Learning Patterns in Multidimensional Space Using Interval Algebra

A. Osmani

LIPN-CNRS UMR 7030,
99, Avenue J.-B. Clément 93430 Villetaneuse- FRANCE
ao@lipn.univ-paris13.fr

Abstract. In this paper we propose a machine learning formalism based on generalized intervals. This formalism may be used to diagnose breakdown situations in telecommunication networks. The main task is to discover significant temporal patterns in the large databases generated by the monitoring system. In this kind of applications, time duration is relevant to the alarms identification process. The shapes of the decision boundaries are usually axis-parallel with constraints. The representation of examples in our formalism is similar to the representation described in the Nested Generalized Exemplar theory [Sal91]. This theory of generalization produces an excellent generalization with interpretable hypotheses [WD95] in domains where the decision boundaries are axis-parallel.
Using Allen qualitative relations between intervals, firstly we will give an adapted organization of the set of relations, then we will define an operator of generalization and we will give a table of qualitative generalization. Finally we suggest two learning algorithms. The second one uses a topologic lattice between relations to optimize the first one.

1 Introduction

One interesting solution to predict or to explain the behavior of a system is to build a general model able to imitate the system. It is typically the case in model-based diagnosis approaches where the problem is to detect, to localize and to identify failures in the system. In several approaches, breakdown situations are simulated in the model [Osm99b], and a set of observations is treated, and then archived. These observations are ordered in time. Sometimes, durations are required: the alarm CT1(technique center) is happened 5 minutes 3 seconds before the alarm CM2(switch). In other situations only the order is required: the alarm SCS is observed after CT1. In this paper, we propose a machine learning approach based on generalized intervals[BCdCO98,Lig91,Lad86] to treat the observations in order to simplify the identification process.

In 1991, Salzberg [Sal91] proposed a learning theory, using hyperrectangles, and showed the relevance of this theory in three cases. In the model of representation NGE (Nested Generalized Exemplar), Salzberg proposed a new way to describe concepts by using hyperrectangles. He associates a weight parameter for each hyperrectangle He considers that the function of distance between hyperrectangles can change dynamically and he takes into account the generalization with exception by using the Thornton [Tho87] results. The work proposed in [WD95] analyses the performance of the NGE

D. Scott (Ed.): AIMSA 2002, LNAI 2443, pp. 31–40, 2002.

approach compared to the $k-$nearest neighbors (knn) approach in 11 domains. He shows that the performance of the hyperrectangle approach is poor, except when the application is adapted; in which case the approach presents good performances with some appreciates advantages like the level of abstraction and the quantity of information necessary to describe the examples.

It is especially the case with the applications related to the learning of alarm sequences for the faults discrimination in a telecommunication network. In order to model the behavior of the telecommunication network when an abnormal operation occurs, we use the simulator proposed in [Osm99a]. This simulator of breakdown situations in telecommunication networks can reproduce components behavior and messages propagation between components in the network. This simulator builds generalized intervals (see example 2), each generalized interval became a learning example of an abnormal situation. This application is an excellent example for the use of the learning approach based on generalized intervals.

In this paper, we introduce a learning approach based on generalized intervals and which uses both quantitative and qualitative relations between intervals. Our goal is to implement a global solution of learning using generalized intervals with constraints.

In section 1, we give a real example that justifies our interest in the machine learning approach using this kind of representation. Section 2 gives some definitions and shows our relations between intervals extracted from Allen's relations. Section 3 defines the operator of generalization and introduces the table of generalization using our relations between intervals. Section 4 presents two learning algorithms; the first one uses naively the operator of generalization, the second one uses the property of the relation's lattice between interval relations defining a conceptual neighborhood structure to organize examples in order to optimize the first learning algorithm. Section 5 concludes this paper.

