Emerging Pattern Based Classification in Relational Data Mining

Michelangelo Ceci, Annalisa Appice, and Donato Malerba

Dipartimento di Informatica, Università degli Studi di Bari
via Orabona, 4 - 70126 Bari - Italy
{ceci,appice,malerba}@di.uniba.it

Abstract. The usage of descriptive data mining methods for predictive purposes is a recent trend in data mining research. It is well motivated by the understandability of learned models, the limitation of the so-called "horizon effect" and by the fact that it is a multi-task solution. In particular, associative classification, whose main idea is to exploit association rules discovery approaches in classification, gathered a lot of attention in recent years. A similar idea is represented by the use of emerging patterns discovery for classification purposes. Emerging Patterns are classes of regularities whose support significantly changes from one class to another and the main idea is to exploit class characterization provided by discovered emerging patterns for class labeling. In this paper we propose and compare two distinct emerging patterns based classification approaches that work in the relational setting. Experiments empirically prove the effectiveness of both approaches and confirm the advantage with respect to associative classification.

1 Introduction

Discovering a characterization of classes has been a challenge for research in machine learning and data mining. Emerging patterns (EPs) discovery is a descriptive data mining task which aims at characterizing classes by detecting significant differences between objects of distinct classes. EPs are introduced in [6] as a particular kind of patterns (or multi-variate features) whose support significantly changes from one data class to another: the larger the difference of pattern support, the more interesting the pattern. Change in pattern support is estimated in terms of support ratio (or *growth rate*). EPs with sharp change in support (high growth rate) can be used to characterize object classes.

Originally, EPs discovery has been investigated for capturing difference between separate classes of objects which are stored in a single relational table. Each tuple in the table represents an attribute-value vector describing an object of one of the classes and each EP is discovered in form of a conjunction of attribute values. Anyway, real-world data typically involve complex and heterogeneous objects with different properties which are modeled by as many relational tables as the number of object types. Mining data scattered over the multiple tables of a relational database (*relational data*) is a challenge in several domains,

e.g., spatial domains, process mining and, in general, in domains where the effect of an attribute is not limited to a specific unit of analysis. In this case, when properties of some units (reference objects) are investigated, attributes of related units (task-relevant objects) must be taken into account as well. Traditional EPs discovery methods do not distinguish task-relevant from reference objects, nor do they allow the representation of any kind of interaction. Recently, EP discovey has been framed in the context of relational data mining [8] in order to deal with both relational data and relational patterns [2].

Although (relational) emerging patterns discovery is specially designed for a descriptive task, in this work, we exploit potentialities of EPs to deal with the (relational) classification task. The rationale behind this undertaking can be found in the goal of classification, that is, to learn the concept associated to each class by finding regularities which characterize the class in question and discriminate it from the other classes. The idea is to learn a relational classifier by exploiting the knowledge implicitly encoded in the (relational) EPs that permits to discriminate between different classes sharply.

In this paper, we propose and compare two EPs-based relational classifiers, namely Mr-CAEP and Mr-PEPC. The former upgrades the EP-based classifier CAEP [7] from the single relational table setting (propositional setting) to the relational setting. It computes a membership score of an object to each class. The score is computed by aggregating a growth rate based function of the relational EPs covered by the object to be classified. The largest score determines the object's class. The latter (Mr-PEPC) maximizes the posterior probability that an object belongs to a given class. Probability is estimated by using a set of relational EPs to describe an object to be classified and, then, to define a suitable decomposition of the posterior probability *à la* naive Bayesian classifier to simplify the probability estimation problem.

The paper is organized as follows. In the next section related works are discussed and the present work is motivated. The problem of relational emerging patterns discovery is faced in Section 3. The two relational EP-based classifiers are presented in Section 4. Lastly, experimental results are reported in Section 5 and then some conclusions are drawn.

2   Related Works and Motivations

Data mining research has provided several solutions for the task of emerging patterns discovery. In the seminal work by Dong and Li [6], a border-based approach is adopted to discover the EPs discriminating between separate classes. Borders are used to represent both candidates and subsets of EPs; the border differential operation is then used to discover the EPs. Zhang et al. [20] have described an efficient method, called ConsEPMiner, which adopts a level-wise generate-and-test approach to discover EPs which satisfy several constraints (e.g., growth-rate improvement). Recently, Fan and Ramamohanarao [9] have proposed a method which improves the efficiency of EPs discovery by adopting a CP-tree data structure to register the counts of both the positive and negative class. All these

methods assume that data to be mined are stored in a single data table. An attempt of upgrading the emerging pattern discovery to deal with relational data has been reported in [2], where the authors proposed to adapt the *levelwise* method described in [15] to the case of relational emerging patterns.

