ELSEVIER

Knowledge-Based Systems 10 (1998) 449–461

Knowledge-Based SYSTEMS

A Data Mining methodology for cross-sales

S.S. Anand[a,*], A.R. Patrick[a], J.G. Hughes[a], D.A. Bell[b]

[a]Northern Ireland Knowledge Engineering Laboratory, Faculty of Informatics, University of Ulster, Shore Road, Newtownabbey, County Antrim BT37 0QB, UK

[b]School of Information and Software Engineering, Faculty of Informatics, University of Ulster, Shore Road, Newtownabbey, County Antrim BT37 0QB, UK

Received 7 August 1997; accepted 3 November 1997

Abstract

In this paper we discuss the use of Data Mining to provide a solution to the problem of cross-sales. We define and analyse the cross-sales problem and develop a hybrid methodology to solve it, using characteristic rule discovery and deviation detection. Deviation detection is used as a measure of interest to filter out the less interesting characteristic rules and only retain the best characteristic rules discovered. The effect of domain knowledge on the interestingness value of the discovered rules is discussed and techniques for refining the knowledge to increase this interestingness measure are studied. We also investigate the use of externally procured lifestyle and other survey data for data enrichment and discuss its use as additional domain knowledge. The developed methodology has been applied to a real world cross-sales problem within the financial sector, and the results are also presented in this paper. Although the application described is in the financial sector, the methodology is generic in nature and can be applied to other sectors. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Cross-sales; Data Mining; Characteristic rule discovery; Deviation detection

1. Introduction

Data Mining, the semi-automated search for hidden, previously unknown and interesting knowledge has over the last decade been an area of great interest to academia as well as to industry [3]. Already a number of real world applications of Data Mining exist in different sectors of industry from shopping basket analysis [4] to space exploration [5].

This paper discusses the use of Data Mining in the previously unexplored area of customer cross-sales (see Section 2 for a detailed definition and analysis of cross-sales). The structure of the paper follows an eight-stage Data Mining process (Fig. 1) used by the authors when tackling real world problems using Data Mining [6]. The first stage of the process is discussed in this section. The following sections of the paper tackle the other stages with respect to the authors' experiences in solving the problem of cross-sales. A hybrid methodology is developed from analysis of the cross-sales problem and a technique is presented for measuring how useful the rules discovered are in the light of having only positive examples of the target product customers. The use of externally procured data is discussed and a novel way of refining the domain knowledge available, to maximise the interest of the rules produced, is described.

In Stage 1 of the Data Mining process, *Human Resource Identification*, the human resources that should be involved in the project and their respective roles are identified. In most real world Data Mining problems the human resources required are the domain expert, the data expert and the Data Mining expert. Normally, Data Mining is carried out in large organisations where the prospect of finding a domain expert who is also an expert in the data stored in the organisation is rare. For example, in the cross-sales project described here, management were particularly interested in targeting Household Insurance customers. The domain expert was a marketing employee involved in selling that product while the data expert was an I.T. employee used to providing database support to the marketing department. The Data Mining experts (the authors in this case) would normally belong to a consultancy organisation deployed by the bank for the purpose of solving the problem at hand.

The format of the rest of the paper traces the remaining stages of the Data Mining process. Section 2 discusses the *Problem Specification* stage, analysing the cross-sales problem in detail. Section 3 describes the *Data Prospecting*

* Corresponding author. Tel.: 44 1232 366671; fax: 44 1232 366068; e-mail: ss.anand@ulst.ac.uk

0950-7051/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII: S0950-7051(98)00035-5

stage, which involves analysing the available data, and choosing a promising subset of data to mine. Stage 4 of the process, *Domain Knowledge Elicitation*, is detailed in Section 4. This is where useful knowledge already known about the problem area is elicited from the domain expert for incorporation in the mining process. Section 5 describes the *Methodology Identification* stage, where the mining paradigms most appropriate for the problem are chosen, and in Section 6 the *Data Pre-processing* stage is described, where the data is transformed into a state that is fit for mining. The *Pattern Discovery* stage, at which the actual discovery of knowledge is carried out, is discussed in Section 7. Here parameter settings for the paradigms chosen are set to initial values and are refined in an iterative manner after analysis of the discovered knowledge. The final stage, *Knowledge Post-processing*, is described in Section 8 where the rules discovered in the previous stage are filtered to find the best rules. The Data Mining process is recursive in nature and continual refinement of the techniques used is carried out until useful knowledge is discovered. Refinement of domain knowledge is one aspect of the Refinement Process that precedes each iteration of the Data Mining process. Section 9 describes a novel refinement carried out on the domain knowledge utilising the knowledge discovered from the previous Data Mining iteration.

2. Problem Specification

Problem Specification is the second stage of the Data Mining process. During this stage the domain, data and Data Mining experts identified in Stage 1 of the process analyse the problem at hand. The problem must be disassembled into smaller tasks, so that those tasks that require a Data Mining solution can be focused upon. We refer to these as Data Mining tasks (DMT). Each DMT is classified into one of the various Data Mining goals (DMG) identified in literature, for example, discovery of associations [4], classification rule discovery [7], sequence rule discovery [8] and characteristic and discriminant rule discovery [1]. As each DMG requires a different Data Mining paradigm to be used, this classification of DMTs into DMGs is crucial. We now analyse the cross-sales problem with a view to arriving at a specification for it.

In the present competitive environment within the service industries, retaining customers and maximising profit from an existing customer base have become at least as important as attracting new customers. Most companies are involved in providing more than one service or product to their customers. Since they have access to information stored about their existing customers through data collected previously, they can make a more informed choice on which customers to target with products that they do not already have. The advantages of such a strategy are threefold. First, targeting customers with products they are more likely to buy should increase sales and therefore increase profits. Second, reducing the amount of people targeted through more selective targeting should reduce costs. Finally, it is an established fact in the financial sector that loyal customers (from the perspective of their length of relationship with the organisation) normally possess more than two products on average; therefore, persuading customers to buy more than one product should increase customer loyalty.

Cross-sales is the term given to the process of a company targeting a product at existing customers who are not already customers of that particular product (see Fig. 2). The company's customer base can be subdivided into a

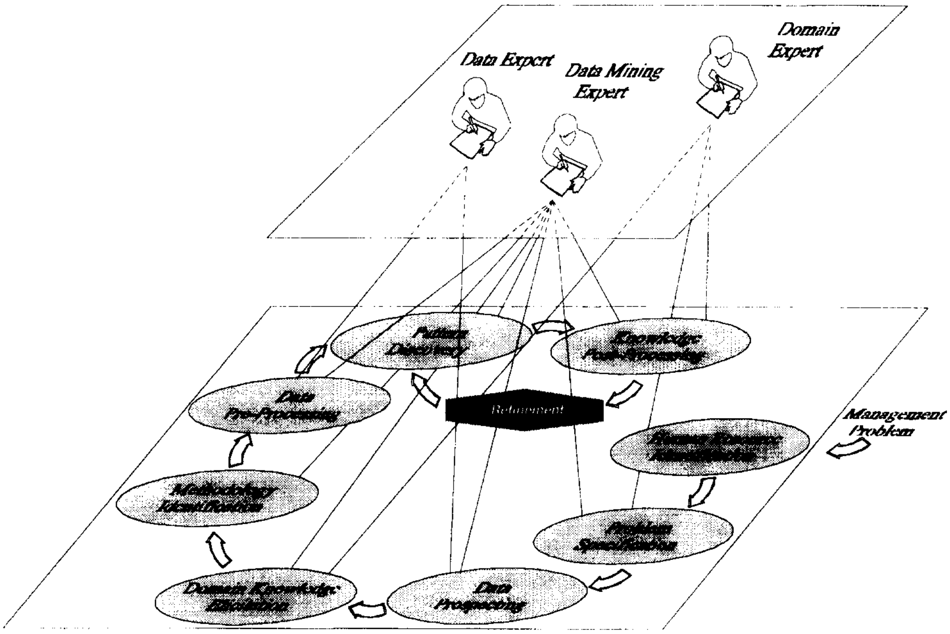

Fig. 1. The Data Mining Process.

number of customer groups based on the products they have already purchased. Cross-sales is the targeting of a product bought by one customer group on the subset of the customer base represented by the customer groups of the other products.

