Manag.* **2014**, *34*, 682–688. [CrossRef]
6. Bilenko, M.; Richardson, M. Predictive Client-Side Profiles for Personalized Advertising. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; ACM Press: New York, NY, USA; pp. 413–421.
7. Ceri, S.; Dolog, P.; Matera, M.; Nejdl, W. Model-Driven Design of Web Applications with Client-Side Adaptation. In Proceedings of the International Conference on Web Engineering, Munich, Germany, 26–30 July 2004; Koch, N., Fraternali, P., Wirsing, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3140, pp. 201–214.
8. Humbert, M.; Trubert, B.; Huguenin, K. A Survey on Interdependent Privacy. *ACM Comput. Surv.* **2020**, *52*, 1–40. [CrossRef]
9. Squicciarini, A.C.; Shehab, M.; Paci, F. Collective Privacy Management in Social Networks. In Proceedings of the 18th International Conference on World Wide Web (WWW '09), Madrid, Spain, 20–24 April 2009; ACM Press: New York, NY, USA; p. 521.
10. Thomas, K.; Grier, C.; Nicol, D.M. unFriendly: Multi-party Privacy Risks in Social Networks. In Proceedings of the International Symposium on Privacy Enhancing Technologies, Berlin, Germany, 21–23 July 2010; Atallah, M.J., Hopper, N.J., Eds.; Springer: Berlin/Heidelberg Germany, 2010; pp. 236–252.
11. Wachter, S. Ethical and Normative Challenges of Identification in the Internet of Things. In Proceedings of the Living in the Internet of Things: Cybersecurity of the IoT-2018, London, UK, 28–29 March 2018; pp. 1–10.
12. Rantos, K.; Drosatos, G.; Demertzis, K.; Ilioudis, C.; Papanikolaou, A. Blockchain-based Consents Management for Personal Data Processing in the IoT Ecosystem. In Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018), Porto, Portugal, 26–28 July 2018; Volume 2, pp. 572–577.

13. Rantos, K.; Drosatos, G.; Demertzis, K.; Ilioudis, C.; Papanikolaou, A.; Kritsas, A. ADvoCATE: A Consent Management Platform for Personal Data Processing in the IoT Using Blockchain Technology. In *Innovative Security Solutions for Information Technology and Communications*; Lanet, J.-L., Toma, C., Eds.; Springer International Publishing: Cham, Germany, 2019; pp. 300–313.
14. Rantos, K.; Drosatos, G.; Kritsas, A.; Ilioudis, C.; Papanikolaou, A.; Filippidis, A.P. A Blockchain-Based Platform for Consent Management of Personal Data Processing in the IoT Ecosystem. *Secur. Commun. Netw.* **2019**, *2019*, 1431578. [CrossRef]
15. Hernández-Serrano, J.; Muñoz, J.L.; León, O.; Mikkelsen, L.; Schwefel, H.-P.; Bröring, A. Privacy risk analysis in the IoT domain. In Proceedings of the 2018 Global Internet of Things Summit (GIoTS), Bilbao, Spain, 4–7 June 2018; pp. 1–6.
16. Kraijak, S.; Tuwanut, P. A Survey on IoT Architectures, Protocols, Applications, Security, Privacy, Real-World Implementation and Future Trends. In Proceedings of the 11th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2015), Shanghai, China, 21–23 September 2015; pp. 1–6.
17. Li, C.; Palanisamy, B. Privacy in Internet of Things: From Principles to Technologies. *IEEE Internet Things J.* **2019**, *6*, 488–505. [CrossRef]
18. Crompton, S.; Jensen, J. Towards a Secure and GDPR-Compliant Fog-to-Cloud Platform. In Proceedings of the 2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion), Zurich, Switzerland, 17–20 December 2018; pp. 296–301.