In the last years, studies on emerging patterns discovery have generally resulted in methods for building classifiers through a careful selection of high quality emerging patterns discovered in training data [7,10,11,13]. Indeed, the idea of using descriptive data mining methods for predictive purposes has its roots in studies on *associative classification*, where the goal is to induce a classification model on the basis of discovered association rules.

Differently from emerging patterns based classification, associative classification has been studied not only in the propositional setting [14,19], but also in the relational one [4]. Interesting aspects of associative classification can be summarized in the following points:

1. Differently from most of tree-based classifiers, whose univariate tests cause the "horizon effect", associative classifiers are based on association rules that consider the simultaneous correspondence of values of different attributes, hence promising to achieve better accuracy [3].
2. Associative classification makes association rule mining techniques applicable to classification tasks.
3. The user can decide to mine both association rules and a classification model in the same data mining process [14].
4. The associative classification approach helps to solve the *understandability* problem [16] that may occur with some classification methods. Indeed, many rules produced by standard classification systems are difficult to understand because these systems often use only domain independent biases and heuristics, which may not fulfill user's expectation. With the associative classification approach, the problem of finding understandable rules is reduced to a post-processing task [14].

All these points are also valid in the case of emerging patterns based classifiers. In addition, we argue that emerging pattern based classifiers take advantages from the fact that EPs provide features which better discriminate classes than association rules do.

3   Relational EPs Discovery

The problem of discovering relational emerging patterns can be formalized as follows:

Given:

- a database schema S with h relational tables $S = \{T_0, T_1, \ldots, T_{h-1}\}$
- a set PK of primary key constrains on tables in S
- a set FK of foreign key constrains on tables in S
- a target relation $T \in S$ [1]

---

[1] Objects in T play the role of reference objects, while objects in $S - \{T\}$ play the role of task relevant objects.

  − a target discrete attribute $y \in T$, different from the primary key of T, whose
    domain is $C = C_1, C_2, \ldots, C_r$
  − a couple of thresholds $0 < minSup \leq 1$ and $minGR \geq 1$

The problem is to find a set of relational emerging patterns.

Formally, a relational emerging pattern is a conjunction of predicates of two
different types:

Definition 1 (Structural predicate). *A binary predicate $p/2$ is a structural*
predicate[2] associated to a table $T_i \in S$ if a foreign key in S exists that connects
T_i and a table $T_j \in S$. The first argument of p represents the primary key of T_j
and the second argument represents the primary key of T_i.

Definition 2 (Property predicate). *A binary predicate $p/2$ is a property*
predicate associated to a table $T_i \in S$ if the first argument of p represents the
primary key of T_i and the second argument represents another attribute in T_i
which is neither the primary key of T_i nor a foreign key.

Definition 3 (Relational Emerging Pattern).
A Relational Pattern is in the form:

  $\langle S \rangle \{ \langle attr(A) \rangle \}_{0..n} \{ \langle rel(A, R_k) \rangle \} \{ \langle attr(R_k) \rangle \}_{0..n} \}_{0..n}$
    $\{ \langle rel(R_k, R_k) \rangle \} \{ \langle attr(R_k) \rangle \}_{0..n} \}_{0..n}$ *where*

  − $attr/1$ *represents the predicate associated to the target table T (key predi-*
    cate). The argument represents the primary key of T.
  − $rel/2$ *represents a generic structural predicate*
  − $attr/2$ *represents a generic property predicate*

A pattern P in this form is a relational pattern if the property of linkedness [12]
is satisfied (e.g. each variable should be linked to the variable in the key predicate
by means of structural predicates). A relational pattern P is a relational emerging
pattern if $\exists t \in C$ $GR_{\bar{t} \to t}(P) > minGR$ and $sup_t(P) > minSup$.

Definitions of support and growth-rate are formally provided below.

