View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by MAT Journals

MAT JOURNALS
Journal of Data Mining and Management
Volume 2 Issue 3

Framework for Predictive Analytics with Complex Event Processing

Ajay Acharya
Assistant Professor
Department of Computer Science
KLS Gogte Institute of Institute of Technology
Belagavi, Karnataka, India
aacharya@git.edu

Dr Nandini S. Sidnal
Professor
Department of Computer Science
KLE MSSCET,
Belagavi, Karnataka, India
sidnal.nandini@gmail.com

Abstract

Event processing is a technique of tracking and scrutinizing of information about the events that happen and deriving a conclusion from them. Complex Event Processing(CEP) is a technology which allows combining of data from multiple sources and infers complicated patterns from the events. Complex events require real time detection in order to have time for appropriate reactions. However, there are several events which should be prevented rather than responding to them after they have occurred. This can be achieved using Predictive analysis. Predictive analysis(PA) enhances the performance of CEP. In this paper, we define CEP and PA technology and provide a conceptual framework which provides synergy between CEP and PA. This framework can be the basis of general design pattern in future.

Keywords: *Complex Event Processing, Predictive Analytics*

INTRODUCTION

The paper presents the synergy between two promising technologies viz Complex Event Processing and Predictive Analysis. The mentioned technologies can be integrated to produce results which are efficient than either of the technology. Complex Event processing is the ability to import data from multiple sources and apply complex rules and derive outbound actions. Complex Event Processing can be employed in diverse application ranging from financial analysis, healthcare, real time business intelligence, fraud detection etc. CEP processes stream of data to produce complex events in real-time.

Predictive analytics is a methodology which uses certain data mining, machine learning techniques to predict the future events. It deals with analyzing historical or current data to predict or foresee future. examples of PA are weather fore casting where data from various sources has to be analyzed and appropriate predictive model has to be applied which helps in anticipating the weather. Credit card fraud detection is another example wherein the fraudulent activities are to be detected in real time.

CEP does not predict the complex events but requires the events to occur in order to analyze complex events and PA can predict events based on historic data. By combining these two methodologies we aim at enhancing the analysis and prediction process.

BACKGROUND

Basic definitions: An event is anything that happens, or is contemplated as happening. An event can be visualized as object that represents encodes or records an event, generally for the purpose of computer processing. An event can be a virtual or a real-world scenario. An event can be a simple event or a complex event. A simple event is the one which is not a composition or abstraction of other events. A simple event cannot be further broken down into simpler events. A complex

Page 1-6 © MAT Journals 2017. All Rights Reserved

event on other hand is an abstraction of simple events viz a complex event is composed of several simple events. An event source provides events as the input to the event processing system. Examples of event sources are simple RFID sensors and actuators, business flows, and CICS applications. An event cloud is a partially ordered set of events (poset), either bounded or unbounded, where the partial orderings are imposed by the causal, timing, and other relationships between the events. An Event processing network is a set of event processing agents (EPAs) and the channels they use to communicate. Event processing agents is a software module that processes events.

Complex Event Processing

Complex Event Processing or CEP, is an event processing technique that takes in data from various sources as input, analyses and processes it to produce more complicated circumstances as output. The aim of Complex Event Processing is to identify meaningful and relevant events (such as opportunities or threats) and respond to them as fast as possible. CEP has a potential to define, manage and predict events or situations that can be an opportunity or a threat in complex networks. CEP deals with detecting real time complex events.

CEP executes data on stored queries unlike the traditional database systems where query is executed on stored data. CEP queries are applied on potentially infinite stream of data. The inputs are processed instantaneously and the data which is not relevant is discarded. Once the system matches the input events to predefined sequence or pattern the output is generated. This aspect leads to the CEP's real- time analytics capability. CEP's real-time analytics capability makes it suitable to be used in high demand continuous intelligence applications to enhance situation awareness and real- time decision making.

