Designing the automatic web tools for small business scale with Case-based reasoning

Sandeep Rathee¹ and Dr. Gundeep Tanwar²
¹Student, M.Tech (CSE), Bhiwani Institute of Technology & Science, Bhiwani, INDIA
²Associate Professor, BRCM College of Engineering & Technology Bhiwani, INDIA

Abstract
This paper shows how humans and machines can better collaborate to acquire adaptation knowledge (AK) in the framework of a case-based reasoning (CBR) system whose knowledge is encoded in a semantic wiki. Automatic processes like the CBR reasoning process itself, or specific tools for acquiring AK are integrated as wiki extensions. These tools and processes are combined on purpose to collect AK. Users are at the center of our approach, as they are in a classical wiki, but they will now benefit from automatic tools for helping them to feed the wiki. In particular, the CBR system, which is currently only a consumer for the knowledge encoded in the semantic wiki, will also be used for producing knowledge for the wiki. A use case in the domain of cooking is given to exemplify the man-machine collaboration. The managers should be skilled to make better decisions in the business organization. They also need a supportive environment where they won't be unfairly criticized for making wrong decisions. Decision making increasingly happens at all levels of a business. Business Intelligence delivers the appropriate data at the proper time and in the precise arrangement. It offers user-friendly information openly to users where they can work, team up, and make resolutions. Although the business intelligence have own button necks, hence the performance of business intelligence system goes down rapidly. In this paper, we are applying case-based reasoning (CBR) technologies in the business intelligence that overcomes the deficiencies of existing system. The business decision are being taken through the concept of case base reasoning approach.

Keywords: Small Scale Business, Web tools, Case-based reasoning, Business Intelligence, Automatic Web business tool

I. INTRODUCTION

Small businesses are normally privately owned corporations, partnerships, or sole proprietorships. What constitutes "small" in terms of government support and tax policy varies by country and by industry. Small businesses can also be classified according to other methods such as sales, assets, or net profits. Small businesses are common in many countries, depending on the economic system in operation. Typical examples include: convenience stores, their small shops (such as a bakery or delicatessen), hairdressers, tradesmen, lawyers, accountants, restaurants, guest houses, photographers, small-scale manufacturing, and online business, such as web design and programming, etc.

Small scale industrial units are those engaged in the manufacture, processing or preservation of goods and whose investment in plant and machinery (original cost) does not exceed Rs.1 crore. The investment limit of Rs. 1 crore for classification as SSI has been enhanced to Rs.5 crore in respect of certain specified items under hosiery, hand tools, drugs & pharmaceuticals, stationery items and sports goods by the Government of India. Independence is another advantage of owning a small business. One survey of small business owners showed that 38% of those who left their jobs at other companies said their main reason for leaving was that they wanted to be their own bosses. Freedom to operate independently is a reward for small business owners. In addition, many people desire to make their own decisions, take their own risks, and reap the rewards of their efforts. Small business owners have the satisfaction of making their own decisions within the constraints imposed by economic and other environmental factors. However, entrepreneurs have to work for very long hours and understand that ultimately their customers are their bosses.

Small businesses often face a variety of problems related to their size. A frequent cause of bankruptcy is undercapitalization. This is often a result of poor planning rather than economic conditions - it is common rule of thumb that the entrepreneur should have access to a sum of money at least equal to the projected revenue for the first year of business in addition to his anticipated expenses. Failure to provide this level of funding for the company could leave the owner liable for all of the company's debt should he end up in bankruptcy court,
under the theory of undercapitalization.

Figure 1: Small scale Business Components

In addition to ensuring that the business has enough capital, the small business owner must also be mindful of contribution margin (sales minus variable costs). To break even, the business must be able to reach a level of sales where the contribution margin equals fixed costs. When they first start out, many small business owners underprice their products to a point where even at their maximum capacity, it would be impossible to break even. Cost controls or price increases often resolve this problem.

