Skilled Agent Application as Software Detectives

Anupam

College Of Advance Studies, Azamgarh.
anupam.sri.2008@gmail.com

Abstract— In today's environment each persons and computer are connected with network, a new entity has evolved: skilled agent software. Over the past few decade agents have emerged as a new software paradigm; they are in part distributed systems, autonomous programs and artificial life. The concept of agents is an outgrowth of years of research in the fields of AI and robotics. They represent concepts of reasoning, knowledge representation, and autonomous learning. Agents are automated programs and provide tools for integration across multiple applications and database running across open and closed networks. They are a means of retrieving, filtering, managing, monitoring, analysing and disseminating information over the Internet, intranets and other proprietary networks.

Agents represent a new generation of computing systems and are one of the more recent developments in the field of AI. Agents are specific applications with predefined goals, which can run autonomously; for example, an Internet-based agent can retrieve documents based on user – defined criteria. They can also monitor an environment and issue alerts or go into action based on how they are programmed. In the course of investigative data mining projects, for example agents can serve the function of software detectives, monitoring, shadowing, recognizing and retrieving information for analysis and case development or real – time alerts.

Agents can be used by investigators and analysts to work on their behalf; for example FinCEN, the U.S. Treasury agency set up to detect money laundering, must review all cash transactions involving dollar amounts of above $ 10000. This amount is roughly 10 million transactions a year, which cannot be manually monitored. The FinCEN Artificial Intelligence System users, an agent to weed through this large data space and search for abnormalities and fraud through the use of neural network and link analysis.

Keywords— *Open Source Agents, Data Mining Agents*

1. Agent or Representative

An intelligent agent is software that assists users and acts on their behalf. Agent autonomously perform tasks delegate by their creators and users. Agents can automate repetitive tasks, remember events, summarize complex data, learn and make recommendations. For example an agent can be used to monitor and search for a suspect's name from multiple government and commercial databases or it can be set up to assemble evidence for use in a prosecution case.

Intelligent agents continuously perform three main functions, which differentiates them from other software programs:

1. They are capable of perceiving dynamic conditions in an environment.
2. They can take action to affect conditions in an environment.
3. They can reason to interpret finding, solve problems, draw inference and determine future actions.

For example, agent software can act on behalf of investigators and thus reduce their workload by sifting through large amounts of data for evidence gathering. Agents have the capability to interact with the external environment and perceive changes in it; hence they can then either inform investigators of changes, such as that a suspect on the INS list has entered the country. Or, they can be set up to react dynamically to finding, issuing an alert at the point-of-entry station, once a match of a suspect on the INS list is found. While there are multiple definitions of intelligent agents, this is their essential characteristic: a software agent is a computing entity that performs user-delegated tasks autonomously. An agent can perform many tasks; however, for investigative data mining the most dominant ones are likely to be information monitoring, retrieval, organization and reporting.

Agent technology is not a single, new technology but rather the integrated application of a number of network, Internet and AI technologies. As such, developers normally do not set out to construct an agent; more commonly they set out to add new functionality to a new or existing application that posses agent-like features. These agent programs possess various forms of learning, creating and modifying rules of behaviour and developing strategies for collaborating among other programs, databases, networks and users and even other agents. Agents can be integrated with other applications, enabling investigators and analysts to automate many tasks. We will propose a system using agent technology for the integration of human investigators and machine-learning algorithms to create a new type of evolutionary investigative system resulting in a fusion of human and machine intelligence.

2. Agent Features or Appearance

In order to define the characteristics of an agent further and to distinguish them from any other type of program the following list en-numerates the attributes and features required of them:

2.1 Autonomy

Being able to carry out tasks independently is the most important feature of an agent; this differentiates an agent from any other computing technique or program. Traditional computer applications only respond to direct manipulation via user instructions. On the other hand intelligent agents can perform actions without human interventions. An agent principally operates without direct intervention, typically in the background, to the extent of the user's specified delegation. The autonomous ability of an agent can range from being able to initiate a lookup in a database to issuing alerts or collecting and assembling a file from multiple networked sources. The search agent will take the input and perform the search independently without user intervention. With the widespread use of the Internet, intranets and other electronic and wireless networks, stationary and mobile agents can be used for investigative data mining applications for detection and deterrence. Stationary agents can send scripts and receive data via networks, but cannot they move. While all agent are not mobile, there have been significant trends toward developing nimble and mobile agents. Mobile agents have the capacity to navigate through networked architectures in the performance of their tasks and to report their finding to various wireless devices or appliances.