CEP finds its application in algorithmic trading, security and fraud detection, health care etc.

In algorithmic trading [1] space, an algorithm means sequence of steps by which pattern in real-time market data can be recognized and responded in order to detect trading opportunities and place and manage orders in market. Before algorithmic trading the traders manually carried out the process of building and managing a trading strategy. The traders in trading stations had to watch out four to eight screens to watch real-time market. The traders had to manually analyses patterns possibly in a spreadsheet and work out where possibly to place orders (buying and selling orders). With the advent of algorithmic trading the traders can now initiate and manage hundreds of algorithms without actually being involved. In this way, the trader can be more productive.

Algorithmic trading can be advantageous if the time between algorithm conception and implementation can be reduced. The traditional approach of developing trading strategies in C++ or java take a long period of time to implement. However, using CEP can reduce this time. CEP processes stream of continuously changing data and produces results in almost real time. This significantly increases the efficiency of the strategy.

CEP also finds its application in the IOT space. With the advancement in internet of things more and more data is collected from the sensors and smart devices. This huge amount of data collected should be analyzed and response should be generated as quickly as possible. Since CEP can analyses and predict data in real time,

Page 1-6 © MAT Journals 2017. All Rights Reserved

integrating CEP with IOT could result in efficient systems.

Predictive Analytics

Predictive analytics extracts useful information from data (past or present data) and analyzes the information to predict the future events. The patterns found in historical data can be used to identify and anticipate risks and opportunities. The task of the predictive model is to estimate that a similar unit in a different sample will exhibit the specific performance. Predictive analytics is used in several fields: financial services, insurance, telecommunication, retail, travel, healthcare, pharmaceutical industry, etc. The techniques used in PA are regression techniques which include linear regression, logistic regression etc. Machine learning techniques enable the computers to learn from data that is fed into the system. Machine learning techniques are widely used in PA because of its self-learning capability. A predictive process will follow the four steps

First the raw data is collected from source and is pre-processed

This pre-processed data is transformed in an appropriate structure so that it can be given as an input to the machine learning technique.

Then, the learning model or the training set is constructed based on the transformed data.

The predicted output is reported using the previously created training set.

Predictive modelling describes a methodology where the underlying relationships between the historical data are recognized and analyzed in order to make the data and the predictions. The predictive model's analysis, test and validate the data in order to make the best possible prediction of the probability of an outcome A model is reusable and is created by a training algorithm by analyzing the historical data and saving the model. The fact that the model is saved is because there is always a fair probability that the model can be applied to a similar data in the future. The predictive models use certain algorithms to achieve this task. Some of the algorithms are

1. Time series algorithms – here predictions are performed time based.
2. Regression algorithms-which predicts continuous variables.
3. Clustering algorithms which cluster observations into groups of similar groups.

ML

Machine learning algorithms are classified into three categories:

1. Supervised Machine learning algorithms- these algorithms make predictions based on some given set of samples i.e. prediction of patterns where value labels are assigned to the data points.
2. Unsupervised Machine learning algorithms- here there are no labels assigned to the data points. These algorithms make the complex data simpler by organizing data into groups.
3. Reinforcement algorithms- these algorithms operate i.e. take some actions on the data points and later learn whether the decision was right.

Some machine learning algorithms are

1. Naïve Bayes Classifier Algorithm- This algorithm is used when the data or the values can be classified or categorized based on some attributes. For example, the Naïve Bayes Classifier Algorithm is used in spam filtering where in emails are filtered and if a spam is found, it is put into spam category
2. Linear regression algorithm – this algorithm describes the relationship between the two variables and also describe how the change in one

Page 1-6 © MAT Journals 2017. All Rights Reserved

variable may affect the other. The variables may be dependent or independent. This algorithm shows the impact of the dependent variable on the changing independent variable. This is algorithm is termed as the most interpretable algorithm as it is easy to explain to others. Its performance output is considerably good as it runs fast making it the most widely used algorithm.