**Entrepreneurial Myth**

Another problem for many small businesses is termed the 'Entrepreneurial Myth' or E-Myth. The mythic assumption is that an expert in a given technical field will also be expert at running that kind of business. Additional business management skills are needed to keep a business running smoothly. Still another problem for many small businesses is the capacity of much larger businesses to influence or sometimes determine their chances for success.

**Small business bankruptcy**

When small business fails, the owner may file bankruptcy. In most cases this can be handled through a personal bankruptcy filing. Corporations can file bankruptcy, but if it is out of business and valuable corporate assets are likely to be repossessed by secured creditors there is little advantage to going the expense of a corporate bankruptcy. Many states offer exemptions for small business assets so they can continue to operate during and after personal bankruptcy. However, corporate assets are normally not exempt, hence it may be more difficult to continue operating an incorporated business if the owner files bankruptcy.

**Social Responsibility**

Small businesses can encounter several problems related to Corporate social responsibility due to characteristics inherent in their construction. Owners of small businesses often participate heavily in the day-to-day operations of their companies. This results in a lack of time for the owner to coordinate socially responsible efforts. Additionally, a small business owner's expertise often falls outside the realm of socially responsible practices contributing to a lack of participation. Small businesses also face a form of peer pressure from larger forces in their respective industries making it difficult to oppose and work against industry expectations. Furthermore, small businesses have more personal relationships with their patrons and local shareholders they must also be prepared to withstand closer scrutiny if they want to share in the benefits of committing to socially responsible practices or not.

**Job Quality**

While small businesses employ over half the workforce and have been established as a main driving force behind job creation the quality of the jobs these businesses create has been called into question. Small businesses generally employ individuals from the Secondary labor market.

**II. Web Tools**

In dynamic composition, automated tools are used to analyze a user query, and select and assemble Web service interfaces so that their composition will solve the user demand. From a user perspective, the composite service will continue to be considered as a simple service, even though it is composed of several Web services.

In order to support greater automation of service selection and invocation, recognition is growing of the need for richer semantic specifications of Web services, so as to enable fuller, more flexible automation of service provision and use, support the construction of more powerful tools and methodologies, and promote the use of semantically well-founded reasoning about services. As a result, Web services have semantic descriptions in addition to their traditional standard syntactic description (WSDL). This is referred to as semantic Web services.

Semantic Web services solve Web service problems semantically and address Web services descriptions as a whole. Semantic markup languages such as OWLS, WSDL-S and SAWSDL describe Web service capabilities and contents in a computer-interpretable language and improve service discovery, invocation, composition, monitoring, and recovery quality. Several methods and tools have been proposed for dynamic Web service composition. The majority of researches conducted in dynamic composition have their origins in the realm of artificial intelligence.
The World Wide Web (Web) has become the best known example of a hypermedia system. To date, numerous organisations across the Globe have developed thousands of commercial and/or educational Web applications. The Web has been used as the delivery platform for two types of applications: Web hypermedia applications and Web software applications. A Web hypermedia application is a non-conventional application characterised by the structuring of information using nodes (chunks of information), links (relations between nodes), anchors, access structures (for navigation) and the delivery of this structure over the Web. Whereas a Web software application represents any conventional software application that depends on the Web, or uses the Web's infrastructure, for execution. Typical applications include legacy information systems such as databases, booking systems, knowledge bases etc. Many e-commerce applications fall into the latter category.

![Figure 2: Web tools working](image)

Our research focus is on proposing and comparing development effort prediction models for Web hypermedia applications. These applications have a potential in areas such as software engineering, literature, education, and training, to mention but a few. Readers interested in effort estimation models for Web software applications. This paper has two objectives. The first is to describe the application of case-based reasoning to estimating the effort for developing Web hypermedia applications. The second is to compare the prediction accuracy of several CBR configurations, based on two datasets of Web hypermedia projects.

Those objectives are reflected in the following research questions:

1. Will different CBR configurations generate significantly different prediction accuracy?
2. Which of the CBR configurations employed in this study gives the most accurate predictions for the datasets employed?

