Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-5, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

Predicting User's Age from Their Web Behaviour

Vaishali Pujari, Abhijeet Gavhane, Vinay Lala, Harshad Shinde & Aman Dudhal
G.H.Raisoni Institute of Engineering and Technology, Pune

Abstract— *An online user can be categorized on some particular entities but majorly his age can be the defining entity of his behavior and can be predicted easily using an Artificial Neural Network. Neural networks is used because it requires less formal statistical training and has ability to implicitly detect complex nonlinear relationships between dependent and independent variables. We treat the user's web browsing behavior as a variable to propagate the age information between different users. The web content is filtered by the ANN depending on the user's age. The user is divided into three categories namely Teenager, Adult and Senior Citizen. The system is around 93% accurate.*

Keywords— *ANN, Age Detection, Web Behaviour*

1. Introduction

The project aims to implement a system of automatically blocking web content and thus provide parental control. Also implement a system where products are recommended not pattern and his age. This will help in providing a more accurate collection of products and services to user[1].

An important example is that every time when we search for a product on an online shopping portal and proceed to buy it then later we get notifications suggesting more products and offers. These products and offers are suggested on the basis of your purchase and search history, which at many times can be irrelevant to the end user. If the end user is to be made interested in the product then it's important that the product excites the user and this can be ensured if it is suitable for his age. So if the end user is a teenager he will be suggested products like T-shirts; Casual Shoes; Music: EDM, Rock; Movies: The Hunger Games, Iron Man; Books: Harry Potter, Percy Jackson. Also unsuitable content will be filtered.

If end user is Adult he will be suggested Formal Shirts; Jobs; Music: Jazz, 70's; Movies: Star Wars, The Godfather;

Books: Game of Thrones, Da Vinci Code.

If end user is a Senior Citizen he will be suggested Medicines; Devices for measuring Blood Pressure, Sugar; Music: Old Music, Bhajans(Religious songs); Movies: Mackenna's Gold, Good Bad and Ugly; Books: Ramayana, Bhagavad Gita.

An Artificial Neural Network is used to implement the project. Artificial neural network (ANN) is a machine learning approach that models human brain and consists of a number of artificial neurons. Neuron in ANNs tend to have fewer connections than biological neurons. Each neuron in ANN receives a number of inputs. An activation function is applied to these inputs which results in activation level of neuron (output value of the neuron).Knowledge about the learning task is given in the form of examples called training examples[4].

2. Literature Survey

Use of Neural Networks: Neural networks, with their notable capability to develop meaning from complex or inaccurate data, can be used to extract patterns and identify trends that are too complicated to be noticed by either humans or other processing techniques. A trained neural network can be thought of as an "expert" in the type of information it has been given to analyze. This expert can then be used to offer projections given new situations of importance and answer "what if" questions[4].

The Advantages of ANN are:

Adaptive learning: A skill to learn how to do tasks based on the data given for training or early experience.

Self-Organization: An ANN can create its own association or version of the information it receives during learning time.

Real Time Operation: ANN computations may be carried out in parallel, and unique hardware devices are being designed and manufactured which can take full advantage of this capability.

Fault Tolerance via Redundant Information Coding: Limited destruction of a network leads to the equivalent degradation of performance. On the other hand, some network capabilities may be retained even with major network damage[5].

Neural Networks opposed to Conventional Computers: Neural networks take a diverse approach

to problem solving than that of conventional computers. Conventional computers make use of an algorithmic approach in which the computers follow a set of instructions in order to resolve a problem. Unless the exact steps that the computer needs to follow are known the computer cannot resolve the problem. That restricts the problem solving potential of conventional computers to problems that we already understand and know how to solve. Computers would be so much more practical and useful if they could do things that we don't exactly know how to do. Neural networks process data in a similar way the human brain does. The network is a collection of a large number of highly interconnected and organized processing elements (neurons) working in parallel to solve a specific problem. Neural networks are trained by example. They can't be programmed to carry out a specific task. The examples must be chosen carefully or else useful time is wasted or yet worse the network might be functioning imperfectly. The drawback is that because the system finds out how to solve the problem by itself, its operation can be volatile. On the other hand, conventional computers use a cognitive approach to problem solving; the way the problem is to be solved should be identified and declared in small explicit instructions. These instructions are then transformed to a high level language program and then into machine code that the computer can comprehend. These machines are completely predictable; if anything is erroneous is due to a software or hardware fault. Neural networks and conventional algorithmic computers are not in competition but balance each other. There are tasks, more suitable to an algorithmic approach like arithmetic operations and tasks that are more suitable to neural networks like machine learning. Especially, a large number of tasks require systems that use a combination of the two approaches (generally a conventional computer is used to administer the neural network) in order to perform at highest efficiency[5].