There are four component tasks that can be identified for the cross-sales problem. They are:

1. Finding sets of attributes that identify customers in the customer base that are most likely to buy a particular product (in this case Household Insurance);
2. Choosing the best of these sets of attributes, to identify customers to target in a marketing campaign of some sort (for example, a mail shot);
3. Carring out this marketing campaign and analysing the results to see if a high "hit rate" was achieved; and
4. Feeding back results into the customer database, to carry out refinement of the rules used for targeting customers with the product.

Of these four tasks, tasks one and two can be identified as DMTs where we are trying to discover the best sets of attributes within the banks database that identify customers of Household Insurance. The next stage is to identify the DMG of each of these tasks.

Consider the following scenario. A bank wants to promote its Household Insurance product by targeting existing customers who have not already bought the product. At present most marketing departments are groping in the dark, targeting some of their customers based on certain hunches held by their marketing experts or sometimes in a totally random manner. Within their customer database, the bank has data on two types of customers:

1. Type 1: Those that have Household Insurance; and
2. Type 2: Those that do not have Household Insurance.

At first glance this seems like a simple classification problem but on closer examination it is obvious that the problem at hand is actually not a classification problem at all. For classification rule discovery data on three types of customer are required:

1. Type 1: Those that have Household Insurance;
2. Type 2: Those that have refused to purchase Household Insurance; and
3. Type 3: Those that do not have Household Insurance but have not refused it.

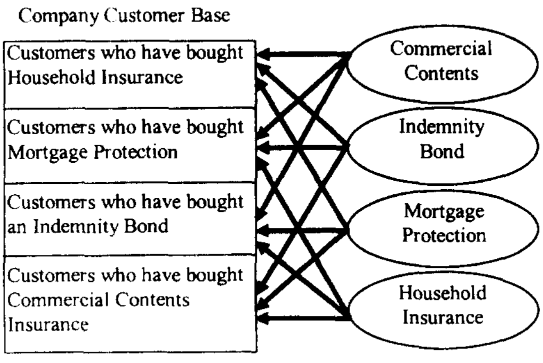
Company Customer Base
Customers who have bought Household Insurance
Customers who have bought Mortgage Protection
Customers who have bought an Indemnity Bond
Customers who have bought Commercial Contents Insurance
Commercial Contents
Indemnity Bond
Mortgage Protection
Household Insurance

Fig. 2. The Cross-Sales Problem.

The first type of customer forms the positive example set, the second type of customer forms the negative example set and the third type forms the target data set.

In cross-sales we only have a positive example set and a target data set. For this reason, the first DMT has a characteristic rule discovery DMG rather than a classification rule discovery DMG. Characteristics are attribute value pairs that are prevalent in a particular group of records, which in our case is the group of records that pertain to customers who have Household Insurance. Given these characteristics, customers in the target data set with similar characteristics can be targeted in the sales campaign. Fig. 3 gives an overview of the process of using characteristic rules to identify prospective customers.

The following is an example of a characteristic rule. The rule indicates that a large proportion (12.79%) of Household Insurance customers (the positive example set) have a skilled occupation, an Hon-Commits status and that the net credit turnover on their account is less that £4000 a year. The main problem with such a rule is that it only describes prevalent characteristics of Household Insurance customers. Potentially, these characteristics could be equally as prevalent in the negative example set; however, the second DMT described later in the section deals with such a scenario.

if Household Insurance = Y,
then occupation = SKILLED and status = Hon-Commits and net credit turnover > 4000
with support = 12.79% and interest 0.74.

The corresponding rule discovered by a classification rule algorithm, would be:

if occupation = SKILLED and status = Hon-Commits and net credit turnover > 4000
then Household Insurance = Y
with support = 12.79%.

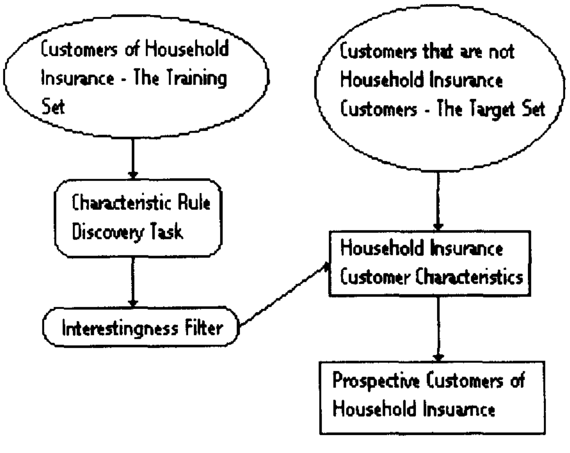
Customers of Household Insurance - The Training Set
Customers that are not Household Insurance Customers - The Target Set
Characteristic Rule Discovery Task
Household Insurance Customer Characteristics
Interestingness Filter
Prospective Customers of Household Insuarnce

Fig. 3. Cross-Sales using Data Mining.

Table 1
Customer table

Customer *n*	Tenancy	Occupation	Status	Date of birth	Sex	Marital status	Dependants	Access
10001	Owner	Professional	Undoubted character	19–10–70	M	M	0	Yes
10002	Tenant	Chef	New	21–1–63	F	M	11	Yes
10002	Owner	Skilled	New	2–4–22	M	S	4	No
10002	With parents	Student	New	14–5–68	F	M	7	No

Here, the antecedent characteristics discriminate between customers likely and those unlikely to buy Household Insurance. Therefore, these characteristics cannot be prevalent in the negative example set and the problem associated with the characteristic rule discovery approach does not arise.

Characteristic rules are not as accurate as classification rules, as classification rule discovery takes both the negative and positive examples into account and discovers rules that discriminate between these two types of examples. However, characteristic rules have the benefit of providing domain experts with an insight into how their business plans are working in the real world, possibly identifying areas within the customer base that seem to be responding less favourably to their present policies of targeting customers. Based on this discovered knowledge, the marketing experts may decide to intensify their efforts with respect to targeting these parts of the customer base.

However, there are two problems that can arise from using a characteristic rule discovery approach. First, the number of characteristic rules discovered using such an approach is exponential, producing an overwhelming number of rules, making manual sifting to discover the best rules difficult, and thus defeating the purpose of Data Mining. Second, in the absence of negative examples, the discovered characteristics may also be characteristics of the customer base in general, rather than of Household Insurance customers only, as they represent necessary but not sufficient conditions for the membership of the positive example set.

While the number of rules generated can be controlled using support and uncertainty thresholds and domain knowledge [9], the second DMT tackles the problem of selecting the most interesting characteristic rules of Household Insurance customers. One way of measuring which rules are most interesting is to use deviation detection [2]. This involves detecting deviations in the Household Insurance characteristics from characteristics of the overall customer base. Thus, the second DMT has a deviation detection DMG.

