Optimization Technique to Development E-market Security Via Social Media

Sam M. K.

The higher institute of Telecommunication & Engineering,

Abstract

One of the best reasons for your business to be marketing through social media is that your customers are spending time on these channels. We proposed a framework which depends on Grey Wolf Optimization Technique also three machine learning algorithms: Random Forest, Linear Regression, and Artificial Neural Networks are used. Data preprocessing is done so that data is reduced and normalized before training the models. The results of this model are compared with results from different machine learning models. The proposed model produces very accurate results Framework estimation and its powerful and stability.

Keywords: Optimization, Information Security, H2o, Facebook, Trust, Grey Wolf Optimizer

1. Introduction

Although the business deal is impersonal, the activities taking place online are the result of human behaviors that can be used as trust measurements [1]. One study considered only the seller’s perspective on trust for online selling. The researchers considered only social media online market while recognizing that sellers, the intermediary, and the buyers all have a unique point of view when deciding if they trust a site enough to use the site again.

Optimization Technique

Nowadays, Social networks, Search engines, E-Commerce, Blogs, Sensors, and biomedical records are generating large volumes of data (zeta bytes) daily [2]. This huge amount of data considered as Big Data is different from traditional data in terms of volume, velocity, and variety. Volume specifies the amount of data (data size), velocity specifies the data generation rate, and variety stands for data structure in general. To deal with this data, we need optimization techniques. One of the most important is Grey Wolf, and also, we can use H2O to have Distributed File System [3].

H2O

H2O is open-source software for big-data analysis. It is a Java-based software for data modeling and general computing. The primary purpose of H2O is providing a distributed (many machines), parallel (many CPUs), in memory (server hundred GBs Xmx) processing engine. The main goal of H2O is allowing simple horizontal scaling to a given problem for producing a faster solution.
Purchasing Process

One risk during the purchase process is that money will be paid but the item or items will never be delivered. Another risk is "escrow service fraud" when a seller convinces the buyer to go to a third-party escrow service fraud to carry out the payment [4] The supposed seller or buyer can prepare a fake escrow service site without the other party knowing the difference. Payment is sent to the escrow service. But no merchandise is received. Or in opposite instance the seller sends the merchandise but never receives the payment Eastern Europe example do not have to pay all at once and can pay with wire transfers with no money transfer control number needed but one party has the others full name and full address [5].

Feedback rating

Feedback is a method that rates previous users' experiences with comments and with ratings such as from satisfactory to unsatisfactory [6]. Even if there are only a small number of comments or ratings that are not honest, the result can lead to misleading feedback that other users rely upon

High Trust

Higher levels of trust are demonstrated by participating in social media online market often or with regularity. Research has identified a positive correlation between trust and frequent participation on social media online market. Four dimensions for online failures and satisfaction (a) customer service, (b) security/privacy, (c) fulfillment/reliability and (d) website design and interaction were identified as having a relationship with trust. [7]

Earlier research determined that some feelings of trust or a trustful point of view must be present in order to Internet users to use social media online market even though there is no personal contact with the seller. Finding the information, they want, when they want it, on a website has been show to positively impact trust. On the other hand, websites, displaying warranties do not have very much influence on trust but instead it is third party endorsements that are more effective. [8]

2. Review and related work

Several researchers have studied the effect of the environment surrounding an individual. Davies and Powell wrote, People are neither deterministically controlled by their environments nor entirely self' determining. Instead they exist in a state of reciprocal determinism whereby they and their environments influence one another in a perpetual dynamic interplay." [9] The model assumes that the environment is the social media online market. Explains psychosocial functioning in terms of a triadic reciprocal causation whereby an individual's internal psychological factors, the environment which they are in and the behavior they exhibit, all operate as interacting determinants that influence one another bi-directionally. [10]
3. The proposed Framework

The theory is based on the fact that interactions on the social media online market sites are governed by human behavior, therefore research' needs to be carried out on the cognitive behavior of the individuals involved in the interactions. The behavior of bidders on business-to-consumer (B2C) social media online markets are the topic investigated in the research. B2C is applied to social media online markets that are also businesses or part of a bigger business that sell to consumers. In this case the B2C takes place online but the definition of B2C is applied to real life situations, mail order and any other purchase interaction when a business sells to a consumer.

The proposed Framework will be as follow: as shown in figure

A- Data Collection (Feature Selection) using Grey wolf optimizer
B- Data Processing (Data scaling) using Grey wolf optimizer
C- Estimation Module Using H2O

![Figure 1. The proposed Framework](image-url)

Data Collection (Feature Selection) using Grey wolf optimizer:
Grey wolves are one of apex predators, apex predators are at the top of the food chain. Grey wolves usually prefer to live in groups. The group size varies from 5–12 on average. The most interesting thing about grey wolves is that they have a very strict social leading hierarchy. The leaders called alphas, are a male and a female. The alpha is responsible for making important decisions to the group such as hunting, sleeping place, and so on. The whole group should obey the alpha’s orders. On the other hand, it had been observed that sometimes the alpha follows the other wolves in the group, which can be considered as some kind of democratic behavior. The alpha wolf is also called the leader wolf since the whole group should follow his/her orders. Interestingly, it is not necessarily for the alpha to be the strongest member among the group, but the best in terms of leading the group. Which implies that the discipline and organization of the group is more important than its strength.

\[ D = |\vec{C} \cdot \vec{F}_p(t) - \vec{F}(t)| \]  

\[ \vec{F}(t+1) = \vec{F}_p(t) - \vec{A} \cdot \vec{D} \]  