The Neuron: It is the fundamental building block of the neural network. A neuron is a communication channel that both accepts input and provides output. The neuron gets its input either from other neurons or the user program. Likewise, the neuron sends its output to other neurons or the user program.

The most common type of artificial neural network consists of three groups, or layers: a layer of "input" units is linked to a layer of "hidden" units, which is linked to a layer of "output" units. The activity of the input units represents the unprocessed information that is fed into the network. The action of each hidden unit is determined by the behavior of the input units and the weights on the connections between the input and the hidden layer. The performance of the output layer depends on the activity of the hidden units as well as the weights between the hidden and output units. This simple type of network is fascinating because the hidden units are free to construct their own representations of the input. The weights connecting the input and hidden units decide when each hidden unit is active, and so by changing these weights, a hidden unit can decide what it represents.

Neuron Connection Weights: The connections stated formerly are not the same, and can be assigned individual weights. These weights are what grant the neural network the capacity to identify certain patterns. Adjust the weights, and the neural network will distinguish a different pattern. Alteration of these weights is a very significant operation[3].

3. Algorithm

The memorization of patterns and the following response of the network can be categorized into two broad paradigms: Associative mapping in which the network learns to create a particular prototype on the set of input units whenever another particular prototype is applied on the set of input units. The associative mapping can usually be broken down into two types:

Auto-association: An input pattern is coupled with itself and the states of input and output units coincide. This is used to offer pattern completion that is to produce a pattern every time a portion of it or a distorted pattern is presented. In the next case, the network essentially stores pairs of patterns building a relationship between two sets of patterns[6].

Hetero-association: is related to two recall mechanisms i.e. nearest-neighbor recall, where the output pattern produced corresponds to the input pattern stored, which is closest to the pattern on hand, and Interpolative recall, in which the output pattern is a similarity dependent interpolation of the patterns stored matching to the pattern on hand.Yet another model, which is a variant associative mapping, is classification, that is when there is a unchanging set of categories into which the input patterns are to be classified. Regularity detection in which units learn to react to particular properties of the input patterns. While in associative mapping the network stores the associations among patterns, in regularity detection the response of each unit has a actual 'meaning'. This type of learning mechanism is vital for feature detection and knowledge representation. All neural network possesses knowledge which is stored in the values of the connections weights. Modifying the knowledge stored in the network as a function of understanding and experience implies a learning rule for altering the values of the weights[6]. All learning methods used for adaptive neural networks can be classified into two key categories:

Supervised learning: which incorporates an external trainer, so that every output unit is told what its

desired response to input signals must be. During the learning process global information may be necessary. Paradigms of supervised learning comprise error correction learning, reinforcement learning and stochastic learning. An essential issue regarding supervised learning is the difficulty of error convergence, i.e. the minimization of error between the preferred and computed unit values. The aim is to find out a set of weights which minimizes the error. One well-known technique, which is common to many learning paradigms, is the least mean square (LMS) convergence.

Unsupervised learning: It uses no external trainer and is based solely upon local information. It is also referred to as self-organization, in the sense that it self-organizes information presented to the network and detects their growing collective properties. Types of unsupervised learning are Hebbian learning and competitive learning. A neural network learns on-line if it learns and operates in parallel. Typically, supervised learning is performed off-line, whereas unsupervised learning is performed on-line.[1]

4. Proposed System

The system will mainly be based on

- Neural Network Training: A Data set can be created which will comprise of some pre-defined data sets and will train the neural network. This process is called Learning Phase and the ANN implements it via back propagation algorithm.
- Input Layer: The ANN takes input from the database created by system which is given to hidden layer. We have two neurons in input layer[4].
- Hidden Layer: The number of hidden layers must be selected in such a way that it not only reduces complexity but also increases it its computational power. As we have three age categories the number of neurons in hidden layer is 3.
- Output Layer: The number of nodes in the output layer is decided according to the application's output. Since the neural network is used to predict the age group of the user the output node is one.
- Back Propagation Formula: The Back Propagation is done using[6] the following formula:-

$$W_{ij}=W_{ij}+\Delta W_{ij}$$

$$\Delta W_{ij} = \frac{-\eta \delta E}{\delta W ij}$$

Presently as everyone is online it is very important to provide them content which is suitable for them in a timely manner.
Some users tend to browse internet content which can be different from than their real age. Our project aims to provide them web content and recommend them services depending not only on their history but their browsing pattern.
We try and detect the mental age of the user and suggest them movies, music, health products etc.