Definition 4. *The support $sup_t(P)$ of the pattern P for the class $t \in C$ is:*
$sup_t(P) = |O_t(P)|/|O_t|$, where: O_t denotes the set of reference objects labeled
with class t, while $O_t(P)$ denotes the subset of reference objects in O_t which are
covered by the pattern P.

Definition 5. *The growth rate $GR_{\bar{t} \to t}(P)$ of the pattern P for distinguishing*
the reference objects labeled with class t from the reference objects labeled with a
class different from t (in $\bar{t} = C - \{t\}$), is computed as follows:

$$GR_{\bar{t} \to t}(P) = \frac{sup_t(P)}{sup_{\bar{t}}(P)} \tag{1}$$

$GR_{\bar{t} \to t}(P) = \frac{0}{0} = 0$ *and* $GR_{\bar{t} \to t}(P) = \frac{\geq 0}{0} = \infty$.

---

[2] "/2" indicates the predicate arity.

The relational emerging pattern discovery is performed by exploring level-by-level the lattice of relational patterns ordered according to a generality relation ($\geqslant$) between patterns. Formally, given two patterns $P1$ and $P2$, $P1 \geqslant P2$ denotes that $P1$ ($P2$) is more general (specific) than $P2$ ($P1$). Hence, the search proceeds from the most general pattern and iteratively alternates the candidate generation and candidate evaluation phases (levelwise method). In [2], the authors propose an enhanced version of the level-wise method [15] to discover emerging patterns from data scattered in multiple tables of a relational database. Candidate emerging patterns are searched in the space of relational patterns satisfying linkedness, which is structured according to the θ-subsumption generality order [17].

Definition 6 (θ-Subsumption). *Let $P1$ and $P2$ be two relational patterns on a data schema S. $P2$ θ-subsumes $P1$ if and only if a substitution θ exists such that $P1\ \theta \subseteq P2$.*

Having introduced θ-subsumption, generality order between relational patterns can be formally defined.

Definition 7 (Generality Order Under θ-Subsumption). *Let $P1$ and $P2$ be two relational patterns. $P1$ is more general than $P2$ under θ-subsumption, denoted as $P1 \geqslant_\theta P2$, if and only if $P2$ θ-subsumes $P1$, that is $P1\ \theta \subseteq P2$ for some substitution θ.*

θ-subsumption defines a quasi-ordering, since it satisfies the reflexivity and transitivity property but not the anti-symmetric property. The quasi-ordered set spanned by $\geqslant_\theta$ can be searched according to a downward refinement operator which computes the set of refinements for a relational pattern.
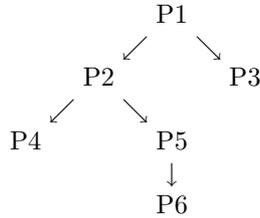
Definition 8 (Downward Refinement Operator Under θ-Subsumption). *Let $\langle G, \geqslant_\theta \rangle$ be the space of relational patterns ordered according to $\geqslant_\theta$. A downward refinement operator under θ-subsumption is a function ρ such that $\rho(P) \subseteq \{Q \in G | P \geqslant_\theta Q\}$.*

The downward refinement operator is a refinement operator under θ-subsumption. In fact, it can be easily proved that $P \geqslant_\theta Q$ for all $Q \in \rho(P)$. This makes possible to perform a levelwise exploration of the lattice of relational patterns ordered by θ-subsumption.

Example 1. Let us consider the relational patterns:
   P1: molecule(M).
   P2: molecule(M), atom(M,A).
   P3: molecule(M), logp(M,[1.8,3.7]).
   P4: molecule(M), atom(M,A), charge(A,[1.7,2.1]).
   P5: molecule(M), atom(M,A), bond(B,A).
   P6: molecule(M), atom(M,A), bond(B,A), type(B,covalent).

They are structured in a portion of a lattice ordered by θ-subsumption, that is:

```
                    P1
                  ╱    ╲
              P2          P3
            ╱    ╲
        P4          P5
                    ↓
                    P6
```

Emerging patterns for distinguishing reference objects labeled with class t (target class) from reference objects labeled with class b (background class) are then discovered by generating the pattern space one level at a time starting from the most general emerging pattern (the emerging pattern that contains only the key predicate) and then by applying a breadth-first evaluation in the lattice of relational patterns ordered according to $\geqslant_\theta$.