Once the best rules discovered are chosen and used to identify customers in a marketing campaign, the bank can keep a record of those customers that were targeted but did not buy the product. These records can be used to refine the characteristic rules making them more accurate or they may be used as negative examples—the characteristic rule discovery task being transformed into a classification rule discovery task.

3. Data Prospecting

Data Prospecting is the next stage in the Data Mining process. It consists of analysing the state of the data required for solving the problem at hand. There are four main considerations within this stage: What are the relevant attributes? Is the required data stored electronically and if so is it accessible? Are the data attributes required populated? Is the data distributed, heterogeneous, stored in legacy systems or is it centrally stored? If heterogeneous, are there any semantic inconsistencies that the data expert can account for or do they need to be "discovered" before the data can be used for discovering knowledge for decision support?

Two main sources of data for mining were available to the present study. These were the data held by the bank relating to their customers and externally procured survey data relating to the population of Northern Ireland in general, where the majority of the bank's customers reside.

With respect to the data held by the bank, two sections of the customer data were identified as being relevant to the cross-sales problem. These were the personal information about the customer, for example, demographic information, sex, occupation and marital status stored in the customer table (Table 1) and transactional information on the different accounts held by the customers stored in the relevant account transaction table (Table 2). An important aspect of the data was identified at this stage. While data on customers who bought Household Insurance were available from the bank's databases, no information was available on customers that did not require a Household Insurance product at all, or had not taken up the Household Insurance product with the bank but had Household Insurance with a competitor organisation. This fact confirms that the classification of the first DMT identified in the previous section has a characteristic rule discovery goal as opposed to a classification goal.

Three externally available data sets were believed to be relevant to the cross-sales problem, the Robson's

Table 2
Account transaction table

Customer *n*	Account number	Date/time	Transaction amount (£)	Balance (£)
10001	263	14–5–97:09:20	−30	330.76
10001	264	14–5–97:14:30	+168	1687.97
10001	263	15–5–97:17:45	−45	285.76
10002	265	14–5–97:13:24	−100	243.16

Deprivation Index [10] data, the Acorn classification data and lifestyles survey data. Each of these data sets provided information based on geographical location, which was not available from the banks internal databases. Robson's Index utilises 18 different indices of deprivation to produce an overall deprivation index at the census enumeration district (ED) level of granularity. The indices used include number of children in unsuitable accommodation, long-term illness, households on income support, 18–24 year-olds with no qualification, properties without public sewage, unemployment and standardised mortality rate. The Acorn classification, also at the ED level of granularity, classifies Northern Ireland geographically into 60 distinct areas. This is based on factors such as the age of people in the area, the accommodation they live in, how they are employed and even how they travel to work. The lifestyle data, at the postcode level of granularity, include the mean and median level of income, the ranking of the area in relation to the rest of the UK and the level of home ownership within the area. Fig. 4 shows the enumeration districts defined in Northern Ireland, along with a geographical map of the bank's customers. The black areas on the map depict the geographical location of customers of the bank while Household Insurance customers are shown in white.

A number of potential problems were identified in the available data. First, the data expert was aware that some attributes, e.g. the income attribute in the customer table, did not contain up-to-date values. Also some attributes had a default value stored that was also a valid value of the attribute For example, the number of dependants had a default value of zero, while zero is clearly a valid value as well. It was observed that over 60% of the Household Insurance customers had number of dependants stored as zero. The varying granularity of the available data—postcodes versus ED—was also a problem with the data considered to be useful within the cross-sales exercise. Problems such as these are dealt with at the Data Pre-processing stage, and therefore a discussion on how these problems were tackled is delayed until Section 6.

4. Domain Knowledge Elicitation

The next stage is that of Domain Knowledge Elicitation. During this stage of the Data Mining process, having identified the data that are relevant to the problem being tackled, the Data Mining expert attempts to elicit any domain knowledge that the domain expert may be interested in incorporating into the mining process.

Domain Knowledge is useful in a number of ways. It can be used for making patterns more visible, for constraining the search space, for discovering more accurate knowledge and for filtering out uninteresting knowledge. Domain Knowledge may be broadly classified into three types: hierarchical generalisation trees (HG-trees), attribute relationship rules (AR-rules) and Environment-based constraints [9]. HG-Trees are the most commonly used domain knowledge that provides various levels of generalisation for attribute values within the discovery view, and are useful in making patterns more visible. However, as

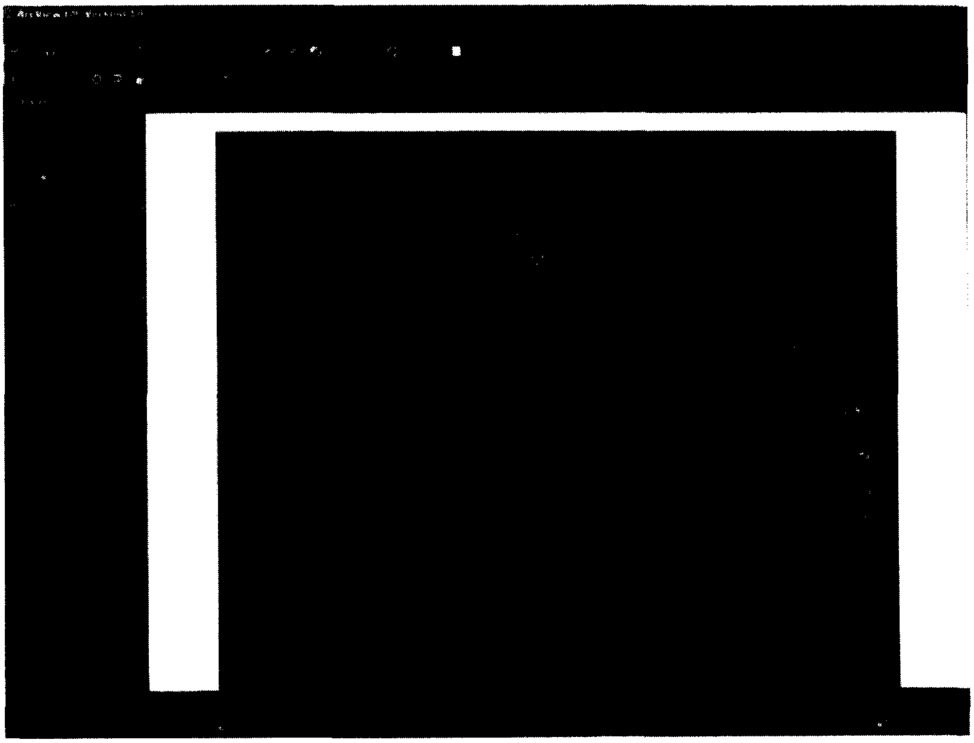

Fig. 4. A Geographical plot of the customer base.

