AC 2010-1879: WRMT CASE STUDY: GIS WITH RULE-BASED EXPERT SYSTEM

Andrew Ernest, Western Kentucky University

Jana Fattic, Western Kentucky University

Ni-Bin Chang, University of Central Florida

Shalini Chitrapu, Western Kentucky University

Paige Davenport, Western Kentucky University

WRMT Case Study: GIS with Rule-based Expert System for Optimal Planning of Sensor Network in Drinking Water Systems

Abstract

This paper provides a case study in the application of the concepts of the Water Resource Management Technologies technology transfer concept presented at the 2009 conference.

The Technology Transfer Model[1]. The Center for Water Resource Studies at Western Kentucky University promotes and facilitates a public/private sector partnership focused on the development and transfer of water resource management technologies that specifically target the small to medium sized industry market. This market includes municipalities, water and wastewater utilities and districts serving populations less than 25,000, local and state government agencies, commercial and non-profit organizations providing engineering, scientific, technical, financial, managerial and analytical services, and the industry relevant trade associations. The size and scope of services needed/offered in this market sector tend to limit the rate of return of product development investment, and, as a result, only a few companies tend to invest in this sector. The partnership integrates small technology startup entrepreneurial firms with the end-user/target market sector and the water resource technology development capacity of the partnering universities to facilitate the translation of market need into technological concept, development, transfer and commercialization. The needs and capacity of the target market dictate a high-volume, low-margin approach to be commercially feasible, and so is typically under-served. By relying on small startup firms for commercialization, and minimizing licensing burden, the partnership promotes a technology development and transfer model process that is sustainable. The partnership focuses on developing processes for rapid identification, development, transfer and commercialization of incremental advances in technology that have an immediate benefit on the target market. The key to success of the partnership is the adoption of the successful "Red Hat Business Model," that relies on integrating the end-user into an open product development process. Key principles incorporated into such a business model include a "first-to-market" philosophy, a highly responsive consumer feedback process, and a balanced reliance on product service and intellectual property protection for commercialization.

The Technology. Emergency Response Plans (ERP) for water supply systems are now required under the Bioterrorism Act of 2002[2]. Community water systems the U.S. are required to perform a Vulnerability Assessment (VA) and incorporate it into their ERP. An increased reliance by rural communities on sensor networks over traditional manual sampling campaigns is a likely outcome of this requirement[3]. The technical goal of this work is to develop a rules-based decision support tool integrated into in a Geographical Information System (GIS). This system will serve as the platform for incorporating sensor placement, sensitivity and system responsiveness rules that will simplify the adoption of low-cost Supervisory Control and Data Acquisition (SCADA) systems by small utilities as a tool in the ERP.

The Case Study. This paper tracks the development of this technology from concept, through an evolutionary process involving sequential refinement and adaptation as different funding opportunities and priorities are realized into its current form and commercialization stage under the WRMT model.

Background

In 2006, an idea for a software application targeted specifically at small and rural water utilities was conceived by collaborators from the Center for Water Resource Studies (CWRS) at Western Kentucky University (WKU), the University of Central Florida (UCF) and Spatial Data Integrations, Inc (SDI). The CWRS, in partnership with the UCF, was awarded a grant from the Kentucky Science and Engineering Foundation in 2007 to refine and commercialize this concept, building on expert system technology and hydraulic modeling capacity the partnering institutions.

Center for Water Resource Studies. The CWRS is a unit of the Applied Research and Technology Program (ARTP) at WKU focused on capacity development for small and rural water resource management entities. The ARTP, a student engagement and entrepreneurship stimulation initiative of WKU, serves as a catalyst for regional economic development by developing a creative workforce and providing technological and environmental services.

University of Central Florida. The Multiscale Sustainable Systems Engineering program at UCF focuses on the concept of sustainability, developing and applying forward-looking, risk-informed, and cost effective decision-making models that combine social and economic factors with broad-based considerations of potential future environmental impacts.

Spatial Data Integrations, Inc. Spatial Data Integrations, Inc. (SDI), a full service geospatial firm located in Louisville, Kentucky, offers a wide variety of geospatial solutions that are enabled by utilizing a host of services to collect, analyze, integrate, design and manage data and information linked to location.