When a level of the latticeis explored, the candidate pattern search space is represented as a set of enumeration trees (SE-trees)[20]. The idea is to impose an ordering on atoms such that all patterns in the search space are enumerated. Practically, a node g of a SE-tree is represented as a group comprising: the *head* ($h(g)$) that is the pattern enumerated at g, and the *tail* ($t(g)$) that is the ordered set consisting of the atoms which can potentially be appended to g by ρ in order to form a pattern enumerated by some sub-node of g. A child g_c of g is formed by taking an atom $i \in t(g)$ and appending it to $h(g)$, $t(g_c)$ contains all atoms in $t(g)$ that follow i (see Figure 1). In the case i is a structural predicate (i.e., a new relation is introduced in the pattern), $t(g_c)$ contains both atoms in $t(g)$ that follow i and new atoms that can be introduced only after i has been introduced (according to linkedness property). Given this child expansion policy, without any pruning of nodes or pattern, the SE-tree enumerates all possible patterns and avoids generation and evaluation of candidates equivalent under θ-subsumption to some other candidate.

As pruning criterion, the monotonicity property of the generality order $\geqslant_\theta$ with respect to the support value (i.e., a superset of an infrequent pattern cannot be frequent) [1] can be exploited to avoid generation of infrequent relational patterns. Let P' be a refinement of a pattern P. If P is an infrequent pattern
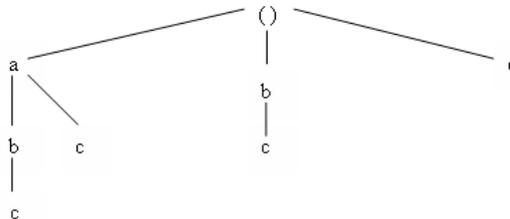

Fig. 1. The enumeration tree over the atoms $A = \{a, b, c\}$ to search the atomsets a, b, c, ab, ac, bc, abc

for a class $t \in C$ $(sup_t(P) < minSup)$, then P' has a support on O_t that is lower than the user-defined threshold ($minsup$). According to the definition of emerging pattern, P' cannot be an emerging pattern for distinguishing O_t from $O_{\overline{t}}$, hence it is possible to avoid the refinement of patterns which are infrequent on O_t. Unluckily, the monotonicity property does not hold for the growth rate: a refinement of an emerging pattern whose growth rate is lower than the threshold $minGR$ may or may not be an emerging pattern. However, growth rate can be also used for pruning. In particular, it is possible to stop the search when it is not possible to increase the growth rate with additional refinements [2].

Finally, as stopping criterion, the number of levels in the lattice to be explored can be limited by the user-defined parameter $MAX_L \geq 1$ which limits the maximum number of predicates in a candidate emerging pattern.

4   Relational Classification

Once relational emerging patterns are extracted, they are used to mine classifiers. Two EPs-based relational classifiers are proposed in the following: Mr-CAEP and Mr-PEPC.

4.1   Mr-CAEP

Mr-CAEP (Multi-Relational Classification based on Aggregating Emerging Patterns) upgrades the EP-based classifier CAEP [7] to the relational setting. It computes a membership score of an object to each class. The score is computed by means of growth rate based function of the relational EPs covered by the object to be classified. The largest score determines the object's class.

The score is computed on the basis of the subset of relational emerging patterns that cover the object to be classified. Formally, let o be the description of the object to be classified (an object is represented by a tuple in the target table and all the tuples related to it according to foreign key constraints), $\Re(o) = \{R_k \in \Re | \exists \theta \ R_k \theta \subseteq o\}$ is the set of relational emerging patterns that cover the object o.

The score of o on the class C_i is computed as follows:

$$score(o, C_i) = \sum_{R_k \in \Re(o)} \frac{GR_{\overline{C_i} \to C_i}(R_k)}{GR_{\overline{C_i} \to C_i}(R_k) + 1} sup_{C_i}(R_k) \qquad (2)$$

This measure may result in an inaccurate classifier in the case of unbalanced datasets that is, when training objects are not uniformly distributed over the classes. In order to mitigate this problem, we follow the same solution proposed in [7] and we normalize this score on the basis of the median of the scores obtained from training examples belonging to C_i.

Formally, the classification is based on the following equation:

$$class(o) = argmax_{C_i \in C} \frac{score(o, C_i)}{median_{to \in TS}(score(to, C_i))} \qquad (3)$$

where TS represents the training set.