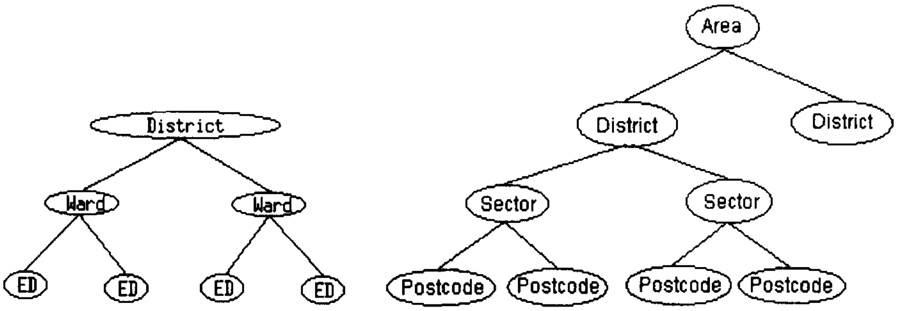

Fig. 5. Hierarchical Generalisations for geographical attributes.

shown in Section 9, the use of such generalisation trees can often be detrimental to the value of the discovered knowledge, due to overgeneralisation. AR-rules are rules that are known to the domain expert about existing dependencies within the data being mined. These rules can be used to filter out uninteresting rules and deal with incompleteness in the data [9]. Environment-based constraints include support and uncertainty constraints, syntactic constraints and inter-attribute dependency constraints (see Section 7). This type of domain knowledge includes information that is domain and goal dependent and is not available from any source other than the domain expert. Table 3 shows the initial domain knowledge, mainly of the HG-tree type, provided by the domain expert. Environment-based constraints used are described in Section 7.

Two common hierarchical generalisations defined on geographical location attributes are shown in Fig. 5. However, the externally procured data available at the ED level of coarseness may be interpreted as additional HG-tree type domain knowledge defined in the EDs, as opposed to ordinary attributes within the discovery view. The externally procured data sets define new hierarchical generalisations on EDs based on each of the attributes within these data sets. An example generalisation is given in Fig. 6.

5. Methodology Identification

The main task of the Methodology Identification stage is to find the best Data Mining methodology to solve the specified mining problem. Often a combination of mining paradigms is required to solve the problem at hand. For example, clustering or data partitioning may be required before the application of a classification algorithm. The chosen paradigm depends on the type of information required, the state of the available data, the domain of knowledge being discovered and the problem at hand. The problem at hand in this case is to carry out the DMTs and their corresponding DMGs identified at the Problem Specification stage of the process.

The first DMT of the cross-sales problem was identified as having a characteristic rule discovery goal. Such rules may be discovered by using an association rule discovery algorithm, incorporating the syntactic constraint, "the Household Insurance indicator attribute is the only antecedent attribute of interest". The Mining Kernel System [11], a prototype Data Mining system developed by the authors, includes the evidence-based association rule (EAR) discovery algorithm [12]. The EAR algorithm is a generalisation of earlier Association algorithms. EAR

Table 3
Initial domain knowledge

Attribute name	Generalised values
Postcode	URBAN|POST_RURAL|POST_OTHERS
Date of birth	Transformed to the following age brackets: 22_24|25_40|41_55|56_95
No. dependants	0|1_2|3_5|6_GTR
Status	UNDOUBTED_CHAR|SAT_CREDIT_LIMIT|HON_COMMITS|NEW|OTHER
Marital status	MARRIED|OTHER
Access	CARD_HELD|OTHER
Net average balance (yearly)	41 equal width intervals of £1000
Net credit turnover (yearly)	48 equal width intervals of £1000
Account [1..3] average balance (yearly)	28 equal width intervals of £1000 for account 1
	27 equal width intervals of £1000 for account 2
	17 equal width intervals of £1000 for account 3
Account [1..3] credit turnover (yearly)	14 equal width intervals of £1000 for account 1
	13 equal width intervals of £1000 for account 2
	4 equal width intervals of £1000 for account 3
Account [1..3] type	CURRENT|LOAN|SAVINGS|PEP|TESSA
Occupation	SKILLED|PROFESSIONAL|ST_NON_EARNING|STUDENT_OTHER|UNEMPLOYED|APPRENTICES|ENT|CHEF|MACHINIST|MANUAL|WAITRESS|PAINTER|SECURITY_OTHERS|FARMER_GTR_50|NE|RE

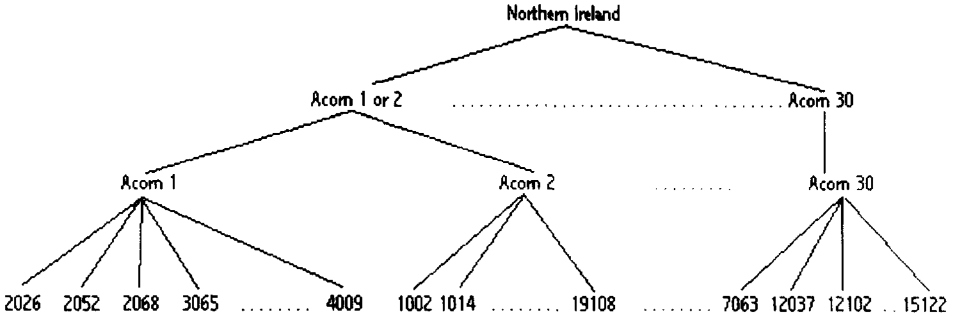

Fig. 6. An HG-Tree defined on Enumeration District based on the Acorn Classification attribute.

allows the incorporation of support and uncertainty thresholds and syntactic constraints [7]. In addition to the simple syntactic constraints of the type defined by Agrawal et al. [7], EAR allows the definition and incorporation of inter-attribute dependency (IAD) constraints. An example of such a constraint is "a rule that contains an expression pertaining to the account average balance of a customer is valid only if it also contains an expression regarding the account type". In addition, the EAR algorithm can discover knowledge from multivalued attributes rather than just binary attributes as in the case of previous algorithms [4]. It allows the incorporation of domain knowledge and can handle missing values in the data. The EAR algorithm requires attributes to be discrete. Therefore, the domain expert must provide interval bands for continuous variables and domain specific hierarchies for a number of other attributes. However, in Section 9 we discuss how such hierarchies can be induced from the data itself. The number of characteristic rules discovered by EAR can be controlled using a threshold of minimum support and uncertainty. The higher this threshold, the fewer is the number of rules discovered.

The second DMT identified was to select the "best" characteristic rules discovered for targeting prospective customers. Some measure of which rules are "best" is required and should be based on the uniqueness of the rule for Household Insurance customers. The interestingness measure used is normally dependent on the problem at hand [13,14]. In cross-sales, we clearly do not want to target customers with a product based on customer characteristics that are actually characteristics of the companies customers in general. Thus, the interestingness measure we use is based on the deviation from the "norm" of the characteristic rules discovered for the product being targeted. The "norm" in our case is the support for these characteristics within the complete customer base of the company, i.e. a characteristic rule is interesting if it is a characteristic of the customer of a product rather than the customer base in general. Thus, we define the interestingness measure for characteristics c, Interest$_c$, as:

$$\text{Interest}_c = \frac{S_p - S_o}{\max\{S_0, S_p\}}$$

where, S_p is the support for the characteristics, c, in the positive example data set, S_o is the support for the characteristics c in the complete customer base. The expression in the denominator, $\max\{S_0,S_p\}$, is called the *normalising factor* as it normalises the interest measure onto the scale $[-1, 1]$.

6. Data Pre-processing

The next stage is that of Data Pre-processing. Depending on the state of the data, this process may constitute the stage where most of the effort of the Data Mining process is concentrated. Data Pre-processing involves removing outliers in the data, predicting and filling in missing values, noise modelling, data dimensionality reduction, data quantisation, transformation and coding and heterogeneity resolution. Outliers and noise in the data can skew the learning process and result in less accurate knowledge being discovered. They must be dealt with before discovery is carried out. Missing values in the data must either be filled in or a methodology used that can take them into account during the discovery process so as to account for the incompleteness of the data model. Data dimensionality reduction is an important aid to improved efficiency of the discovery algorithm as most of the algorithms have execution times that increase exponentially with respect to the number of attributes within the data set.