These issues are investigated using data gathered through two case studies where sets of suggested metrics for effort prediction were measured. These metrics reflect current industrial practices for developing multimedia and Web hypermedia applications. Both case studies gathered data on Web hypermedia projects developed by Computer Science Honours or postgraduate students attending a Hypermedia and Multimedia Systems course at the University of Auckland. The first case study took place during the first semester of 2000 and the second, a replication of the first, took place during the first semester of 2001. The metrics collected were classified as size, effort and confounding factors.

Both case studies looked at effort prediction for design and authoring processes, adopting the classification proposed by Lowe and Hall. Using this classification, authoring encompasses the management of activities for the actual content and structure of the application and its presentation. Design covers the methods used for generating the structure and functionality of the application, and typically does not include aspects such as application requirements elicitation, feasibility consideration and applications maintenance.

In this paper we chose to use CBR as a prediction technique for the following reasons:

- An early study comparing CBR to other prediction techniques, CBR presented the best prediction accuracy when using a dataset of Web hypermedia projects.
- A second study, using a different dataset, also showed good prediction accuracy using CBR. CBR is an intuitive method and there is evidence that experts apply analogic reasoning when making estimates.
- CBR is simple and flexible, compared to algorithmic models.
- CBR can be used on qualitative and quantitative data, reflecting closer types of datasets found in real life.

The rationale for CBR is to characterise the project, for which the estimate is to be made, relative to a number of attributes (e.g. application complexity, link complexity etc). This description is then used to find other similar already finished projects, and an estimate for the new project is made based on the known effort values for those finished projects.

### III. RELATED WORK

To our knowledge, there are relatively few examples in the literature of studies that investigate effort prediction models generated using data from Web hypermedia applications. Most research in Web/hypermedia engineering has concentrated on the proposal of methods, methodologies and tools as a basis for process improvement and higher product quality.
Mendes et al. describe a case study involving the development of Web sites structured according to the Cognitive Flexibility Theory (CFT) principles in which simple size metrics were collected. Several prediction models are generated (linear regression, stepwise regression and case-based reasoning) and compared. Results show that the best predictions were obtained using case-based reasoning, confirming similar results using non-hypermedia applications. Their metrics can be applied to any hypermedia application, and represent an initial step towards the proposal of development effort prediction models for Web hypermedia applications. However, some of these metrics are subjective, which may have influenced the validity of their results.

Mendes et al. describes a case study evaluation in which 37 Web hypermedia applications were used. These were also structured according to the CFT principles and the Web hypermedia metrics collected were organised into five categories: length size, complexity size, reusability, effort and confounding factors. The size and reusability metrics were used to generate top down and bottom up prediction models using linear and stepwise regression techniques. These techniques were then compared based on their predictive power and stepwise regression was not shown to be consistently better than multiple linear regression. A limitation of this study is that it only compared prediction models generated using algorithmic techniques. The same dataset was also used as input to a case based reasoning tool to predict effort, where results obtained were most favourable.

Mendes et al. presents a case study where size attributes of Web hypermedia applications were measured. Those attributes correspond to three size categories, namely Length, Complexity and Functionality. For each size category they generated prediction models using linear and stepwise regression. The accuracy of the predictions for those six models was compared using boxplots of the residuals. Results suggested that all the models offered similar prediction accuracy. The limitation of this study is also that it only compared prediction models generated using algorithmic techniques.

Fewster and Mendes propose a prediction model for authoring and designing Web applications and for project risk analysis using a General Linear Model. They suggest two types of metrics: those applied to static Web hypermedia applications and those applied to dynamic Web hypermedia applications. The former metrics can also be used to measure other types of hypermedia and Web applications. Their results were also an initial step towards the proposal of effort prediction models for Web hypermedia development.

IV. Case-based Reasoning

Case-based reasoning is a problem solving paradigm that in many respects is fundamentally different from other major AI approaches. Instead of relying solely on general knowledge of a problem domain, or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to utilize the specific knowledge of previously experienced, concrete problem situations (cases). A new problem is solved by finding a similar past case, and reusing it in the new problem situation. A second important difference is that CBR also is an approach to incremental, sustained learning, since a new experience is retained each time a problem has been solved, making it immediately available for future problems.