3. K Means Clustering Algorithm- It is an unsupervised machine algorithm. It uses an iterative method and operates on a given data set through predefined number of clusters. For example, consider a Wikipedia search for the word jaguar. The output will contain information relating to the word like jaguar as a car, jaguar as an animal etc. So, by using K means clustering algorithm, the output can be more structured i.e. the information relating to jaguar as a car is grouped into one cluster and so on.

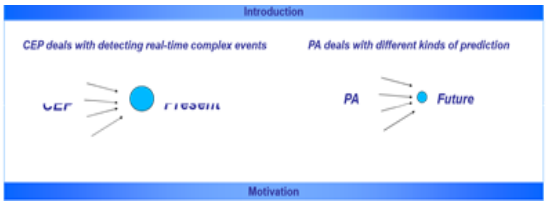

Fig 1: *CEP vs PA*

CONCEPTUAL FRAMEWORK

As the name describes, CEP is concerned with processing of complex events i.e. it detects the complex events in the real time. Predictive analytics is concerned with predicting the future events i.e. long term or short-term events using predictive models which take historical data as input. So, the combination of both the approaches can give rise to many useful applications. Both the approaches can be combined so as to predict complex events in real time or even before the event happens. CEP is built up on a set of queries which are fed with the real-time data for detection of the complex events. These queries are developed using simple events which are predefined. By the connection of CEP and PA, complex events can be predicted.

As the name describes, CEP is concerned with processing of complex events i.e. it detects the complex events in the real time. Predictive analytics is concerned with predicting the future events i.e. long term or short-term events using predictive models which take historical data as input. So, the combination of both the approaches can give rise to many useful applications. Both the approaches can be combined so as to predict complex events in real time or even before the event happens. CEP is built up on a set of queries which are fed with the real-time data for detection of the complex events. These queries are developed using simple events which are predefined. By the connection of CEP and PA, complex events can be predicted.

The requirements to be taken care of while building such a model are

The future occurrences are to be predicted correctly by the PA and sent to the CEP.
The predictors and the Primary complex events are the key aspects required by the PA. So, CEP should ensure this.

There should be some level of transparency maintained between the key components of the system i.e. the modification of components of CEP should not trigger the modification of PA components and vice versa.

Synergy should not affect the maintainability of the CEP system

Page 1-6 © MAT Journals 2017. All Rights Reserved

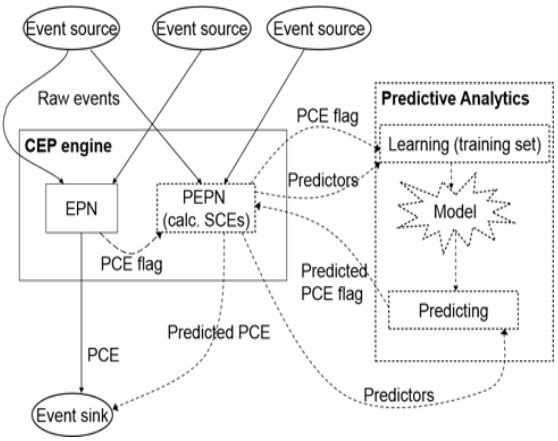

Fig 2: *Framework for PA with CEP*

The Complex Event Processing with Predictive Analytics has two parts: CEP engine and Predictive Analytics.

