A STUDY OF MACHINE LEARNING USING GENETIC ALGORITHM

Satish Kumar Gupta[1], Dr. Yashpal Singh[2]

[1]Research Scholar, OPJS University, Churu, Rajasthan
[2]Assistant Professor, OPJS University, Churu, Rajasthan

ABSTRACT—

These genetic algorithms are a kind of optimising algorithms that combine the most efficient survival of the genetic process with a simplified version. Whether machine learning is an issue for the application of GAs has not yet been shown. This is why the study investigates the application of GA in machine education. It conducted a comprehensive research on the effectiveness of GAs in maschine learning[4], but only compared. The article uses Chess as an example for using GA and offers new methods that may replace current approaches for applying GA to machine learning. The suggested approach has shown resilient, making learning rather than an algorithmic natural process. The article is based on the unpredictability of GAs and their capacity to converge the population with a fitness function to the target point and combines it with the feedback notion similar to that of neural networks.

KEYWORDS— Genetic Algorithm, Machine learning, Classifier, Supervised learning.

INTRODUCTION

Heuristic search helps with genetic algorithms. Whether GA's can be used to machine learning is a controversial issue. In the following work, the topic was investigated and illustrated using an example of chess play. In the first part, the definition and kinds of classifier systems were described and machine learning was explained. We saw the exponential rise of data across the globe for many years. Experts say that in the past two years, 90% of global data was produced. This huge quantity of data cannot be handled by humans, thus the computer comes first for treatment and extraction of relevant information. We shall concentrate in this article on a specific kind of data: individuals. Due to the variety of its sources this data may be extremely diverse. Data is obtained from many sources: public (social media, forums...) or private sources (employee database, customer database, etc.). The data gathered are processed in the same manner notwithstanding its diversity: one or more profiles are matched by every user (a genuine individual). A profile may include global (town, gender...) information or particular (job history...) information. The amount of information may be dense or scarce. This article varies from previous studies on the recognition of profiles on social networks (Rawashdeh and Ralescu, 2014), as it does not concentrate on similarity across social networking profiles but between social networks. Even though our research inspired current solutions, such the usage of Vector Space

Model (VSM, Salton, 1968), they are not directly linked. The issue is to identify the same actual individual from various profiles. The purpose is to instruct a computer to answer the question automatically: "Are these two profiles about the same actual person?." Just like for a human being, the instruction is divided into two stages. The computer uses a series of human data to train throughout the first phase. Within this training set, the aforementioned question was answered for every conceivable combination of profiles. Different profiles of diverse sources should be used for the training. The computer can anticipate a resemblance between two profiles after the training stage. The performance is calculated by analysing specified prediction criteria. This research explores how a person's profile may be determined by combining natural language processing, genetic algorithm, and machine learning. We also suggest a novel method for reproduction, called Best Together (BT). The new play is contrasted with previous techniques such as Wheel and Binary. The results show BT as a successful profile recognition technique.

LITERATURE REVIEW

S. Ibrahim's safa (2013) In this article the neural networks are trained via constructive learning. The findings are produced using neural networks however their results are not understandable in form of blackbox. Our aim is to utilise an essential and desired model to discover input variable sets that lead to the intended outcome. This model can contribute to finding an excellent collection of challenging input variables. Accuracy. Genetic algorithms are utilised to understand the achievement of reversal of the neural network. On the other hand, the neural network reversal allows one or more input patterns to be found which meet a particular output. For neural system explanation facility construction, the input patterns generated from the genetic algorithm may be utilised.

Bhasin Harsh (2011) The Genetic Algorithms (GAs) are a kind of optimization algorithms that combine fittest survival with a reduced Genetic Process version. Whether machine learning is an issue for the application of GAs has not yet been shown. The study therefore investigates the use of GAs in machine learning. R. D. King, R. Henery, C. Feng, and A. Sutherland conducted a comprehensive research on the effectiveness of GAs in maschine learning[4], but only compared. The article uses the example of Chess to apply the GA system and offers a new approach that can replace current methods by applying GA to machine learning. The suggested approach has shown resilient, making learning rather than an algorithmic natural process. The article is based on the unpredictability of GAs and their capacity to converge the population with a fitness function to the target point and combines it with the feedback notion similar to that of neural networks.