Intended Hybrid Business Model. A hybrid product/service business model was initially conceived as the intended commercialization approach for the technology under development. Under this scenario, a functional toolset would be developed. However, unlike traditional licensing commercialization models, the IP protected toolset would be deployed within a business environment that encouraged the engagement of students is technical service provision to the end-user while maximizing commercial profitability. Although units such as the CWRS within the ARTP environment are able to engage students in contract service activities, as tax subsidized entities, their ability to compete commercially is limited. More competitive models that encourage student entrepreneurship and engagement were sought. Two concepts were identified for further development – an employee-owned commercial services business entity paralleling the university-house research/technology development unit, and a for-profit subsidiary of an applicable component of the host university research enterprise (e.g. research foundation, technology incubator, etc).

Technical Objectives. The research team aim to help rural communities efficiently and cost-effectively complete the deployment of sensors and sensor networks for monitoring water supply systems in a user-friendly environment with the aid of a rule-based expert system embedded in a geographical information system (GIS) platform. The design-basis involves developing a customized sensor and control/response network, a rule-based expert system, and the analysis, display and response methods that enhance a unique spatial visualization leading to promote the interactions between the designers, the operators, and the end-users. System costs will be minimized by using the expert system to design a sensor network that triggers response based on changes in overall system state, rather than the more expensive route of detecting specific intrusions with vector-specific sensors.

Technical Project Details

There are two primary technical aspects of this project.

- Identification of water demand patterns, contamination scenarios, and simulation analysis.
- Development of the rules-based expert system.

Identification of water demand patterns, contamination scenarios, and simulation analysis was conducted by the UCF team. The EPANET[4] toolkit was used to setup simulations of hydraulic and water quality scenarios. Outputs from EPANET were analyzed by engineers and experts for creating rules and assessment of sensor deployment. Types, amount, and locations of sensors were assessed based on the historical water usage data and the simulations' outcomes from the EPANET. Hardin County District #1 (HCWD), located in Radcliffe, Kentucky, provided monthly historical water usage data and spatial data. Spatial data of the water distribution network (piping network) were also provided. The spatial data of the network was converted and exported into the EPANET file format. Simulations were conducted based on multi-dimensional scenarios. EPANET offers various capabilities including hydraulic and water quality assessments. Spatial data is imported into EPANET software as .INP file format. Sources of water, pumps, and tanks are manually added into the model. Pumps' characteristic can be adjusted according to the actual specifications. Storage tanks can be individually characterized. The simulation was set to run for 72 hours. Once complete simulations of the network were assessed, the next step is to deploy sensors based on the needs.

Contamination scenarios were arranged with respect to the actual conditions that are highly likely to happen in real life. In this study, three routine contamination scenarios were organized as possible events whereas three more were designed as special (highly unlikely) events - events of waterborne disease outbreak and terrorist attack. Within three routine contamination scenarios, one of them was configured for reflecting the impacts of lead release in WDSs. The other two routine contamination scenarios were related to chlorination and its impact due to the by-products.

During the process of creating the scenarios and their simulations, EPANET was integrated into the GIS platform. Tools were developed to correct problems associated with working with real world spatial data. Problems included: mains not split at intersections, mains not split at junctions (meters, fittings, etc), fittings missing at intersections of mains, elevation data missing, missing mains (unconnected pipe network), junctions not snapped to mains, and meter locations do not necessarily match locations in water usage file.

A rule-based expert system is software that provides the knowledge of an expert. This knowledge can be used to answer a specific problem that would have otherwise required a human expert to be consulted. This is accomplished by using a rule base, facts, and an inference engine. The rule base is the knowledge of an expert. The facts are the information about the real world problem or environment. The inference engine analyses the rules and facts to produce a result to answer the problem.

Within a programming environment, logic is frequently repeated multiple times throughout a system. As code is edited and updated, changes must be kept up to date. Logic will get out of sync and will become hard to maintain. By creating and using a rules-based system which maintains the expert logic within a central location this duplication of effort and difficulty in maintenance can be avoided. Ideally a truly centralized logic location for our data is optimal.
Rules may be implemented by writing programming code. Rules may be additionally coded within a scheme described by metadata (such as XML); the actual processing though is done by code. A rule is either broken or not broken, a Boolean condition. This indicated whether the condition was satisfied. Simple rules based engines simply chain procedural logic together in an order that you specify. Most offers sophisticated matching algorithms like Rete[5] to connect facts with rules, determine which rules should be run and in what order. The Rete algorithm creates a node tree based on the rules. Facts are used as parameters of the rules. The facts begin at the nodes at the top of the tree. They then go down the tree checking for conditions in which they match. Rules based systems are designed to allow the user flexibility in changing and adding rules without the

need of a complete software rewrite. The rules are meant to be more malleable and adaptable to change. Rules then become more dynamic. Using rule grouping the rules may be used in whole or in part by a system using the same rules based system as a backend.