4.2   Mr-PEPC

In Mr-PEPC (Multi-Relational Probabilistic Emerging Patterns Based Classifier), relational emerging patterns are used to build a naïve Bayesian classifier which classifies any object o by maximizing the *posterior probability* $P(C_i|o)$ such that o is of class C_i, that is, $class(o) = arg\ max_i P(C_i|o)$. By applying the Bayes theorem, $P(C_i|o)$ is reformulated as: $P(C_i|o) = \frac{P(C_i)P(o|C_i)}{P(o)}$. Since $P(o)$ is independent of the class C_i, it does not affect $class(o)$, that is, $class(o) = arg\ max_i P(C_i)P(o|C_i)$. Under the conditional independence assumption (*naïve Bayes assumption*), the likelihood $P(o|C_i)$ can be factorized: $P(o|C_i) = P(o_1, \ldots, o_m|C_i) = P(o_1|C_i) \times \ldots \times P(o_m|C_i)$, where $o_1, \ldots, o_m$ represent the set of attribute values. Surprisingly, naïve Bayesian classifiers have been proved accurate even when the conditional independence assumption is grossly violated [5].

The formulation reported above for naïve Bayesian classifiers is clearly limited to propositional representations. In the case of structural representations, some extensions are necessary. The basic idea is that of using a set of relational emerging patterns to describe an object to be classified, and then to define a suitable decomposition of the likelihood $P(o|C_i)$ *à la* naive Bayesian classifier to simplify the probability estimation problem. $P(o|C_i)$ is computed on the basis of the subset $\Re(o) \subseteq \Re$:

$$P(o|C_i) = P(\bigwedge_{R_k \in \Re(o)} R_k|o_i). \qquad (4)$$

The straightforward application of the naïve Bayes independence assumption to all literals in $\bigwedge_{R_k \in \Re(s)} R_k$ is not correct, since it may lead to underestimate the probabilities for the case of classes for which several emerging patterns are found.

To prevent this problem we resort to the logical notion of factorization [18] which is given for clauses (i.e., disjunctions of literals) but we adapt it to the notion of relational pattern.

Definition 9. *Let P be a relational pattern, which has a non-empty subset $Q \subseteq P$ of unifiable literals with most general unifier (*mgu*) θ. Then $P\theta$ is called a* factor *of P.*

A factor of a pattern P is obtained by applying a substitution θ to P which unifies one or more literals in P, and then deleting all but one copy of these unified literals. In our context we are interested in particular factors, namely those that are obtained by substitutions θ which satisfy three conditions:

1. $Domain(\theta) = \bigcup_{R_k \in \Re(o)} Vars(R_k)$ that is, the domain of θ includes all variables occurring in the relational pattern $R_k \in \Re(o)$,
2. $Domain(\theta) \cap Range(\theta) = \oslash$, that is, θ renames all variables occurring in the association rules $R_k \in \Re(o)$ with new variable names,

3. $\theta_{|Vars(R_k)}$ is injective, that is, the restriction of θ on the variables occurring in R_k is injective.

For each pattern P, a factor always exists. In the trivial case, it coincides with P up to a redenomination of variables in P. A factor $P\theta$ is minimal, when there are no other factors of P with less literals than $P\theta$.

As stated previously, a straightforward application of the naïve Bayes independence assumption may result in totally unreliable probability estimates because of the presence of redundant literals. For this reason, we impose that $P(o|C_i) = P(F|C_i)$ for any minimal factor F of $\bigwedge\limits_{R_k \in \Re(o)} R_k$.

By separating the contribution of the conjunctions of literals corresponding to structural predicates $(struct(F))$ from the contribution of the conjunction of literals corresponding to property predicates $(props(F))$ we have:

$$P(o|C_i) = P(struct(F)|C_i) \times P(props(F)|struct(F) \wedge C_i) \qquad (5)$$

Under the naïve Bayes independence assumption, $P(struct(F)|C_i)$ can be factorized as follows:

$$P(struct(F)|C_i) = \prod_{rel_j(A,B) \in \ struct(F)} P(rel_j(A,B)|C_i), \qquad (6)$$

where $P(rel_j(A,B))$ is computed on the basis of the relative frequency, computed on the training set, that two tuples are associated according to foreign key constraints.