As noted at the Data Prospecting stage, a number of problems exist within the data sets available for this cross-sales exercise. These problems are dealt with at the Data Pre-processing stage. Attributes with out-of-date values, such as the income attribute, were discarded from the discovery view. On the other hand, the "number of dependants" attribute contained default values of zero which is also a valid value for the attribute. This attribute was, however, retained within the data set as there was a similar proportion of the overall customer base with no dependants recorded. Therefore, that particular value of the attribute would be filtered out by the interest measure, while other attribute values could potentially prove to be interesting. Also, the attributes pertaining to the additional customer accounts, i.e. account type2, account2 average balance, account2 credit turnover, etc., had a low population, as few customers of the bank owned more than one account. However, as the missing values in these fields represented the non-existence of additional accounts, they were treated as a distinct attribute value rather than as a missing value. Data in the account transaction tables were summarised to account type, account average balance and account yearly turnover and joined with the customer table.

The inconsistency between postcode and ED-based data was removed by transforming the ED data to the correct level of granularity using a spatial join between the data at the postcode level of granularity and data at the ED level of granularity. This pre-processing of the data resulted in the initial discovery view shown in Table 4.

7. Pattern Discovery or Model Development

The Pattern Discovery stage follows the Data Preprocessing stage. It consists of using algorithms that automatically discover patterns from the pre-processed data. For the Household Insurance product provided by the bank, the positive subset of records from the overall customer base was extracted, i.e. those records corresponding to Household Insurance customers. This subset was mined to discover characteristic rules for the product. Next, the overall customer database was mined and deviation detection used to identify the best characteristic rules discovered. For this particular exercise, the data set used was a random sample of the overall customer base consisting of over 20,000 customer records with 430 Household Insurance customers.

Table 4
Initial discovery view

Attribute name
Customer number
Tenancy
Occupation
Status
Date of birth
Sex
Marital status
Dependants
Access
Net average balance (yearly)
Net credit turnover (yearly)
Account type1
Account1 average balance (yearly)
Account1 credit turnover (yearly)
Account type2
Account2 average balance (yearly)
Account2 credit turnover (yearly)
Account type3
Account3 average balance (yearly)
Account3 credit turnover (yearly)
District
Mean income
Robson Deprivation Index
Acorn classification
Retired (ratio of retired persons in the post code population, ranked 0–4)
Children (children per family ratio in the post code population, ranked 0–4)
Income distribution (spread of income in the post code population, ranked 0 = narrow, 1 = medium, 2 = wide)

The procedure followed was to initially set a high threshold value for support and then to reduce it iteratively and constrain the rules discovered to those attributes that seem interesting through knowledge discovered in previous iterations. Initial iterations tend to discover a number of obvious relationships that the domain expert is normally aware of. These rules can be used as AR-rules type domain knowledge in future iterations to filter out such obvious knowledge.

During the discovery of the characteristic rules from the Household Insurance customers data, three inter-attribute dependency constraints were defined:

1. IAD1: if account1 average balance or account1 credit turnover appears in a rule account type1 must appear in the rule as well;
2. IAD2: if account2 average balance or account2 credit turnover appears in a rule account type2 must appear in the rule as well;
3. IAD3: if account3 average balance or account3 credit turnover appears in a rule account type3 must appear in the rule as well.

The only syntactic constraint defined was:

> "the only antecedent attribute on interest is the Household Insurance indicator".

Example characteristic rules discovered are shown below:

1. if Household Insurance = *Y*
   then occupation = SKILLED
   with support = 26.51% and interest 0.53
2. if Household Insurance = *Y*
   then occupation = SKILLED and status = Hon-Commits
   with support = 21.86% and interest 0.68
3. if Household Insurance = *Y*
   then occupation = SKILLED and status = Hon-Commits and net credit turnover > 4000
   with support = 12.79% and interest 0.74
4. if Household Insurance = *Y*
   then occupation = SKILLED and status = Hon-Commits and net credit turnover > 4000 and account type1 = CURRENT
   with support = 12.56% and interest 0.74
5. if Household Insurance = *Y*
   then dependants = 0_Dep and net average balance = Zero_1500 and CHILDREN = 4
   with support = 10.93% and interest = − 0.77

The effect of the interest value threshold on the number of rules discovered is shown in Table 5. The rules were constrained to a maximum size of seven consequent attributes.

Table 5
Number of rules discovered vs the interest measure

Interest threshold	Number of rules
0.84	7
0.80	217
0.70	1178
0.50	1739
0.00	3737

Fig. 7 presents some of the rules graphically. Here, the oval nodes represent attributes used to specialise the rule specified by the path from the root node to the node preceding the oval node, while the rectangular nodes represent the specialisation attribute value. The numbers shown in the rectangular node are the support and interest of the rule. The light grey nodes represent rules that have an interest value less than or equal to another rule that the present rule is a specialisation of, while the darker grey nodes represent rules where the specialisation attribute has improved the interest value of the rule. Non-shaded nodes represent rules where the specialisation has decreased the interest in the rule.

8. Knowledge Post-processing

The last stage of the Data Mining process is Knowledge Post-processing. Trivial and obsolete information must be filtered out and discovered knowledge must be presented in a user readable way, using either visualisation techniques or natural language constructs. Often this knowledge filtering process is domain as well as user dependent. Another aspect of the Knowledge Post-processing stage is to validate the knowledge before using it for critical decision support.

Due to the fact that the data used as input to the Data Mining process are often dynamic and prone to updates, the discovered knowledge has to be maintained. Knowledge Maintenance may involve reapplying the already set up Data Mining process for the particular problem or using an incremental methodology that would update the knowledge as the data change, keeping them consistent.

In Section 2 we described why the cross-sales problem cannot be solved using a classification algorithm. We also suggested that after using a characteristic rule discovery approach, customers could be targeted in a sales campaign the results of which, including negative responses could be stored in the database. This new database could then be used to generate a classification tree. However, we also pointed out certain advantages of the characteristic approach i.e. the fact that these rules provide the expert with insights into the customer base that are not in general provided by the classification algorithm. In this section we discuss how the characteristic rules discovered may be refined to take into account the negative examples.

Refining the characteristic rules generated from the positive examples consists of two steps. First, discovery of characteristic rules for the negative examples. Second, determination of deviation in support for a set of characteristics in the negative examples from the support for the same set of characteristics in the positive examples. The first step may be carried out using the techniques used for discovering rules from the positive examples discussed in this paper.

To illustrate the second step, consider the following example. Suppose the following rule is discovered from the negative example set:

if Household Insurance = *N*
then marital status = M and occupation = Skilled and access = CARD_HELD