There are two types of chain procedural logic for a rule system: Forward Chaining and Backward Chaining. Forward chaining is when the facts are searched against the inference rules until the desired result is found. Backward chaining is where the inference engine begins with a list of goals and then tries to find conclusions to these goals by working backwards.

The rule-based expert system was created by modifying a rule engine, jDREW[6] and the extension OO (Object-Oriented) jDREW. jDREW is a deductive reasoning inference engine written in Java. It is a powerful, configurable engine designed for facts and rules. jDrew has several application programmer's interfaces (APIs) that can be used to embed its engine into other applications. A Graphical User Interface (GUI) of the engine is shown in Figure 1. It can be utilized on a server, client, or as a part of a larger application written in Java. jDREW uses Prolog and RuleML formats. It is both backward and forward reasoning. OO jDREW, a deductive reasoning engine for the RuleML rule language, is an Object Oriented extension to jDREW. OO jDREW implements Object Oriented extensions to RuleML which are Order Sorted Types, Slots, and Object Identifiers. The RuleML format is a XML schema, which is easy to define, validate, and convert data.

OO jDREW contains a rule base, a place for facts, and an induction engine. Rules and facts are stored in the RuleML format. The rule base is already created based on the knowledge of an expert. The facts are dynamically created with each project based on the spatial data provided by the user. The inference engine processes the data using forward chaining and/or backwards chaining to produce the result.

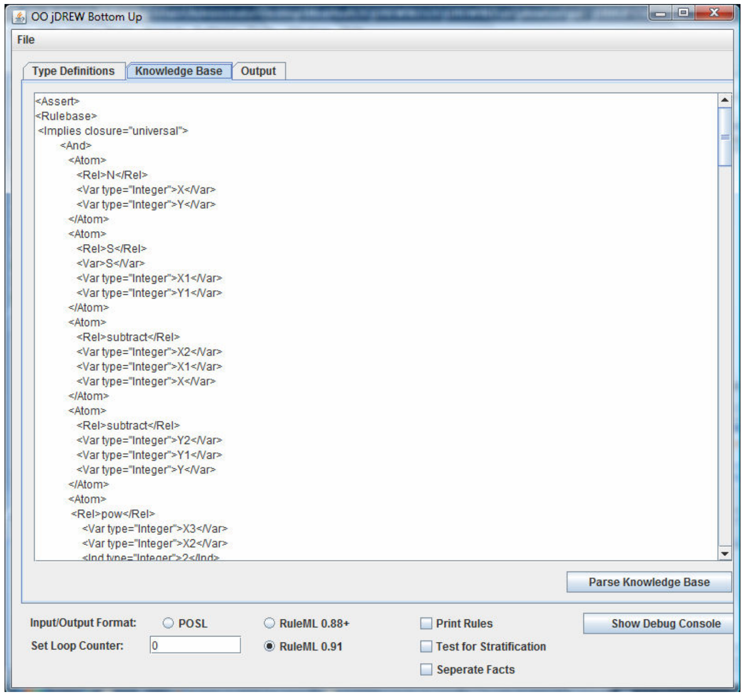

Figure 1 – OO jDrew inference engine GUI

New tools were developed and integrated into the rule-based system to increase usability and cross functionality with many other existing programs. This was accomplished by adding conversion capabilities to the program. The conversion tool was created with the OGR Simple Features Library which is an open source library written in C++ programming language. This tool allows users to convert existing GIS data into many different formats.

These formats include ESRI Shapefile, MapInfo file, BNA, CSV, KML, GMT, GPX, GXT, and GML. ESRI Shapefile is geospatial vector data format used by ESRI GIS software. Mapinfo file is used by MapInfo Professional software. BNA is a format by Atlas Boundary used by Softwright Software and Golden Software. CSV is used for the digital storage of data structured in a table of lists form, where each associated item (member) in a group is in association with others also separated by the commas of its set. This format can be used by anyone with a text editor. KML format is used to display geographic data in an Earth browser such as Google Earth, Google Maps, and Google Maps for mobile. GMT is a part of NetCDF (network Common Data Form) which is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. This format is open source for any software product to use. GPX can share GPS data (waypoints, routes, and tracks) across different applications and Web services on the Internet written in a a light-weight XML data format. It is used in many software products, a few free. GXT format is used by GeoConcept Company.