5. Implementation

We implement the website using certain modules:
1) Login-
This is used by all type of users. They need to create an account in which they have to provide email id and password. This will enable administrator to monitor the content being accessed by user. If user does not have an account he has to register with the website.
2) Search - This can check the entire website and return the product/service which the user has searched for. ANN stores the search word and depending on the word user is categorized into age category. Depending on the detected age of the user certain products will get a preference and be displayed on a higher priority.
3) ANN – This would act as the brain of the system where all the calculations of user age is carried out. We design 3 arrays consisting a set of keywords each pertaining to certain age category. Every time user hits one of those keywords his/her age is adjusted to that particular category. Through Back propagation and training the ANN, after some usage we can detect age of user.
The other modules like Mail and Download Files provide additional services for the user which can be monitored to identify the age. The files can be songs, videos or documents.

6. Architecture

The following architecture shows how the system is implemented

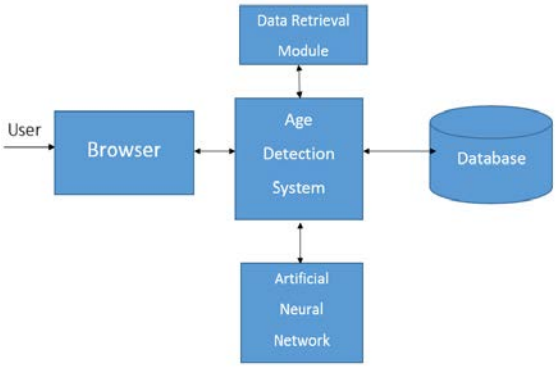

Fig 1: Architecture of the system

The components are:-
Browser: It is the way the user interacts with the system. There will be a web portal where user needs to sign up and get the recommended web content.

Data Retrieval Module: This collects the data and processes it so that it can be easily stored in the database and provide faster access to the system.
Artificial Neural Network: As stated earlier an ANN is used to calculate and detect user's age. The architecture used is a 2-3-1 type.
Database: It comprises of all the data which is to be displayed on the website which is stored on the server. The larger the database the better the chances of getting the user involved in the website and get a higher probability of predicting his exact age. There will be an Admin with full access to database[7].

7. Future Scope

There is potential to increase the accuracy of the system and approach 100% accuracy. This would require extensive programming of the ANN and also extensive user interaction with the website.
In the future a system can be implemented in which along with age the gender of user can also be detected [2].

8. Conclusions

Throughout this paper, a technique has been proposed to predict the internet user's age depending on their browsing pattern using Artificial Neural Network. It establishes suitability of non-linear ANN as analytical tool for internet user demography. Final result gives an accuracy of around 93%. When implemented as a website, this tool can be used to restrict teenagers from accessing censored contents online.

9. Acknowledgements

We would like to thank our project guide Mrs.Vaishali Pujari, Assistant Professor at G. H. Raisoni Institute of Engineering & Technology, Wagholi, Pune.

10. References

[1] Misha Kakkar, Divya Upadhyay , "*Web Browsing Behaviours Based Age Detection*", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-1, March 2013.

[2] Indre Zliobaite and Bogdan Gabrys, "*Adaptive Preprocessing for Streaming Data*", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 2, FEBRUARY 2014.

[3] Ms.Dharmistha D.Vishwakarma, "*Genetic Algorithm based Weights Optimization of Artificial Neural Network*", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 1, Issue 3, August 2012.

[4] B Yegnanarayana, "*Artificial neural networks for pattern recognition*", Scidhanci, Vol. 19, Part 2, April 1994, pp. 189-238.

[5] Ajith Abraham, "*Artificial Neural Networks*", Oklahoma State University, Stillwater, OK, USA

[6] Amir Atiya, "*Learning Nettwork for Neural Networks*", California Institute of Technology, Pasdena, California.

[7] Inma Hernandez, Carlos R.Rivero, David Ruiz, Rafael Corchuelo, "*CALA: An unsupervised URL-based web page classification system*", Scidhanci, Vol. 19, Part 2, April 1994, pp. 189-238.