2.2 Perception

The agent needs to be able to affect its environment via some type of programmed mechanism for autonomous operation. It needs to be able to monitor its environment in order to be able to perform its task independently. An agent must be able to perceive events in the environment and react to them as necessary in an appropriate fashion. Agents almost never operate in isolation. They work within a system or in a network. Its environment includes other agents, systems, human users and in certain cases external objects such as sensory devices on factory floors or robots. An agent receives inputs or requests from its environment and sends information back to it.

2.3 Purpose

Agents perform a set of tasks on behalf of a user or other agents that are explicitly approved and programmed by users or organizations. Agents essentially need to be their own bosses and have clearly defined goals that they seek to accomplish. Being goal – driven also entails that an agent be proactive rather than just reactive to an environment. Some of the most sophisticated agents can learn as they perform their tasks with new dynamic rules od behavior evolving as they learn user preferences and users needs for specific types of information or actions. For investigative data miners this means agents can be used for the retrieval of specific suspect – or case – related information at predefined intervals and ranges.

2.4 Communications

An agent needs to be able to interact with the user, receive task delegation instructions and inform the user regarding task status and completion through an agent-user interface or through an agent communication language. Agents need to be able to communicate with other agents and humans. Agent and human communication can be via terminal input, such as keyboards or more sophisticated technologies, such as natural language processing and speech recognition. Multi agent communication can take place using standard or defined protocols. Agents allow for scalability, permit software reuse and can handle software evolution and promote open systems. For an agent to work in this environment it should be able to cooperate with its peers and also coordinate efforts. Agents operate continuously upon achievement of their goals they continue to run in the background and monitor the environment. In this context an agent-based application is not supposed to terminate.

2.5 Intelligence

Lastly an agent needs to be able to interpret monitored events to make appropriate decisions for autonomous operation. Agents need to possess a certain degree of intelligence in order to perceive the working environment and be autonomous when performing their programmed tasks. The level of intelligence exhibited by an agent will depend on its function. The dimension of intelligence equates to the degree to which the agent employs reasoning, learning and other techniques to interpret the data to which it has access. The intelligence of agents equates to the degree to which they are able to perceive their environment and change it dynamically. Some agents can incorporate expert systems with predefined rules however dynamic rules generated from machine learning can make them even more intelligent by instilling them with the ability to learn and evolve. For this reason the intelligence of an agent – that is the rules that it follows to complete its designated tasks – can evolve either from the developer or independently from its environment and built – in algorithms.

3. Importance of the Agents

One of the most compelling uses for agent technology is in the area of information retrieval the explosion of information about individuals and companies on the internet and the databases connected to it is huge. Based on studies form Forrester Research and the Yankee Group there are over 2 billion documents on the visible Web with 8-9 million documents being added on daily basis. What is more important is that the web is becoming increasingly database driven and records in these databases cannot be indexed or retrieved using typical search engines. This is due in part to the rise of new technologies like XML and Active Server Page which conventional search engines omit simply because they cannot retrieve the records from these dynamic databases.

These studies indicate that this dynamic web is 400-500 times larger than the noticeable web of 1-2 billion pages. Agent technologies which support special scripting capabilities have the capability to correspond to different information types and thus to retrieve much more information than normal search engines. In other words, agent can sense the type of data source and adjust and convert the parameters into a query that can be understood by the information source. Of course these types of agents can negotiate and extract information not just from Web-connected database but also from local databases, intranets, extranets and other proprietary networks.

Agents are needed to help analysts and investigators deal with and leverage a tremendous amount of data in the course of their work. Agents are sophisticated programs that as we have discussed possess human-like attributes such as the ability to work independently. Communicate, coordinate, learn and accumulate knowledge to conduct their assigned tasks. When used in conjunction with other data mining technologies agents can assist investigators in accessing, organizing and using current and relevant data for security deterrence, forensic analysis and criminal detection.

Agents are designed to perform in a particular environment such as a closed network or the Internet they can also be categorized according to their functionality, such as information retrieval, information filtering, monitoring and alerting etc. They can also be classified according to their core architecture. For the most part there are two major categories of agent that lend themselves to investigative data mining applications – Open sources (Internet) and Secured sources (Intranet) agents.

4. Open Source Agents

These Internet agents provide search services over the web. There are also server specific agents that provide services such as security at the server level. There are internet agents that can serve as information –filtering agents so that based on the security level of users only certain information is passed to them. There are also notification and special services agents and even mobile agents for execution specific tasks like special alerts to wireless devices internet agents are computer programs that reside on servers performing specific data detection, retrieval and delivery tasks to designated users based on preset parameters, behaving very much like intelligent robots. In this context, intelligent agents can play an integral role in the overall process of investigative data mining.