The CBR field has grown rapidly over the last few years, as seen by its increased share of papers at major conferences, available commercial tools, and successful applications in daily use. All case-based reasoning methods have in common the following process:

- retrieve the most similar case (or cases) comparing the case to the library of past cases;
- reuse the retrieved case to try to solve the current problem;
- revise and adapt the proposed solution if necessary;
- retain the final solution as part of a new case.

There are a variety of different methods for organizing, retrieving, utilizing and indexing the knowledge retained in past cases. Retrieving a case starts with a (possibly partial) problem description and ends when a best matching case has been found. The subtasks involve:

- identifying a set of relevant problem descriptors;
- matching the case and returning a set of sufficiently similar cases (given a similarity threshold of some kind); and
- selecting the best case from the set of cases returned.

![CBR Cycle](image-url)
Some systems retrieve cases based largely on superficial syntactic similarities among problem descriptors, while advanced systems use semantic similarities. Reusing the retrieved case solution in the context of the new case focuses on: identifying the differences between the retrieved and the current case; and identifying the part of a retrieved case which can be transferred to the new case. Generally the solution of the retrieved case is transferred to the new case directly as its solution. Revising the case solution generated by the reuse process is necessary when the solution proves incorrect. This provides an opportunity to learn from failure.

Retaining the case is the process of incorporating whatever is useful from the new case into the case library. This involves deciding what information to retain and in what form to retain it; how to index the case for future retrieval; and integrating the new case into the case library.

A CBR tool should support the four main processes of CBR: retrieval, reuse, revision and retention. A good tool should support a variety of retrieval mechanisms and allow them to be mixed when necessary. In addition, the tool should be able to handle large case libraries with retrieval time increasing linearly (at worst) with the number of cases.

Applications
Case-based reasoning first appeared in commercial tools in the early 1990's and since then has been used to create numerous applications in a wide range of domains:

- Diagnosis: case-based diagnosis systems try to retrieve past cases whose symptom lists are similar in nature to that of the new case and suggest diagnoses based on the best matching retrieved cases. The majority of installed systems are of this type and there are many medical CBR diagnostic systems.
- Help Desk: case-based diagnostic systems are used in the customer service area dealing with handling problems with a product or service.
- Assessment: case-based systems are used to determine values for variables by comparing it to the known value of something similar. Assessment tasks are quite common in the finance and marketing domains.
- Decision support: in decision making, when faced with a complex problem, people often look for analogous problems for possible solutions. CBR systems have been developed to support in this problem retrieval process (often at the level of document retrieval) to find relevant similar problems. CBR is particularly good at querying structured, modular and non-homogeneous documents.
- Design: Systems to support human designers in architectural and industrial design have been developed. These systems assist the user in only one part of the design process, that of retrieving past cases, and would need to be combined with other forms of reasoning to support the full design process.

V. DESIGNING AUTOMATIC TOOLS WITH CBR

Some of the characteristics of a domain that indicate that a CBR approach might be suitable include:
1. records of previously solved problems exist;
2. historical cases are viewed as an asset which ought to be preserved;
3. remembering previous experiences is useful;
4. specialists talk about their domain by giving examples;
5. experience is at least as valuable as textbook knowledge.

Case-based reasoning is often used where experts find it hard to articulate their thought processes when solving problems. This is because knowledge acquisition for a classical KBS would be extremely difficult in such domains, and is likely to produce incomplete or inaccurate results. When using case-based reasoning, the need for knowledge acquisition can be limited to establishing how to characterize cases. Case-based reasoning allows the case-base to be developed incrementally, while maintenance of the case library is relatively easy and can be carried out by domain experts.