19. Subahi, A.; Theodorakopoulos, G. Ensuring Compliance of IoT Devices with Their Privacy Policy Agreement. In Proceedings of the 2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud), Barcelona, Spain, 6–8 August 2018; pp. 100–107.
20. Imtiaz, S.; Sadre, R.; Vlassov, V. On the Case of Privacy in the IoT Ecosystem: A Survey. In Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; pp. 1015–1024.
21. Cañedo, J.; Skjellum, A. Using machine learning to secure IoT systems. In Proceedings of the 14th Annual Conference on Privacy, Security and Trust (PST), Auckland, New Zealand, 12–14 December 2016; pp. 219–222.
22. Shen, M.; Tang, X.; Zhu, L.; Du, X.; Guizani, M. Privacy-Preserving Support Vector Machine Training Over Blockchain-Based Encrypted IoT Data in Smart Cities. *IEEE Internet Things J.* **2019**, *6*, 7702–7712. [CrossRef]
23. Jeong, H.-J.; Lee, H.-J.; Moon, S.-M. Work-in-Progress: Cloud-Based Machine Learning for IoT Devices with Better Privacy. In Proceedings of the 2017 International Conference on Embedded Software (EMSOFT), Seoul, Korea, 15–20 October 2017; pp. 1–2.
24. Guntamukkala, N.; Dara, R.; Grewal, G. A Machine-Learning Based Approach for Measuring the Completeness of Online Privacy Policies. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015; pp. 289–294.
25. Schläger, C.; Pernul, G. Trust Modelling in E-Commerce through Fuzzy Cognitive Maps. In Proceedings of the 2008 Third International Conference on Availability, Reliability and Security, Barcelona, Spain, 4–7 March 2008; pp. 344–351.
26. Salmeron, J.L. Fuzzy cognitive maps for artificial emotions forecasting. *Appl. Soft Comput.* **2012**, *12*, 3704–3710. [CrossRef]
27. Felix, G.; Nápoles, G.; Falcon, R.; Froelich, W.; Vanhoof, K.; Bello, R. A review on methods and software for fuzzy cognitive maps. *Artif. Intell. Rev.* **2019**, *52*, 1707–1737. [CrossRef]
28. Papageorgiou, E.I.; Stylios, C.D.; Groumpos, P.P. Active Hebbian learning algorithm to train fuzzy cognitive maps. *Int. J. Approx. Reason.* **2004**, *37*, 219–249. [CrossRef]
29. Nápoles, G.; Espinosa, M.L.; Grau, I.; Vanhoof, K. FCM Expert: Software Tool for Scenario Analysis and Pattern Classification Based on Fuzzy Cognitive Maps. *Int. J. Artif. Intell. Tools* **2018**, *27*, 1860010. [CrossRef]
30. Nápoles, G.; Leon, M.; Grau, I.; Vanhoof, K. Fuzzy Cognitive Maps Tool for Scenario Analysis and Pattern Classification. In Proceedings of the 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, USA, 6–8 November 2017; pp. 644–651. [CrossRef]
31. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]

32. Strutz, T. *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond*, 2nd ed.; Springer Vieweg: Berlin/Heidelberg, Germany, 2015; ISBN 978-3-658-11455-8.
33. Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63. [CrossRef]
34. Lisovich, M.A.; Mulligan, D.K.; Wicker, S.B. Inferring Personal Information from Demand-Response Systems. *IEEE Secur. Priv.* **2010**, *8*, 11–20. [CrossRef]
35. Palmieri, P.; Calderoni, L.; Maio, D. Private inter-network routing for Wireless Sensor Networks and the Internet of Things. In Proceedings of the Computing Frontiers Conference (CF '17), Siena, Italy, 15–17 May 2017; ACM Press: New York, NY, USA, 2017; pp. 396–401.
36. Hodge, V.J.; O'Keefe, S.; Austin, J. Hadoop neural network for parallel and distributed feature selection. *Neural Netw.* **2016**, *78*, 24–35. [CrossRef] [PubMed]