These web robots operate using different Boolean or vector – space strategies when following links and retrieving documents, based on different prioritized methods and schemes. In fact search agents are the most widely used web services. Using keyword query forms, they are easy to use and provide the user an instant response and a hierarchical list of sources of information. Their indexing provides users a universe of information in an instant. In addition some meta – search engines incorporate the knowledge of where to look for information depending on the attributes of the data such as searching for individuals phone numbers, physical address etc.

5. Secured Sources Agents

Intranet agents can serve as database service providers. They can also automate workflow processes and the collaborate communications to intranet or proprietary closed network users. There are intranet agents that can perform resources allocation services which are IT-specific such as updating a data set or deleting a database. These intranet agents can also be programmed to perform a variety of reports and conduct ad hoc analyses of databases across a network. As with internet agents intranet agents can perform similar data – organization tasks for users in closed proprietary secured agency and departmental networks.

An intranet agent is a software program that resides on an internal agency or departmental server or cluster of servers in a private proprietary network. These types of agents are designed to focus on information dissemination among a team of user involved in special task forces or focusing on specific type of data aggregation and analyses. Typically these intranet agents are programmed to assist in accessing internal databases, data marts and data warehouses or proprietary networks. Some can also provide support via wireless devices to field investigators. They enable information sharing within a designated and authorized group of users. They can also be set up to shield and protect unauthorized access to some users and provide alerts when changes to the data occur.

6. Agents Work

Regardless of whether it is an open or a closed type of agent, or of the function that it performs its benefits are usually in automating some type of repetitive behavior that is either time – based or event – based. They can automate repetitive tasks such as performing a common query against a database. More advanced agents can notify specific users of the arrival or creation of new data ready for their analysis; they can assist users with more advanced analyses, guiding them in processes they are not knowledgeable about and lastly they can perform messaging tasks such as notifying users when a model has been completed. Some data mining tools incorporate agents to automate the process of model construction and analysis.

Specialized database or network agents can on the basis of user requests go out and perform queries assemble the data found into a pre-designed template or process the data through a designed analysis. Database agents provide valuable functions in making information available to users in the most useful form and context. Once the data has been retrieved and assembled or the analysis is complete, such as the creation of a data cube or a data mining model the results can be transmitted to a designated group of users. The entire process can be done in real time, the agent can be programmed to perform the task as required by the agency or department needs. The benefits are clear : agents reduce the workload of investigators, lead to faster decision making by the analysts and increase the productivity of everyone involved.

7. Agents Reason or Distinguish

Men and machines such as agents reason through simple to elaborate networks of rules:

IF X, AND Y, THEN Z

Some of these rules are codified from the domain of experts hence the development of expert system in the early 1990s. However, these systems fell out of popularity after some initial enthusiasm when they proved to be expensive to maintain and brittle in deployment. Expert systems represented a set of rules in such areas as making soup or configuring systems or auditing tax returns. Some expert systems still exist.

However there is a different method by which rules can be constructed; this involves data mining. Replacing expert systems as reasoning engines was the development of neural networks and machine-learning algorithms in the area of AI. Rather than developing rules from experts and taking a top-down approach to knowledge acquisition, rules can be extracted from observations in large databases. This is the inductive method of data analysis, now known as data mining. Which uses machine learning and is a bottom-up approach to knowledge acquisition?

These processes of rule creation are not mutually exclusive; in fact a hybrid system is probably the ideal solution for investigative data mining applications, in which some rules are drawn from years of investigators experience, coupled with rules extracted from hundreds of thousands of cases from large databases. This type of man-machine hybrid system is the topic of a proposed data mining architecture. Agents as engines of inference can use both types of rules. To develop intelligence in agents, certain steps can be taken. Briefly they involve the following type of rule sequencing and construction:

- The user or developer provides a set of rules that describe a desired behavior when X happens then do Y. This can be done using a plain-text editor and then transcribed to code-such as C or Jave.
- The reasoning system is next provided with a set of conditional input events, such as when a match of Entity Z898R from List DEA-01/02/04 happens, do Y.
- The reasoning system is provided with interfaces to perform or initiate various desired actions.
- After the reasoning system is initiated it can wait for an event to arrive. It will extract facts from the event and then evaluate its rules to see if the new facts cause any of them to fire. If one or more rules fire it may cause additional action to be initiated or record to be written or updated.