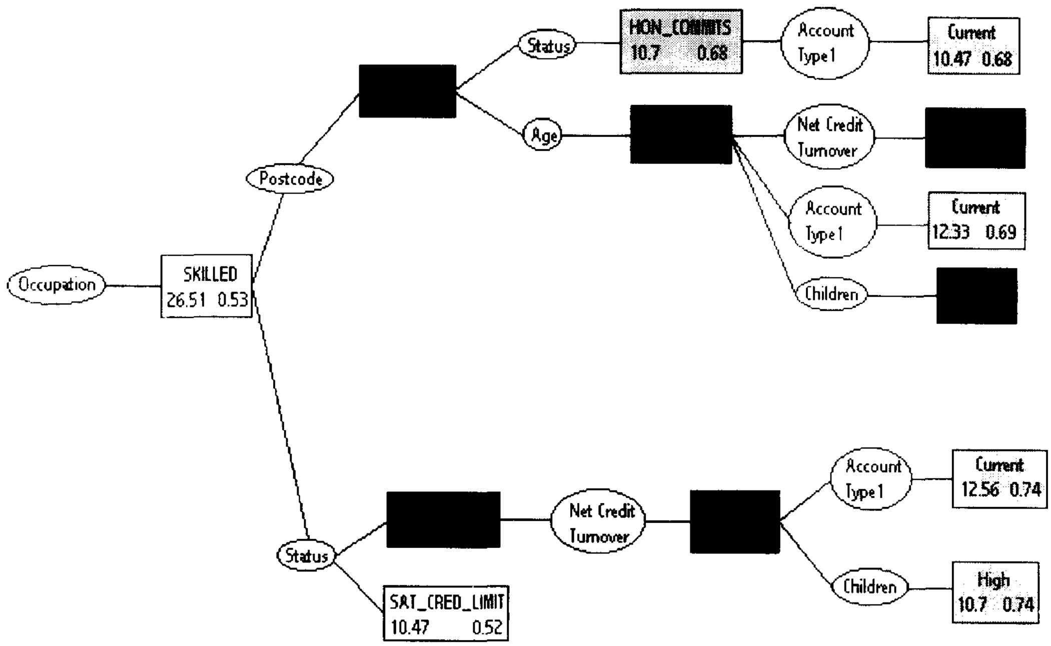

Fig. 7. Characteristic Rule Visualisation.

with support = 20.45% and interest = − 0.628
and the following rule, discovered from the positive example set, already exists:

if Household Insurance = *Y*
then marital status = M and occupation = Skilled and access = CARD_HELD
with support = 68.30% and interest = 0.194

The aim is to reinforce the original rule induced from the positive examples with the new knowledge gained about the same characteristics in the negative examples. The original interest value measured how uniquely the characteristics identified Household Insurance customers from general characteristics of customers. To reinforce the rule this value is replaced with a value that measures how uniquely the characteristics identify Household Insurance customers from customers who would not buy Household Insurance. The new value is calculated using the same calculation of interest as in Section 5. This time however the support for the rule in the overall customer base is replaced with the support for the rule in the negative examples:

$$\frac{68.3 - 20.45}{68.3} = 0.7$$

Therefore, the original rule would be refined to:

if Household Insurance = *Y*
then marital status = M and occupation = Skilled and access = CARD_HELD
with support = 68.30% and interest = 0.7

Note that, if the rule discovered from the negative examples had a support value of 78.68%, the interest value would be 0.266 and the refined rule would have an interest measure of − 0.132. Thus a positive interest measure for rules from the negative example set reduces the interest in the corresponding rule from the positive examples as it signifies the characteristics to be those people who have refused the product being targeted. Similarly, a negative interest value implies that the characteristics are in fact that of customers of the product; thus, the interest in them increases.

9. Refinement Process

It is an accepted fact that Data Mining is iterative. After the Knowledge Post-processing stage, the knowledge discovered is examined by the domain expert and the Data Mining expert. This examination of the knowledge may lead to the Refinement Process of Data Mining. During the Refinement Process the domain knowledge as well as the actual goal of the discovery may be refined. Refinement of the goal of discovery could take the form of redefining the data used in the discovery or a change in the methodology used or the user defining additional constraints on the mining algorithm, or refinement of the parameters of the mining algorithm. Once the refinement is complete the Pattern Discovery and Knowledge Post-processing stages are repeated. Note that the Refinement Process is not a stage of the Data Mining process. Instead it embodies the iterative aspects of the Data Mining process and may make use of the initial stages of the process i.e. Data Prospecting, Methodology Identification, Domain Knowledge Elicitation and Data Pre-processing.

In this section we concentrate on the domain knowledge refinement aspects of the Refinement Process. We discuss the domain knowledge refinement under three separate headings based on the underlying attribute data type, as the Refinement Process is dependent on the attribute type. The iterative aspects of the Pattern Discovery stage, i.e. parameter refinement and constraint refinement, have already been discussed in Section 7.

9.1. Discrete unordered attributes

Although domain knowledge in the form of HG-trees can make patterns in data more visible, they can also reduce the apparent usefulness of the rules discovered. Consider the domain knowledge represented in the HG-tree in Fig. 8. Now, if the Household Insurance customers are distributed unevenly across districts in a generalised geographical area, the positive interest in some districts within the generalised area will be countered by negative interest in other districts within the same generalised area, and thus the rule will appear to be uninteresting. However, if the HG-tree is modified based on information about the Household Insurance customers, a discovery in its own right, the geographical area data could yield interesting rules.

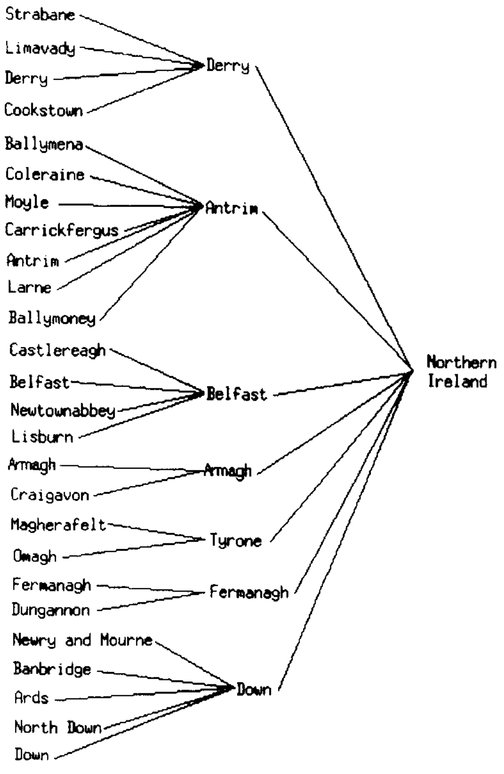

Fig. 8. Initial HG-Tree defined on DISTRICT.

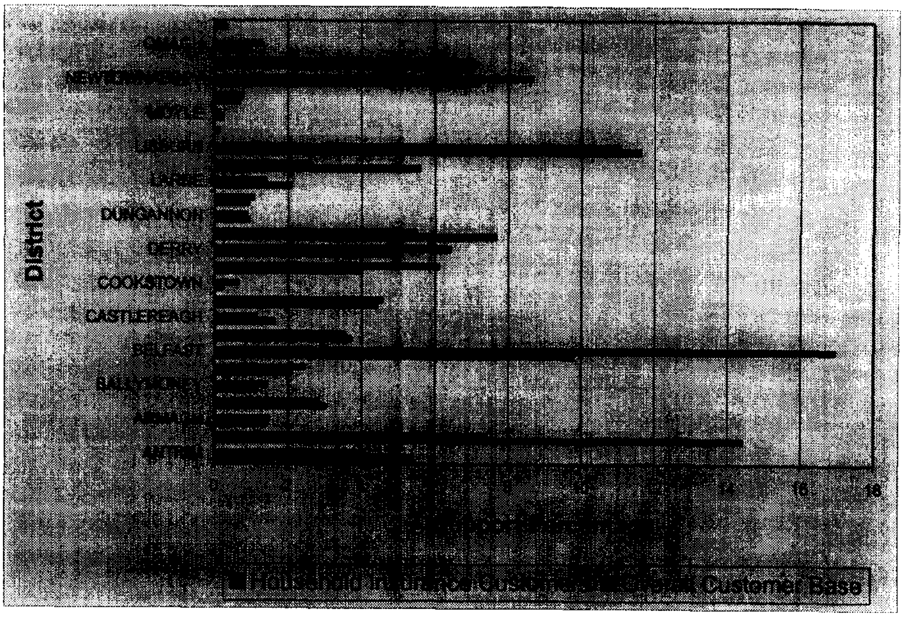

Fig. 9. Distribution of customers across districts.