The GML format was chosen to be the standard format in used in this project. It was chosen because it is the open source standard used by the Open Geospatial

Consortium (OGC) and is a XML schema. Many of the software already mentioned can use the GML format. Several popular open source software, such as GRASS and MapWindow GIS, use the format as well. The rule-based system converts GIS data to the GML format. A parser written in Java programming language was created. The parser takes the GIS data that is in the GML format and translates it into a RuleML file. The RuleML file is used in the rule engine OO jDREW to create a fact base. The parsing component dynamically locates each GIS data set and places them into separate relations within the RuleML file and associates the data sets attributes into relating variables. After OO jDREW creates results based on facts and rules then the RuleML is parsed into GML and added to the GIS as a new layer. Figure 2 represents the workflow of the application. The goal of using the GML format is to increase usability. The GML format is becoming an industry standard. The GML format helps make the rule-based system compatible with many of the most popular GIS software available.

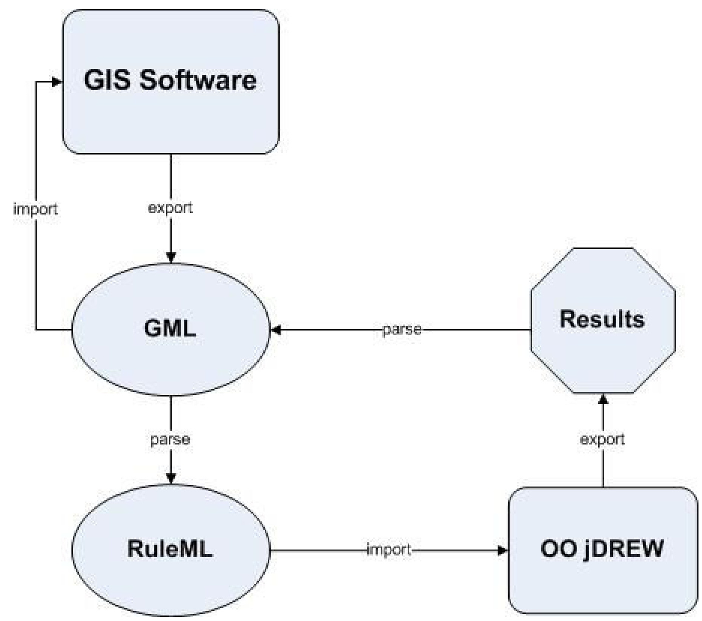

Figure 2 – Diagram of the rule-based engine's workflow

Current Status

The technology development is currently nearing completion, and is entering the prototype deployment and validation phase. The final activity, development of a robust commercialization plan is the find activity to be performed for this project. Open Environment, PLLC (http://open-environment.com), was established as the intended licensee of the technology. The business plan for this company included both professional services and technology deployment as the primary revenue generation sources. Ownership and business organizational structure are yet to be finalized, with the outcome of the commercialization plan development being a primary driver. Of critical importance is the ability to engage students in business activities beyond the traditional professional service role (i.e. entrepreneurial activities).

Acknowledgements

The work presented herein is funded in part through a grant from the Kentucky Science and Engineering Foundation, project number KSTC-144-401-07-016, COMMFUND-713-RFP-006.

Bibliography

1. Ernest, A. N. S., N-B. Chang, R. Fowler, J. R. Fattic, K. Andrew, and J. Ballweber, "Water Resource Management Capacity Development: A Small Systems Technology Transfer Model", 2009 ASEE Annual Conference and Exposition, Austin, Texas, American Society for Engineering Education, 06/2009.
2. Public Health Security and Bioterrorism Preparedness and Response Act of 2002. Public Law 107–188 116 Stat. 682 (2002)
3. Chang, N. B.and Makkeasorn, 2007. A.Rule-based Expert System for Sensor Deployment in Drinking Water Systems for Rural Communities, Environmental Informatics Archives, Volume 5 (2007), 455- 463, EIA07-050
4. United States Environmental Protection Agency, 2000. EPANET2 Users Manual. EPA/600/R-00/057. http://www.epa.gov/nrmrl/wswrd/dw/epanet.html. September 2000
5. Forgy, C. L., "Rete: A Fast Algorithm for the Many Pattern/ Many Object Pattern Match Problem", Artificial Intelligence, vol. 19, pp. 17-37, 1982.
6. Bruce Spencer. jDrew, A Java Deductive Reasoning Engine for the Web. http://www.jdrew.org/jDREWebsite/jDREW.html