The naïve Bayes conditional independence can also be assumed for the computation of $P(props(F)|struct(F) \wedge C_i)$, in which case

$$P(props(F)|struct(F) \wedge C_i) = \prod_{attr_j(A,v) \in \ props(F)} P(attr_j(A,v)|struct(F) \wedge C_i).$$
$$\qquad (7)$$

where $P(attr_j(A,v)|struct(F) \wedge C_i)$ is computed on the basis of the relative frequency, computed on the training set, that the property predicate is satisfied given $struct(F)$ and C_i.

5   Experimental Results

Mr-CAEP and Mr-PEPC have been implemented as modules of the multi-relational data mining system MURENA (MUlti RElational aNAlyzer) which interfaces the Oracle 10g DBMS. We tested the methods on two real world data sets: the dibEmail dataset and the North-West England Census Data. Both data sets include numeric attributes, which are handled through an equal-width discretization to partition the range of values into a fixed number of bins. Experiments aim at both comparing Mr-CAEP vs. MrPEPC and comparing emerging patterns based classification against associative classification approaches in the context of relational data mining.

5.1   Data Sets

Dibemail Data Set. These data concern incoming emails processed by the mail server of the Department of Computer Science at University of Bari in the period between 20-Aug-2007 and 3-Oct-2007.

In all, there are 91,291 incoming messages. For each incoming message some information are stored. In particular, the client that dispatched the email, the ip address of the machine that sent the email, the email size, time and date the email was received, percentage of existing destination accounts, number of specified accounts. For each incoming message, multiple outgoing message can be generated: one for each specified account plus accounts automatically generated (mailing lists). In all, there are 111,787 outgoing message and for each outgoing message, the receiver is stored. For each receiver (there are 188 receivers, one for each employed) some information are available such as: role, number of taught courses (if lecturer). For each course (there are 1,044 stored courses), we considered the number of students attending the course, attending period and academic year.

The training set referred to emails received in the period 20-Aug-2007 : 15-Sep-2007 (52,920 incoming messages), while emails received in the period 16-Sep-2007 : 3-Oct-2007 (38,371 incoming messages) are considered as testing objects. In this application, the goal is to classify emails as spam or no-spam.

The North-West England Census Data. Data were obtained from both census and digital maps provided by the European project SPIN! (http://www.ais.fraunhofer.de/KD/SPIN/project.html). They concern Greater Manchester, one of the five counties of North West England (NWE). Greater Manchester is divided into into 214 census sections (wards). Census data are available at ward level and provide socio-economic statistics (e.g. mortality rate) as well as some measures of the deprivation of each ward according to information provided by Census combined into single index scores. We employed the Jarman score that estimates the need for primary care, the indices developed by Townsend and Carstairs to perform health-related analyses, and the DoE index which is used in targeting urban regeneration funds. The higher the index value the more deprived the ward. The analysis we performed was based on deprivation factors and geographical factors represented in topographic maps of the area. Vectorized boundaries of the 1998 census wards as well as of other Ordnance Survey digital maps of NWE are available for several layers such as urban area (115 lines), green area (9 lines), road net (1687 lines), rail net (805 lines) and water net (716 lines). Elements of each layer are stored as tuples of relational tables including information on the type (TYPE). For instance, an urban area may be either a "large urban area" or a "small urban area". Topological relationships between wards and objects in these layers are materialized as relational tables expressing non-disjoint relations. The number of materialized "non disjoint" relationships is 5313.

In this application, the goal is to classify the DoE index for targeting urban regeneration funds. Results are obtained by means of a 10-fold cross validation.

5.2   Mr-CAEP vs Mr-PEPC

In Table 1, results on the dibEmail Data set are reported. This table reports accuracy results of Mr-CAEP and Mr-PEPC as well as the number of relational EPs discovered by varying values of $minGR$ and $minSup$. Results are in favour of Mr-CAEP when $minGR = 1$. The situation changes for $minGR = 1.5$ when results do not show a clear advantage of one approach over the other. Indeed, it seems that Mr-PEPC suffers from problems coming from the high number of probabilities to be computed and takes advantage from highly discriminating emerging patters.