Consider the distribution of Household Insurance customers shown in Fig. 9. Clearly, the HG-tree in Fig. 10 would provide more useful and interesting knowledge as opposed to the original HG-tree in Fig. 8. In Fig. 10, using the geographical distribution of Household Insurance customers, each of the generalised values in the original HG-tree is now split into two subareas, one with positive interest and another with negative interest. The domain expert may interactively modify the domain knowledge further based on additional information not available from the database. In the application discussed in this paper, the domain expert further refined the hierarchy into eight new regions defined below: N_ANTRIM consisting of {Antrim, Ballymena, Carrickfergus, Larne, Moyle}; S_ANTRIM_ARDS consisting of {Ards, Castlereagh, Down, Lisburn}; DUNGANNON consisting of {Dunganon}; LIMAVADY consisting of {Limavady}; S_DOWN_ARM consisting of {Armagh, Banbridge, Craigavon, Newry and Mourne}, BEL_N_DOWN consisting of {Belfast, Newtownabbey, North Down}; S_WEST consisting of {Cookstown, Derry, Fermanagh, Omagh, Strabane} and N_CENTRAL consisting of {Ballymoney, Coleraine, Magherafelt}.

9.2. *Discrete ordered attributes*

In the case of a discrete ordered attribute, the domain knowledge need not be provided initially by the domain expert. Hierarchical generalisation in the form of bandings for the attribute can be generated directly by the algorithm using the distribution of the Household Insurance customer information as shown in Fig. 11. For the Acorn classification attribute, the following bandings where generated: [1,2], 3, [4,15], 16, [17,22], 22, [23,26], 27, [28,29], 30, 31, 32, 33, [34,36], [37,40], 41, [42,54], [55,56], 57, 58, [59,60]. Bandings were chosen on consecutive Acorn classes such that the classes in each band had the same interest type i.e. positive or negative interest. It is possible that further generalisation of these values may be required so as to increase the support for the individual generalised values so that they are greater than the support threshold set

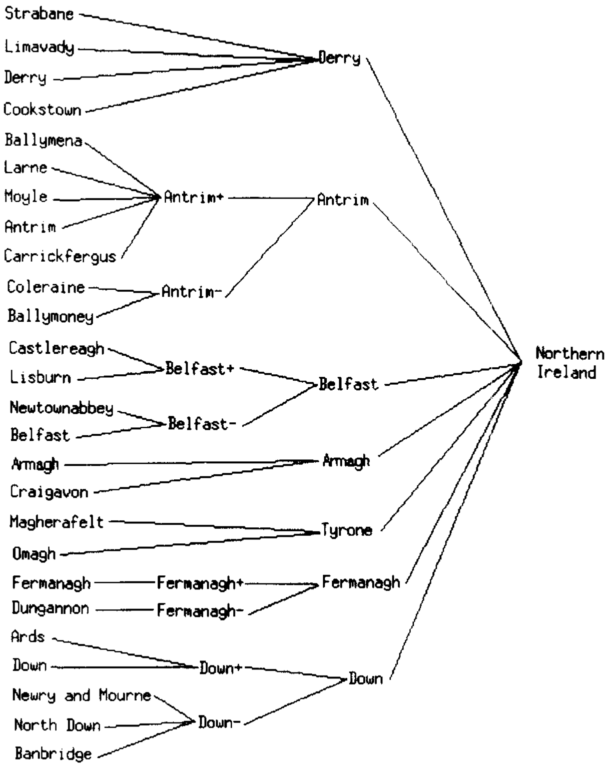

Fig. 10. Refined HG-Tree.

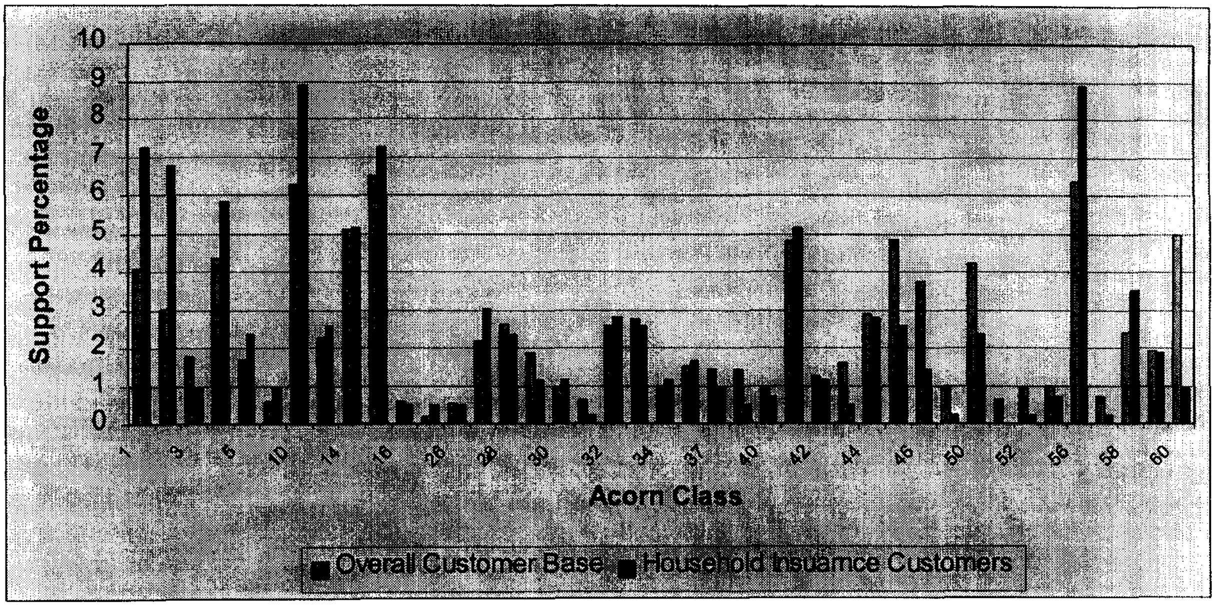

Fig. 11. Distribution of customers across Acorn classes.

by the Data Mining expert during the Pattern Discovery stage.

9.3. *Continuous attributes*

Finally, we discuss the possibility of employing a similar technique for domain knowledge refinement for continuous attributes. Once again no intervals are required from the domain expert though if the expert desires, the intervals provided may be taken into account and refined in an interactive fashion as in the case of discrete ordered variables. In Fig. 12 we present an example of the distribution of the Household Insurance customers as well as the overall customer base with respect to the Robson Deprivation Index. Clearly, the points where the distribution curves for the overall customer base and positive example set intersect are the optimal points for splitting the domain of Robson Index as it is at these points that the interest value changes sign. Mathematically, the interest measure for each interval, $[a, b]$, can be calculated for each area as $\int_a^b[f(x) - F(x)]$ where $f(x)$ is the distribution function for the positive example set, $F(x)$ is the distribution function of the overall customer base. As can be seen from Fig. 12, there are a number of such cross-over points, resulting in very fine intervals being created. Once again the initial intervals produced will need to be merged to produce further generalised values that would have support values greater than the user defined support threshold.

Table 6 shows the refined domain knowledge for some of the attributes used to discover the rules presented in Section 7.