Tian's Haiman (2016) Researchers in image classification have achieved significant progress mostly because of the availability of huge public visual datasets and of strong CNN models. Pre-trained CNN models may be used to learn extensive features from smaller training datasets to enable information to be transferred from a single source to multiple target areas. Curring frameworks include creating preliminary features from the early layers of pre-trained CNN models, making use of the middle and high-level features and finally tuning the pretrained CNN models to serve various target areas. In this study we suggested to create a deep study model selection framework based on the genetic algorithm to solve different problems in detection. This framework automates the process through which pre-trained models for various functions find the most important and valuable characteristics. Depending on the number of layers each model varies in several ways.

CLASSIFIER SYSTEM

A classifier system is a system which, through credit allocation and rule discovery processes, learns basic rules termed classifier. These systems constantly detect fresh environmental information. They build assumptions without changing the skills gained. This may also be utilised to bring a system of experts into being. Classification systems establish the ranking among the population via numerous environment interactions in which force changes are caused by the division of the classification system's credit subsystem[4]. The classifier system's performance relies on the length of the classifier. The output rises with the length. Thus, with a new situation, the plausibility of developing an action rises. Classifiers match messages and actions from which the environment is modified. Detectors and efficients are available in the systems. Detectors pick specific environmental characteristics and convert the data to a binary form for classifier processing[5]. Effectors convert binary into an environmental change form. The classification system components include

A. Input interface: It uses detectors and converts the binary form into standard messages.

B. Classifiers: Classifiers process messages using system procedures.

C. Message list: Message list has the list of current messages

D. Output interface: It generates the output in the desired form.

GENETIC ALGORITHM

Genetic algorithms (GA) are heuristics, based on Charles Darwin's theory of natural evolution, used to solve optimization problems (Hüe, 1997). The general process for a GA is described as follows (Eberhart et al., 1996), (Kim and Cho, 2000):

Step 1: Initialize a population.

Step 2: Compute the fitness function for each

chromosome in the population.

Step 3: Reproduce chromosomes, based on their fitness.

Step 4 : Perfom crossover and mutation.

Step 5 : Go back to step 2 or stop according to a given stopping criteria.

GA may also be used as an algorithm for machine learning (ML) and has been shown to be efficient (Goldberg, 1989). The notion is that natural algorithms may show a better efficiency than human-designed algorithms in the area of ML in certain situations (Kluwer Academic Publishers, 2001). Actual evolutionary processes, as shown by probabilistic reasoning, have been able to solve quite difficult issues (Moorhead and Kaplan, 1967). For each label included in a training set, GA will be utilised to establish an appropriate weighting set. Our training set consists of profile-similarities, two profiles are either similar (output = 1) or unlike (output = 0).

1. Genetic Representation

The genotype is the group of all labels in the training set for each chromosome in the population. Each label is specified as a gene and the allel of the associated gene is weighed by a certain label. The weighting is in [0,1], the significance of the label may be expressed, and its greatest value at 1 is at 0, and the worst at 0.

2. Population Initialization

Two questions arise for population initialization: 'What's the starting population size?' and 'What is the initial population initialization procedure?' The aim is to compromise the global complexity and performance of the solution on population size (Holland, 1992). A small population cannot converge while a big population may need excessive memory and computer time. In order to solve our particular issue, it turns out that the size of the first population must be carefully selected. As a reminder, the GA has to calculate weighting in the training set for some N of the labels. Different population size values were examined, compromised between performance and complexity and the population had to be set at 2 NN. Secondly, there are generally two different methods to initialise a population: heuristic initialization or random initialization. Even though it enables a quicker convergence of solutions, heuristic initialisation is an obstacle to the GA's search solution in a given region and it may not converge to an optimal global (Hüe, 1997). A random initialization helps prevent GA from being trapped in a local optimum. In our instance, the random initialization consists of a random weighting of [0,1] in the genotype of each gene.