Related Work

Several works focus on machine learning using intervals and/or rectangles. Methods that induce logical conjunctions are a good example of orthogonal rectangle bias. The induction of logical conjunctions is one the early machine learning approaches. The original idea was presented by Bruner and al. [BA56], while the first popular implemented IGS (Incremental general-to-specific) algorithm was proposed by Winston [Win75]. This framework has been extended by Mitchell [Mit77] by combining ISG and IGS approaches. Langley's Book [Lan96] gives some complements about this learning approach.
PAC learning of rectangles, have been also studied because they have been experimentally showed to yield excellent hypotheses for several applied problems [ALS97,WD95, WK91].
Nested generalized exemplar theory accomplishes the learning task by storing objects in Euclidian space as hyperrectangle [Sal91]. The hyperrectangles may be nested inside another to arbitrary depth. Some applications results are presented in [WD95].

Other works deal with intervals to learn and to classify patterns in Euclidian space: In Koc[Koc95] an algorithm (COFI) for classification with overlapping feature intervals is proposed. COFI algorithm is an exemplar-based concept-learning algorithm where learning concepts are represented as intervals on the class dimensions for each feature. The knowledge representation is similar to the NGE method, no domain theory is used. In COFI algorithm, learning task is performed in a dynamic environment. The prediction step is based on a majority voting taken among individual predictions based on the votes of the features. The nearest work to the one presented in this paper is the paper presented by Hoppner[Hop01]. Hoppner uses interval algebra to discover temporal rules behavior of multivariate time series.

2 Application Example

This section gives some "generalized examples" from the learning database generated by AutModSim tool[1] describing breakdown situations in telecommunication network. The interpretation of the first line of the matrix is done as follow: if the supervisor of the telecommunication network receives the alarms a_1 in the interval [2,7], a_2 in the interval [2,4] and a_3 in the interval [0,1], probably the network center P_{ct} is in breakdown situation.

$$\begin{array}{cc} & \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \\ \begin{bmatrix} p\_ct \\ p_{ct} \\ p_{cm} \end{bmatrix} & \begin{bmatrix} [2,7] & [2,4] & [0,1] \\ [1,2] & [3,4] & \\ [0,1] & [0,1] & [2,5] \end{bmatrix} \end{array}$$

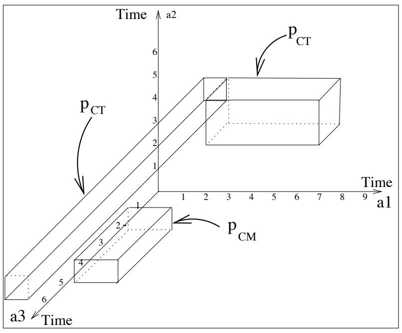

Fig. 1. Illustration.

More generally, the learning database can contain simple examples or generalized examples[2] in the form (breakdown situation, alarm sequence). Initially, we have m known breakdown situations (m is a subset of possible breakdown situations): $\{P_1, \ldots, P_m\}$ and n types of alarms: $\{a_1, \ldots, a_n\}$. These alarms define the dimension

[1] AutModSim simulates breakdown situations in telecommunication networks and generates alarm sequences with time intervals [OL].

[2] A simple example is represented with point in multidimensional space. A generalized interval is represented by generalized interval.

of our Euclidian space. For each breakdown situation P_i, we associate a set of generalized example: the example j of the breakdown situation P_i is the sequence $(a_1, X^j_{i1}, \ldots, X^j_{in})$ where X_{ij} is an interval. In this paper we consider the generalized interval $X^j_i = (X^j_{i1}, \ldots, X^j_{in})$ as the representative of the example j of P_i.

The learning database will be in the following form:

$$\begin{array}{cc} & \begin{bmatrix} a_1 & \ldots & a_n \end{bmatrix} \\ \begin{bmatrix} P_1 \\ \vdots \\ P_1 \\ \vdots \\ P_m \\ \vdots \\ P_m \end{bmatrix} & \begin{bmatrix} X^1_{11} & \ldots & X^1_{1n} \\ \vdots & & \vdots \\ X^{m_1}_{11} & \ldots & X^{m_1}_{1n} \\ \vdots & & \vdots \\ X^1_{m1} & \ldots & X^1_{mn} \\ \vdots & & \vdots \\ X^{m_m}_{m1} & \ldots & X^{m_m}_{mn} \end{bmatrix} \end{array}$$

This example gives one possible application of our formalism. In this paper, we will present the operator of generalization and the training algorithm.