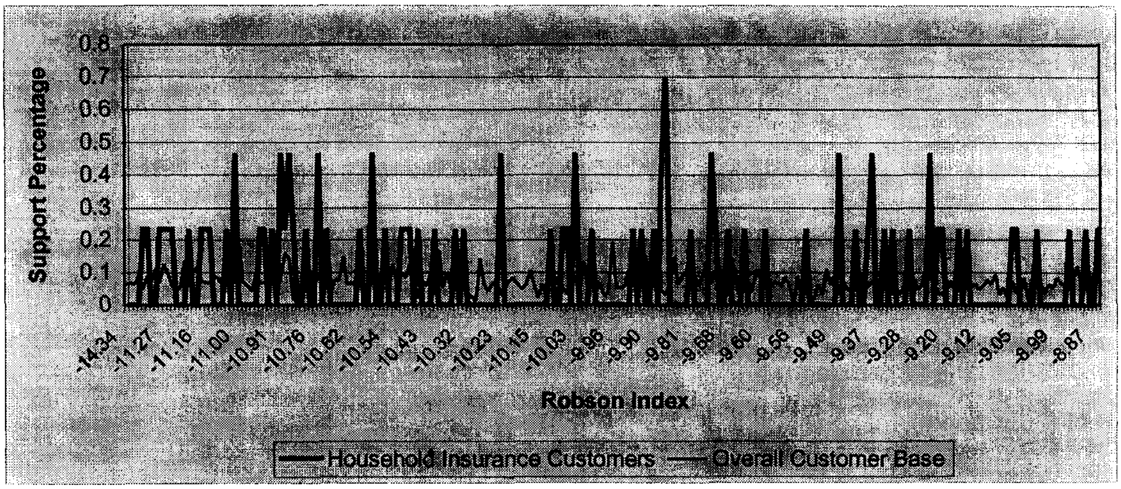

Fig. 12. Distribution of customers across Robson's Index.

Table 6
Refined domain knowledge

<table>
<tr><th>Attribute name</th><th>Generalised values</th></tr>
<tr><td>Net average balance (yearly)</td><td>[0.0, 1500], [1500, 20000000]</td></tr>
<tr><td>Net credit turnover (yearly)</td><td>[0, 5000], [5000, 20000000]</td></tr>
<tr><td>Account [1..3] average balance (yearly)</td><td>[0, 1000]</td></tr>
<tr><td></td><td>[1000, 20000000]</td></tr>
<tr><td>Account [1..3] credit turnover (yearly)</td><td>[0, 4500]</td></tr>
<tr><td></td><td>[4500, 20000000]</td></tr>
<tr><td>Account [1..3] type</td><td>CURRENT|LOAN|SAVINGS|PEP|TESSA</td></tr>
<tr><td>Occupation</td><td>SKILLED|PROFESSIONAL|ST_NON_EARNING|STUDENT_OTHER|UNEMPLOYED|APPRENTICES|ENT|CHEF|MACHINIST|MANUAL|WAITRESS|PAINTER|SECURITY_OTHERS|FARMER_GTR_50|NE|RE</td></tr>
<tr><td>District</td><td>N_ANTRIM|S_ANTRIM_ARDS|DUNGANNON|LIMAVADY|S_DOWN_ARM|BEL_N_DOWN|S_WEST|N_CENTRAL</td></tr>
<tr><td>Acorn classification</td><td>[1,2], 3, [4,15], 16, [17,22], 22, [23,26], 27, [28,29], 30, 31, 32, 33, [34,36], [37,40], 41, [42,54], [55,56], 57, 58, [59,60]</td></tr>
</table>

10. Conclusions

In this paper we have discussed the cross-sales problem and analysed this problem from the view to solving it using Data Mining techniques. The Data Mining solution is based on the discovery of characteristic rules, generalised to allow uncertainty in the rules, from the set of positive examples (customers of the product being targeted). We have highlighted the need for filtering the discovered rules based on some interest measure, and have presented a novel technique based on deviation detection for this purpose. We have also discussed the negative effect that domain knowledge can have on the discovery of interesting rules and introduced a method for the refinement of the domain knowledge using the interestingness filter. We have shown how the characteristic rules might be refined to more accurate rules if negative examples are made available through subsequent data collections. We have also investigated the use of externally procured data for data enrichment and as additional domain knowledge.

The approach discussed in this paper has been applied to a real world Data Mining application in the financial sector and results of its application are also provided. The methodology developed for providing a Data Mining solution to the cross-sales problem resulted in useful knowledge being discovered from the customer database, and it has great potential for exploitation within other service sectors.

Acknowledgements

The authors would like to thank ICL (Ireland) for making available to them the externally procured data sets used in the work discussed in this paper. In particular, we would like to thank Ken Morris, Graham Wilson and Martin Maken for their assistance. We would also like to thank the data and domain experts involved in the exercise described in this paper. The IRTU Technology Development Programme supports NIKEL.

References

[1] Y. Cai, N. Cerecone, J. Han, Attribute-oriented Induction in Relational Databases, Knowledge Discovery in Databases, AAAI/MIT Press, California, 1991, ed. G. Piatetsky Shapiro, W.J. Frawley.

[2] G. Piatetsky-Shapiro, C.J. Matheus, The Interestingness of Deviations, Workshop on Knowledge Discovery in Databases, AAAI Press, California, 1994.

[3] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, P. Uthurusamy, Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, California, 1996.

[4] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th VLDB Conference, Chile, Morgan Kaufman Publ. NC, California, USA, pp. 487–499, ed. J. Bocca, M. Jarke, C. Zaniolu.

[5] U. Fayyad, N. Weir, S. Djorgovski, Automated Analysis of a Large-Scale Sky Survey: The SKICAT System, Workshop on Knowledge Discovery in Databases, AAAI Press, California, 1993

[6] S.S. Anand, A.G. Buchner, Decision Support Using Data Mining, Financial Times Management, London, April 1998.

[7] R. Agrawal, T. Imielinski, A. Swami, Database mining: a performance perspective, IEEE Transactions on Knowledge and Data Engineering (1993) 5(6) pp. 914–925.

[8] R. Agrawal, R. Srikant, Mining Sequential Patterns, IBM Research Report, RJ9910 IBM, 1995.

[9] S.S. Anand, D.A. Bell, J.G. Hughes, The role of domain knowledge in Data Mining, in: Proceedings of the 4th International Conference on Information and Knowledge Management (CIKM'95) ed. N. Pissinou, A. Silberschatz, E.K. Pack, K. Makki, ACM Press, New York, 1995, pp. 37–43.

[10] B. Robson, M. Bradford, I. Deas, Relative Depravation in Northern Ireland, A Government Statistical Publication, UK, 1994.

[11] S.S. Anand, B.W. Scotney, M.G. Tan, S.I. McClean, D.A. Bell, J.G. Hughes, I.C. Magill, Designing a kernel for Data Mining, IEEE Expert Intelligent Systems and Their Applications 12 (2) (1997) 65–74.

[12] S.S. Anand, D.A. Bell, J.G. Hughes, Evidence-based discovery of association rules, Internal Report, Faculty of Informatics, University of Ulster, Newtonabbey, 1996.

[13] M. Kamber, R. Shinghal, Evaluating the interestingness of characteristic rules, in: 2nd International Conference on Data Mining and Knowledge Discovery, ed. E. Simoul, J. Han, O. Fayya, AAAI Press, California, 1996, pp. 263–266.

[14] C.J. Matheus, G. Piatetsky-Shapiro, D. McNeill, An Application of KEFIR to the Analysis of Healthcare Information, Workshop on Knowledge Discovery in Databases, AAAI Press, California, pp. 441–452, 1994.