Table 1. Accuracy results on the dibEmail testing set for different values of $minGR$ and $minSup$ ($MAX_L = 3$)

minGR	minSup	Mr-CAEP	Mr-PEPC	No of discovered EPs
1	0.05	97.93	82.7	254
1	0.1	97.97	83	214
1.5	0.05	98.11	95.53	21
1.5	0.1	97.97	97.7	14

From a descriptive data mining perspective, the following EP has been discovered for the class "spam":

$$inmessage(A) \wedge inmessage\_outmessage(A, B) \wedge$$
$$\wedge inmessage\_percentageExistingAccounts(A, [48.0..57.6]).$$

This pattern has support of 0.06 and growth rate 5.32. It captures the fact that when an incoming message is associated with an outcoming message and the there is a high percentage of not existing destination accounts (percentage of existing account between 48% and 57.6%), it can be used to discriminates spam messages from nospam messages.

The same behavior is observed on North West England Census Data. In particular, in Table 2, it becomes evident the fact that the two classifiers reach the same maximum predictive accuracy for different values of growth rate. Indeed, in Figure 2, we see that Mr-CAEP is able to take advantage from the high number of discovered emerging patterns, while Mr-PEPC reaches the best performance when $minGR = 2$. By considering only Emerging Patterns with growth rate equal to *Infinity* (Jumping Emerging Patterns), the accuracy of both classifiers decreases. This suggests us that some useful information is lost when working only with Jumping Emerging Patterns.

5.3   Associative Classification vs Emerging Patterns Based Classification

In this subsection we compare Mr-PEPC with its associative classification counterpart described in [4] in the context of spatial data mining. In that case, the

Table 2. North West England Census Data. Mr-CAEP Vs Mr-PEPC: 10-Cross Validation average accuracy and average number of discovered EPs for different values of $minGR$ ($minSup = 0.1$, $MAX_L = 3$).

minGR	Mr-CAEP	Mr-PEPC	No discovered Eps
1	93.18	90	1858
1.5	91.36	92.27	820.9
2	90.28	93.18	603.9
2.5	90.28	91.82	549.6
3	90.28	90.91	531.7
3.5	90.28	90.28	518.7
7	90.28	89.37	345
30	90.28	89.83	280.7
100	90.28	89.83	280.7

classification is based on extracted association rules and is performed by resorting to a naïve Bayesian classifier. The comparison is performed on the same cross validation framework.

Results of the associative classifier on North West England Census Data are reported in Table 3. Results vary on the basis of minimum support, minimum confidence and number of literals in extracted association rules. It is noteworthy, by comparing results in Table 2 with results in Table 3 in a predictive data mining perspective, that the emerging pattern based classifier outperforms its associative classification counterpart when $minGR \in [1.5, 2.5]$.

In a descriptive data mining perspective, it is not possible to compare results. Mr-PEPC returns relational emerging patterns, while the associative classifier returns association rules. In the following, an example of discovered relational emerging pattern for the class DoE index = low (zone to be addressed by regeneration funds) is reported:

$$ward(A) \wedge ward\_rails(A, B) \wedge ward\_carstairsidx(A, [-2.30..0.23]).$$

This emerging pattern has support of 0.24 and growth rate *Infinity*. It captures the fact that when a ward is crossed by a railway and has a low value of Carstairs index there is high demand for urban areas' regeneration projects.
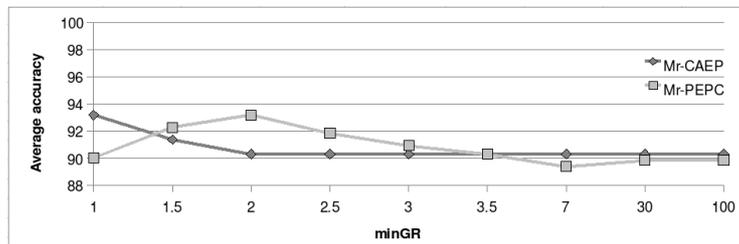

Fig. 2. Mr-CAEP Vs Mr-PEPC: average accuracy varying $minGR$ on North West England Census Data ($minSup = 0.1$, $MAX_L = 3$)

Table 3. DoE Index Associative Classification average accuracy

minsup	minconf	K	Associative Classification
0.2	0.8	5	**90.3**
0.2	0.8	6	**88.3**
0.2	0.8	7	**87.4**
0.1	0.6	5	**91.2**
0.1	0.6	6	**91.2**
0.1	0.6	7	**91.2**

6   Conclusions

In this paper, we presented two emerging patterns based classifiers that work in the multi-relational setting. The first approach classifies objects on the basis of an heuristic evaluation function. The second approach is based on a probabilistic evaluation. By comparing the two approaches, we noted that the probabilistic approach suffers from the high number of considered emerging patters, but takes advantages from more discriminative patterns. The comparison with an associative classification approach, in fair conditions, confirm the advantage of relational emerging patterns discovery. This because associative classification considers many association rules that are not able to discriminate one class from the others.