Where \( t \) represents the current iteration, \( \vec{A} \) and \( \vec{C} \) are coefficient vectors, which are calculated as in Eq. (14) and Eq. (15), \( \vec{F}_p \) indicates the position vector of the prey, and \( \vec{F} \) is the grey wolf’s position vector. In each iteration, \( \vec{F} \) will be updated if there is a better solution.

\[ \vec{a} = 2 - t \left( \frac{2}{Max_{iter}} \right) \]  

\[ \vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \]  

\[ \vec{C} = 2 \cdot \vec{r}_2 \]  

Where \( t \) is the loop counter, \( Max_{iter} \) is the maximum number of iterations in this loop, \( \vec{r}_1 \), \( \vec{r}_2 \) are random vectors in \([0,1]\), and \( \vec{a} \) is linearly decreased from 2 to 0 over the course of iterations. For testing the effects of Eqs. (1) and (2), assume a two-dimensional position vector and some of the possible neighbors as shown in Fig. 2.

**Figure 2. Grey wolf optimizer**

**Data Processing (Data scaling) using Grey wolf optimizer**

After selecting the appropriate features, we standardized the data to change the raw feature vectors into a representation that is more suitable for the different machine learning estimators. For this task, we used Standard Scaler which standardize features by removing the mean and scaling to unit variance.
Table [1] Sample of Dataset records

<table>
<thead>
<tr>
<th>InvoiceNo</th>
<th>StockCode</th>
<th>Description</th>
<th>Quantity</th>
<th>InvoiceDate</th>
<th>UnitPrice</th>
<th>CustomerID</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>622</td>
<td>536414</td>
<td>22139 NaN</td>
<td>56</td>
<td>12/1/2010 11:52</td>
<td>0.00</td>
<td>NaN</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>1443</td>
<td>536544</td>
<td>21773 DECORATIVE ROSE BATHROOM BOTTLE</td>
<td>1</td>
<td>12/1/2010 14:32</td>
<td>2.51</td>
<td>NaN</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>1444</td>
<td>536544</td>
<td>21774 DECORATIVE CATS BATHROOM BOTTLE</td>
<td>2</td>
<td>12/1/2010 14:32</td>
<td>2.51</td>
<td>NaN</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>1445</td>
<td>536544</td>
<td>21756 POLKADOT RAIN HAT</td>
<td>4</td>
<td>12/1/2010 14:32</td>
<td>0.85</td>
<td>NaN</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>1446</td>
<td>536544</td>
<td>21757 RAIN PONCHO RETROSPOET</td>
<td>2</td>
<td>12/1/2010 14:32</td>
<td>1.66</td>
<td>NaN</td>
<td>United Kingdom</td>
</tr>
</tbody>
</table>

the retailer records such kind of entries without a further description? It seems that there is no sophisticated procedure how to deal with and record such kind of transactions. This is already a hint that we could expect strange entries in our data and that it can be difficult to detect them.

**Estimation Module Using H2O**

It can be called from the statistical package R, Python and other environments. It is used for analyzing and exploring cloud computing systems datasets and in Apache Hadoop Distributed File System as well as operating systems like Windows, Linux and macOS. H2O is providing implementations for many of the most popular algorithms such as Random Forest, Deep Neural Networks, GBM, stacked Ensembles, and Word2Vec. H2O is extensible so that developers can add data transformations and custom algorithms of their choice.

H2O’s core code is written in Java. Inside H2O, a Distributed key/value store is used to access and reference data, models, objects, etc. across all nodes and machines. The implementation of the algorithms will be carried out on top of H2O’s distributed Map/Reduce framework and utilizing the Java Fork/Join framework for multi-threading. The data is read in parallel, then, it will be distributed across the cluster and stored in memory in a compressed way in a columnar format.
H2O’s data parser has built-in AI that predicts the schema of the incoming dataset and supports data ingest in various formats from multiple sources. H2O’s REST API allows access to all H2O capabilities from an external program or script via JSON over HTTP. The rest API is used by H2O’s web interface (Flow UI), Python binding (H2O-Python), and R binding (H2O-R).

4. Results and Discussion

Three artificial intelligence algorithms were proposed in our experiment:

- **Artificial Neural Network (ANN)**
  
  An information processing paradigm inspired by biological neural networks. Nowadays, ANN is widely used in many applications in various aspects of engineering and science.

- **Linear Regression**
  
  Multiple Linear Regression is based on Ordinary Least Squares (OLS) which tries to fit the model by minimizing the sum-of-square of the difference between the actual (observed) value and the estimated (predicted) value.

- **Random Forest**
  
  Random Forest Algorithm combines multiple decision trees to improve the generalizability and the robustness over a single decision tree.

Finally, The Random Forest was able to predict the value with the lowest error in comparison with other estimators (RMSE = 0.0123). The next figures illustrate the estimation Algorithm.
<table>
<thead>
<tr>
<th>Estimator</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>0.0435</td>
<td>0.0123</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>0.0750</td>
<td>0.0137</td>
</tr>
<tr>
<td>Artificial Neural Networks</td>
<td>0.2721</td>
<td>0.0954</td>
</tr>
</tbody>
</table>
5. Conclusions

In this research, we proposed framework machine learning tasks on H2O to Development Optimization Technique for social media online market deepen on grey wolf optimizer. We used the following three machine learning algorithms: Random Forest, Linear Regression and Artificial Neural Networks. Random forest produced the lowest error.

Reference