3. Fitness Function

The primary aim of a healthy function, depending on how well it matches the intended solution, is to assess a chromosome (Hüe 1997). In our instance, chromosomes are assessed by the use of the VSM technique to forecast similarities and how well they match the training set. We have used an extensive ML fitness function (Shen, 2005), since our problems may be linked to the issue of regression.

4. Crossover

The artificial technique of reproduction for the GA is crossover. It needs to determine how two chromosomes create a descendant. Crossover is essential for GA, leading to an organised flow of genetic information across solutions that may lead to excellent solutions (Srinivas and Patnaik, 1994).

5. Mutation

Mutation causes some chromosomes to alter and enables the GA to escape the local optima (Hüe, 1997). In our situation, there is the danger of being trapped in a local optima. The goal of the GA is to discover the optimum weighting for a number of labels that comprises of selecting the relevant labels (high weighting) and the non-relevant ones (low weighting). Lokal optima is a variant of the GA that detects the relevance to a certain similarity inside the training set, for instance, using the "lastname" label. This outcome leads to improved fitness. However, owing to the need for the appropriate weighting and ordering for each label, it would be a local optimum. A suitable variety should be specified to maintain a decent weighting for a particular label, but permit variations on other labels. We thus decided to introduce a technique of mutation that improves variety. We have set the mutation rate Mr in addition to the mutation chance Mc, which determines the probability of a child mutating. Should a new chromosome mutate (this means, if the number of the chromosome is random in [0,1] < Mc), each one is subject to different random factor F [1 – Mr, 1 + M]. A repair function is naturally performed to guarantee every allle is kept in [0,1].

6. Reproduction

The technique of reproduction (or selection) aims at improving a new generation's average quality. This technique typically uses the Darwin principle to achieve this objective, namely that the fittest chromosomes tend to replicate more (Hüe, 1997). Two kinds of reproductive techniques are available: proportional and ordinal. The first is based on the worth of fitness and provides a chance to replicate each chromosome proportionately to its value. The wheel is a standard way (Goldberg, 1989). We consider a wheel with a section proportionate to the size of each chromosome. We just "launch" the wheel to pick crossover partners. Every chromosome has the opportunity to become a parent using this technique; but, statistically, the most suited chromosomes may be much better than others. Conversely, the chromosomes are classified according to their fitness for ordinal-based techniques. The selection depends on the population's chromosomal rank. The usage of the tournament technique is a traditional approach (Miller and Goldberg, 1995). This method must have a S (tournament size) argument, which defines the round size. This technique really goes on round after round. S chromosomes are randomly chosen in each round, the highest ranking chromosome is selected around and cross-crossed. We nonetheless came up with a novel approach for that particular issue of profile identification. Then, compared to the state of the art reproductive technique, we will demonstrate the outcomes.

In general, a GA has a minimal probability of mutation (~0,05), resulting in controlled variety. The reproductive technique also monitors variety by allowing a chance to cross across each chromosome worldwide. Our GA is designed to establish a large mutation factor (section 3.5) to guarantee our algorithm is very diverse and to avoid it from being embedded in a local optima. In addition to this, tactics are designed to provide a highly selective way of reproducing all inadequate chromosomes. In summary there will not be enough chromosomes yet a generation will have a large amount of variety. Our technique of mutation ensures that variety, so we just have to choose the reproduction strategy. We have thus developed our own, Best-Together reproduction technique (BT). The notion of considering only an elite set of reproductive chromosomes brought by the advent of partial random key genetic algorithms is an inspiration (Resende, 2010). It is an ordinal technique; the population must be classified accordingly. This technique selects a number X from the best chromosomes for a population of size N.