The framework is depicted in fig 2. The CEP engine contains Event Processing Network and the Predictive Event Processing Network. The Event Processing Network has four fundamental parts i.e., Event Producers, Event Consumers, Event Processing Agents and Event Channel. Event Producers produce events which can be taken as inputs by Event Consumers or an Event Processing Agent. The Event Producers are represented as source nodes and are only capable of producing events. The Event Consumers takes the events as input and performs specific tasks associated with the events. The Event Consumers are represented as sink. Event Channel is a path through which events from the Producers and/or EPAs are transferred to the Consumers and/or EPAs. The Event Channel is accountable to appropriately take in events from the Producers and/or EPAs, order and combine them and deliver it to the appropriate Consumers and/or EPAs. The Event Processing Agents check if the events are in legitimate format. It includes pattern matching, processing and emission. Raw events come from the event sources. The Event Processing Network processes these raw events and detects Primary Complex Events. The detected PCEs are sent to the event sink and the corresponding action is taken. At the same instance EPN sends the PCE flag (PCEF) to the Predictive Event Processing Network (PEPN) to indicate if the PCE has occurred or not. Then the Primary event processing network enhances the training set of the PA using the primary complex event flag. The primary complex event flag is a Boolean variable and it tells whether or not the primary complex event has occurred. Then the predictors are sent by the primary event processing network to the PA and a new prediction(PPCEF) about the primary complex event flag is given by the PA on the basis of the current learning model and the predictors. PPCEF denotes the predicted primary complex event flag and represents the forecasted primary complex event. The predicted events are sent to the event sink. A predictive model used in the PA is of machine learning based. This model works continuously on the extended training set. The PEPN produces the Secondary Complex Events which are the input to the PA. The model needs to be refreshed regularly. Because the model refreshment is a resource consuming task therefore it should be performed seldom and on a separate thread during the implementation.

TOOLS FOR CEP AND PA

Some of the Predictive Analytics tools listed as follows [5]

1. Rapid Miner- Rapid Miner provides a platform where in machine learning, data mining, predictive analytics and statistical modeling are integrated. it uses a client/server model.
2. SAS predictive analytics - SAS is a leading software in the field of business analytics. It provides high performance data mining, statistical analysis, forecasting, visual analytics etc.
3. IBM predictive analytics- IBM provides an easy to use predictive analytics software that meet specific

Page 1-6 © MAT Journals 2017. All Rights Reserved

needs of users. It helps transform data into predictive insights to predict customer needs, how to maximize the profit and productivity, perform statistical analysis using regression techniques, cluster analysis and correlation analysis.

4. KNIME- Konstanz Information Miner is an open source predictive analytics tool and it integrates machine learning and data mining using modular data pipelining concept and provides a better GUI.
5. Oracle Data Mining(ODM)- It provides a powerful platform for data mining and includes many data analysis algorithms which provides regression, classification, prediction. These models stored as objects in the Oracle database and the models are implemented in Oracle database kernel.

Tools for CEP are as follows

1. Microsoft Stream Insight- It has a development platform based on .NET framework which enables rapid implementation of CEP application.
2. Stream Base- It is a platform which allows low cost real time applications on streaming data. Other prominent CEP tools are : Oracle Complex Event Processing, ESPER (open source) and TIBCO Business Event.

CONCLUSION

The paper provides an insight of two important technologies viz CEP and PA along with their respective tools. The paper provides the synergy between the two technologies and provides a conceptual framework which can be a generic model for future development. The paper also highlights the areas which can use these technologies for productive and efficient results. The paper provides a layout of the two technologies and open up's the possibilities of its application in various fields.

REFERENCES

1. http://collaboration.cmc.ec.gc.ca/science/rpn/biblio/ddj/Website/articles/DDJ/2007/0704/070301jb01/070301jb01.html Algorithmic trading. The use of algorithms in automate trading by John Bates.
2. Complex Event Processing Synergies with Predictive Analytics by Árpád Beszédes, Árpád Beszédes, Lóránt Farkas.
3. Predictive Complex Event Processing: A Conceptual Framework for Combining Complex Event Processing and Predictive Analytics by Lajos Jeno Fülöp, Gabriella Tóth, László Vidács.
4. https://docs.oracle.com/cd/E13157_01/wlevs/docs30/get_started/overview.html-Overview of oracle complex event processing.
5. https://www.predictiveanalyticstoday.com/top-predictive-analytics-software/ - Predictive Analytics software.

Page 1-6 © MAT Journals 2017. All Rights Reserved