The above process follows a set structure, leading to the creation and use of conditions rules and logic, which can be coded in a variety of ways. Here is an example

IF (Condition 1)
OR (Content A)
AND (Condition 2)
THEN (Action Z)

This can be demonstrated by the example of a system for issuing alerts to say customs agents at point-of-entry stations, based on conditions gleaned from a plate number input into a network system using models developed from both human investigators experience and machine-learning-generated rules. These data mining rules could well have been developed form an extensive analysis of prior convicted cases of contraband prosecutions :

Condition fields:

IF INSURER is None (Condition 1)
Source: Human Domain
OR YEAR is 1988 (Content A)
Source: DMV Registration Record
AND MAKE is CADILLAC (Condition 3)
Source: Data Mining Model
Prediction # 1: THEN ALERT is Medium (Action Z)
Inspect Trunk

8. Intelligent Agents

Unlike an expert system an agent is embedded in its environment and can perceive and react to it using inputs

of conditions. It can dynamically construct new rules as it works; in other words some agents are capable of using sensors to monitor their surroundings develop new rules and then take action independently. For example Doppelgaenger is an agent developed at MIT's Media Lab that uses sensors that provide many kinds of information about the user population in its computing network environment. These sensors can include active badges that provide location information about users along with login information that detects their arrival and departure from the MIT computer network.

This agent can gauge user actions that reflect the frequency and duration of the use of applications, programs and workstations and telephones. The system monitors what applications and data sets users tend to access in order to construct a profile of user preferences so that as it monitors their behaviour it dynamically creates new rules. Doppelgaenger uses this user information to construct profiles into statistical clusters this type of intelligence is used by the agent to provide users with network-specific information likely to be of interest to them. It also uses this behavioural information to provide them with notification about databases and applications that meet their profile interests. It can also alert specific users about changes to data sets they have an interest in.

9. Data Mining Agents

Investigative data mining represents a powerful new approach to criminal monitoring detection and alert dissemination. In private industry, data mining has primarily been applied to very large corporate databases for such applications as indentifying potential customers. While date mining was originally conceived of as a way of extracting hidden associations from large database when coupled with agent technology it can be used to monitor events, extract important information via the Internet, intranets and other proprietary networks discover new patterns, assemble profiles and deliver alerts to military, medical, law enforcement and intelligence agency personnel. Using sensors, agents can work in tandem with other systems to analyse collected data and then issue real-time alerts to systems or personnel via the internet to proprietary networks even to wireless devices.

For example, IBM developed an agent to work with its intelligent data miner suite. The system consists of five agents:

1. A user interface agent that provides a web interface for users to interact with the data miner and help them perform data mining analysis and display results
2. A coordinator agent that is responsible for delegating and managing various tasks that need to be performed for problem solving
3. A data -set agent that is responsible for keeping track of what data is stored in which data mart or data warehouse and actively maintaining metadata information
4. A data mining agent that executes the user-defined algorithms, performs on-line analytical processing analysis and communicates the results to users or other agents
5. A visualization agent that allows for ad hoc and predefined reporting capabilities and a wide array of graphical reports

Data mining software tools are already incorporating agents into their products to assist the user in minimizing the effort of extracting, preparing, modelling and delivering the results of their analyses. Agent technology can be used with text mining technologies to monitor and retrieve specific information via the web and other networks for criminal and terrorist detection and case development.

References

[1] Caglayan, A. and Harrison, C. (1997) Agent Sourcebook, New York : Wiley Computer Publishing.
[2] Franklin, S. and Graesser, A. (1996) " Is It an Agent, or Just a Program? : A Taxonomy for Autonomous Agents," Proceedings of the Third International Workshop on Agent Theories, Architectures and Languages, Springer – Verlag.
[3] Kotz, D. and Gray, B. (1999) " Mobile Agents and the Future of the Internet," Workshop Autonomous Agents, Seattle, WA.
[4] Wooldridge, M. and Jennings, N. (1995) "Intelligent Agents: Theory and Practice," Knowledge Engineering Review, Vol. 10, No. 2.

Anupam is a Ph.D. Research Scholar in the Research Department of Computer Science and Application, at JJTU,Jhunjhunu (Raj.) . He holds a Master degree in Computer Application from V.B.S. Purvanchal University, Jaunpur. He finished M.Sc. in Information Technology from Punjab Technical University, Jalandhar. Anupam is having 12 years of teaching experience.