MACHINE LEARNING

Machine learning is designed to create algorithms that computers use empirical data to generate behaviour. The computer is taught in the machine. A student may use examples to grasp the important features of its unknown fundamental distribution of probability. The examples show the relationship between input and intended outcome. An important job of master study is automated learning to detect complex patterns and take smart choices based on instances. Any behaviours given all potential inputs are too wide to cover with the number of instances seen. Therefore the learner must look at the examples so that an usable output in new situations may be produced[1]. Tom M. Mitchell claimed 'If the performance on T tasks, as measured by P, increases with the experience E, a computer programme will learn from experience E for a task group T and performance measure P'.' [2] A learner's primary goal is to spread his experience[3]. The teaching examples from his experience typically originate from a pattern that is indeterminate and the student must extract something more general out of it which will enable him to provide helpful responses in new situations.

A. Strategies

There are many techniques and strategies used in machine learning. Some of them are

• Inductive Logic Programming

• Simulated Annealing

• Evolutionary Strategies

• Neural Nets

B. Types of Machine Learning Algorithms

Machine learning algorithms can be classified based on desired outcome of the algorithm. It is of two types: supervised and unsupervised.

• Supervised learning generates a function that maps inputs to desired outputs. For example, in MS Word speech to writing translation is possible. A paragraph is taken and training is provided to the machine i.e. computer in order to learn the pronunciation and ellipses of the speaker.

• Unsupervised learning models a set of inputs, like clustering.

Some more algorithms are based on the above two as described in the following section.

• Semi-supervised learning combines both supervised and unsupervised illustrations to produce a classifier.

• Reinforcement learning learns how to act to a given situation in a particular scenario. Every action has some impact in the environment, and the environment provides feedback in the form of rewards and punishments.

• Transduction tries to predict new outputs based on training inputs

CONCLUSION

The idea is implemented, however if done fully, this will have a better outcome since it would significantly decrease the amount of time needed to look for movements. It also provides a method to include GA into machine education. Our approach enabled the simultaneous identification of both profiles and which profiles match the same individual automatically. This article offers an adaptive and general solution. This adjustment is due to the lack of a fixed set of labels in the model. As long as the labels remain, the mathematical model as well as the genetic process will fit into a new set of labels. We think, therefore, that it may be utilised for sources that include profiles of any type. In addition, this approach may be utilised for other comparable data recognition applications.

REFERENCES

1. Reeves C.R., Bush D.R. (2001) Using Genetic Algorithms for Training Data Selection in RBF Networks. In: Liu H., Motoda H. (eds) Instance Selection and Construction for Data Mining. The Springer International Series in Engineering and Computer Science, vol 608.
2. Bhasin, Harsh & Bhatia, Surbhi. (2012). Application of Genetic Algorithms in Machine learning.
3. Safa S. Ibrahim, (2013), INTERPRETATION TRAINED NEURAL NETWORKS BASED ON GENETIC ALGORITHMS, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.4, No.1

4. Hafezi, R.; Shahrabi, J.; Hadavandi, E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. Appl. Soft Comput. 2015, 29, 196–210.
5. Chiang, W.-C.; Enke, D.; Wu, T.; Wang, R. An adaptive stock index trading decision support system. J. Econ. Policy Res. 2016, 59, 195–207
6. Classifier System and Genetic Algorithm , L.B. Booker, D.E. Goldberg and J.H. Holland ,Computer Science and Engineering, 3116 EECS Building, The University of Michigan, Ann Arbor, MI 48109, U.S.
7. Mitchell, T. (1997). Machine Learning, McGraw Hill. ISBN 0-07-042807-7, p.2.
8. Classifier System and Genetic Algorithm, Robert A. Richards
9. R. D. King, R. Henery, C. Feng, and A. Sutherland. A Comparative Study of Classification Algorithms: Statistical, Machine Learning, and Neural Network. In K. Furukawa, D. Michie, and S. Muggleton, editors, Machine Intelligence, volume 13, pages 311–359. Clarendon Press, 1994.
10. G. K´okai. GeLog - A System Combining Genetic Algorithm with Inductive Logic Programming. In Proceedings of the International Conference on Computational Intelligence, 7th Fuzzy Days, pages 326–345. Springer-Verlag, 2001.