The previous methods tried to find a compact representation of the data that can be used for future prediction. In case-based reasoning, the training examples - the cases - are stored and accessed to solve a new problem. To get a prediction for a new example, those cases that are similar, or close to, the new example are used to predict the value of the target features of the new example. This is at one extreme of the learning problem where, unlike decision trees and neural networks, relatively little work must be done offline, and virtually all of the work is performed at query time. Case-based reasoning can be used for classification and regression. It is also applicable when the cases are complicated, such as in legal cases, where the cases are complex legal rulings, and in planning, where the cases are previous solutions to complex problems.

If the cases are simple, one algorithm that works well is to use the k-nearest neighbors for some given number k. Given a new example, the k training examples that have the input features closest to that example are used to predict the target value for the new example. The prediction can be the mode, average, or some interpolation between the prediction of these k training
examples, perhaps weighting closer examples more than distant examples.

For this method to work, a distance metric is required that measures the closeness of two examples. First define a metric for the domain of each feature, in which the values of the features are converted to a numerical scale that can be used to compare values. Suppose val(e,Xi) is a numerical representation of the value of feature Xi for the example e. Then (val(e1,Xi) - val(e2,Xi)) is the difference between example e1 and e2 on the dimension defined by feature Xi. The Euclidean distance, the square root of the sum of the squares of the dimension differences, can be used as the distance between two examples. One important issue is the relative scales of different dimensions; increasing the scale of one dimension increases the importance of that feature. Let wi be a non-negative real-valued parameter that specifies the weight of feature Xi. The distance between examples e1 and e2 is then
d(e1,e2) = \sqrt{\sum wi × (val(e1,Xi) - val(e2,Xi))^2}.

The feature weights can be provided as input. It is also possible to learn these weights. The learning agent can try to find a parameter setting that minimizes the error in predicting the value of each element of the training set, based on every other instance in the training set. This is called the leave-one-out cross-validation error measure.

VI. IMPLEMENTING BUSINESS INTELLIGENCE IN WEB TOOLS

In case-based reasoning, knowledge representation covers the cases stored in the expert system’s database (case representation), and the adaptation strategies used to adapt these cases to fit a new situation. Each case has a semantic and a syntactic representation. The core idea of CBR is case, which is depiction of analyzed situation (problem) together with its solution (decision, plan, etc.). The case is situation, which occurred in the past, remembered by the computer and later reused in another similar problem. The concept of case here is very wide.

In medicine it may by a set of symptoms occurring during patient examination together with doctor’s diagnosis. In HelpDesk systems it will be description of error in supported product together with specification of what kind of action the user needs to take. In this approach the knowledge is stored without changes. CBR requires selection of good knowledge representation at first. The individual cases, consisting of problems and their solutions should be encoded in a way allowing for efficient search. Moreover, knowledge representation must support the possibility to modify or adjust the existing solution so that it can be applied in the specific case under consideration. Thus the reuse of knowledge must comprise elements of creativity. Some further elements consist in appropriate revision of the final solution and storing it in an appropriate form in the knowledge base for further use.

In practical implementations one can find many methods of representing cases depending on the specific domain or format of existing data. The main ways of representation may by divide into the following classes: simple set of properties of described case with values (database models), object oriented schemes, specific representation (graphs, plans). Most frequently cases are described as sets of attributes with their values.

VII. CONCLUSION AND FUTURE WORK

The case based approach make very impact on performance of business intelligence system. It tries to overcome the bottom necks of business intelligence. It has various effects on business intelligence system. First, it builds the business cases & stores them in its case base. Hence the various business cases are defined at various level of organization. Second, the solutions that are found are mostly appropriate. In Revise phase, the solutions are analyzed & found faults are repaired. CBR can be used to remind decision makers of many specific factors which were considered relevant in past cases which may now have been forgotten. If the combination of human decision maker and computer leads to better decisions, the evolving human-machine relationship may help reduce users’ reluctance to engage with CBR. Hence the faults do not propagate further in future. So the solutions may corporate the business needs and objectives of the company. Third, although there is a lack of high-quality data, then the business decision are made on basis of business cases that make effective analysis of the incomplete information.

REFERENCES