3 Definitions

An interval x can be identified by these two ends $[x^-, x^+]$. In this case, it is called: instantiated interval.

Definition 1. *A generalized interval X is defined by a sequence of instantiated intervals or by a sequence of intervals and a constraints network between these intervals.*

The constraints are expressed by disjunctions of relations representing a subset of Allen's thirteen relations [All83]. These relations are $\{$precedes(p), meets(m), overlaps(o), finish(f), stard(s), during(d) $\}$, their opposites and the relation of equality.

Example 1. Each of the two following examples illustrates the definition of a generalized interval defined by a sequence of instantiated intervals and a generalized interval defined by a sequence of intervals and a constraints network.

$$X = ([1, 2], [3, 5], [2, 6]),\ X = \{(x_1, x_2, x_3), R_X\},\ R_X = \{x_1\, p\, x_2, x_1\, m\, x_3, x_2\, d\, x_3\}$$

Definition 2. *A generalized example is a set of examples describing the same concept. If a point represents a training example of a concept in a given space, an interval, a rectangle, and a cube that describe the same concept are generalized examples*[3].

Allen's relations between intervals [All83] are not necessary used for our learning problem. The figure 2 presents considered relationships between intervals. We consider four groups of relations: the relation disconnect ($disc$) which correspond to the relation precedes, the relation intersect ($inter$) which corresponds to the Allen's relations *meets* and *overlaps*, the relation contain ($cont$) which groups together the relations $finish$,

[3] A simple example is a particular case of a generalized interval. In the rest of the paper, example will indicate both simple and generalized example.

$start$ and $during$ and the relation of equality eq. We note the opposite relations of the relations $disc, inter$ and $cont$ by $disc^{-1}, inter^{-1}$ and $cont^{-1}$, respectively. The relations $DISC,\ CONT$ and $INTER$ indicate the relations $(disc \vee disc^{-1}), (cont \vee cont^{-1}),$ $(inter \vee inter^{-1})$, respectively. The properties of this cutting will be detailed in the next section.

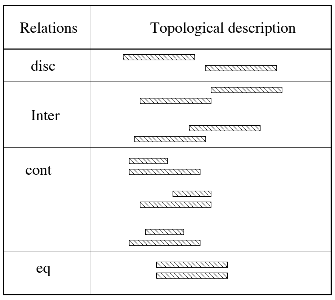

Fig. 2. $disc$, $cont$ and $inter$ relations between simple intervals.

Definition 3. *A minimal covering set of parts of the attributes space for a given concept is a space described by a set of generalized intervals E(X), such that any example describing the concept is member, at least, a generalized interval and the intersection of each side of each element of E(X) with one of the examples which it contains is nonempty. The operator of generalization presented in the next section respect this property.*

4 The Operator of Generalization ψ

This section introduces the operator of generalization ψ. This operator allows making a minimal recovery of the space of the attributes by minimizing the number of generalized intervals characterizing each concept and also by minimizing the recoveries of counterexamples during generalization.

Notation 1 *Let us consider* $X = (x_1..., x_n)$. *We note* Gx_i *an unspecified interval which generalizes* x_i. *We note* $x_{(i)}$ *the generalized interval defined as follow:* $X_{(i)} = (x_1...x_{i-1}, Gx_i, x_{i+1}...x_n)$, *and we note* $X_{(i_1...,i_k)}$ *the generalized interval* X *for which the components* $x_{i_1},\ x_{i_k}$ *are generalized before.*

Let us consider $X = (x_1, ..., x_n)$ and $Y = (y_1, ..., y_n)$ two generalized intervals with the same dimension and defined in the same space.