As future work, we intend to evaluate scalability of the proposed approaches.

References

1. Agrawal, R., Imielinski, T., Swami, A.N.: Mining association rules between sets of items in large databases. In: Buneman, P., Jajodia, S. (eds.) International Conference on Management of Data, pp. 207–216 (1993)
2. Appice, A., Ceci, M., Malgieri, C., Malerba, D.: Discovering relational emerging patterns. In: Basili, R., Pazienza, M. (eds.) AI*IA 2007. LNCS (LNAI), vol. 4733, pp. 206–217. Springer, Heidelberg (2007)
3. Baralis, E., Garza, P.: Majority classification by means of association rules. In: Lavrac, N., Gamberger, D., Todorovski, L., Blockeel, H. (eds.) PKDD 2003. LNCS (LNAI), vol. 2838, pp. 35–46. Springer, Heidelberg (2003)
4. Ceci, M., Appice, A.: Spatial associative classification: propositional vs structural approach. Journal of Intelligent Information Systems 27(3), 191–213 (2006)
5. Domingos, P., Pazzani, M.: On the optimality of the simple bayesian classifier under Zeo-Ones loss. Machine Learning 28(2-3), 103–130 (1997)

6. Dong, G., Li, J.: Efficient mining of emerging patterns: Discovering trends and differences. In: International Conference on Knowledge Discovery and Data Mining, pp. 43–52. ACM Press, New York (1999)
7. Dong, G., Zhang, X., Wong, L., Li, J.: CAEP: Classification by aggregating emerging patterns. In: Arikawa, S., Furukawa, K. (eds.) DS 1999. LNCS (LNAI), vol. 1721, pp. 30–42. Springer, Heidelberg (1999)
8. Džeroski, S., Lavrač, N.: Relational Data Mining. Springer, Heidelberg (2001)
9. Fan, H., Ramamohanarao, K.: An efficient singlescan algorithm for mining essential jumping emerging patterns for classification. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 456–462 (2002)
10. Fan, H., Ramamohanarao, K.: A bayesian approach to use emerging patterns for classification. In: Australasian Database Conference, vol. 143, pp. 39–48. Australian Computer Society, Inc. (2003)
11. Fan, H., Ramamohanarao, K.: A weighting scheme based on emerging patterns for weighted support vector machines. In: Hu, X., Liu, Q., Skowron, A., Lin, T.Y., Yager, R.R., Zhang, B. (eds.) IEEE International Conference on Granular Computing, pp. 435–440 (2005)
12. Helft, N.: Inductive generalization: a logical framework. In: Progress in Machine Learning, pp. 149–157. Sigma Press (1987)
13. Li, J., Dong, G., Ramamohanarao, K., Wong, L.: DeEPs: A new instance-based lazy discovery and classification system. Machine Learning 54(2), 99–124 (2004)
14. Liu, B., Hsu, W., Ma, Y.: Integrative classification and association rule mining. In: Proceedings of AAAI Conference of Knowledge Discovery in Databases (1998)
15. Mannila, H., Toivonen, H.: Levelwise search and borders of theories in knowledge discovery. Data Mining and Knowledge Discovery 1(3), 241–258 (1997)
16. Pazzani, M., Mani, S., Shankle, W.: Beyond concise and colorful: learning intelligible rules. In: Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, pp. 235–238. AAAI Press, Menlo Park (1997)
17. Plotkin, G.D.: A note on inductive generalization. 5, 153–163 (1970)
18. Robinson, J.A.: A machine oriented logic based on the resolution principle. Journal of the ACM 12, 23–41 (1965)
19. Yin, X., Han, J.: CPAR: Classification based on predictive association rules. In: SIAM International Conference on Data Mining (2003)
20. Zhang, X., Dong, G., Ramamohanarao, K.: Exploring constraints to efficiently mine emerging patterns from large high-dimensional datasets. In: Knowledge Discovery and Data Mining, pp. 310–314 (2000)