4.1 $\psi(X, Y)$ Operator

The operator of generalization $\psi_i(x, y)$ defines the generalization of the example X compared to the example Y as follows: $\psi_i(x, y) = x_{(i)}$ such as:

- $Gx_i = [min(x_i^-, y_i^-), max(x_i^+, y_i^+)]$ if X and Y are instantiated and
- $cont(Gx_i, x_i) \wedge \quad cont(Gxi, y_i) \wedge (\forall z_i(\neg \quad cont(G_{x_i}, z_i) \vee \neg disc(x_i, z_i) \vee \neg disc(z_i, y_i)) \wedge (\forall z_i(\neg cont(G_{x_i}, z_i) \vee \neg disc(y_i, z_i) \vee \neg disc(x_i, y_i)))$.

The application of the operator ψ_i on X and Y makes possible to replace X in the database by the generalized example $\psi_i(x, y)$.

According to the same principle, $\psi_{i_1 \ldots, i_k}(x, y)$ defines the operator of generalization of X compared to Y for the sequence of attributes $x_{i_1} \ldots, x_{i_k}$.

$$\psi_{i_1 \ldots, i_k} = \psi_{i_1} \circ \ldots \circ \ldots \psi_{i_k} = \circ_{l=i_1}^{i_k} \psi_l$$

Example 2. Let us consider two generalized examples $X = ([2, 3], [5, 6])$ and $Y = ([5, 7], [1, 3])$. Figure 3 illustrates the application of the operators ψ_1 and $\psi_{1,2}$.

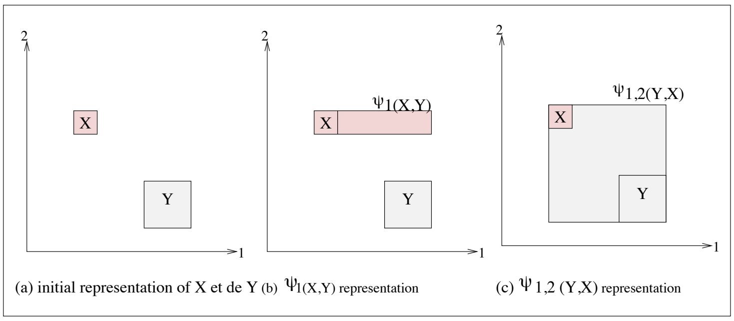

Fig. 3. Illustration of the operators ψ_i et $\psi_{i_1,\ldots,i_k}$.

The operator $\psi(X, Y)$ generalizes the two generalized examples X and Y and produces only one example Z as follows:

$$\psi(X, Y) = Z = ([min(x_1^-, y_1^-), max(x_1^+, y_1^+)] \ldots, [min(x_n^-, y_n^-), max(x_n^+, y_n^+)])$$

Proposition 1. $\psi(X, Y) = GX$ *such that* $GX = \circ_{i=1}^{i=n} \psi_i(X, Y)$

Property 1 *The operators ψ_i and $\psi_{i_1 \ldots, i_k}$ are not commutative. The operator ψ is commutative.*

Example 3. The figure 3(c) gives an example of generalization of Y with X on two attributes. If the description space of the examples contains only two attributes then $\psi(x, y) = \psi_{1,2}(y, x)$.

4.2 $\psi(X, Y, Z)$ Operator

Generalization process operates in a universe of examples and counterexamples. The operator $\psi(x, y, z)$ defines the concept of generalization of two examples X and Y of the same concept, knowing that Z is a counterexample for this same concept.

Before defining this operator, I will start with the definition of the operator $\psi_i(x, y, z)$. This operator defines the generalization of the example X compared to the example Y describing the same concept by knowing the counterexample Z.

- **dimension 1:** it indicates the behavior of $\psi_i(x, y, z) = \psi(x, y, z)$ in an one-dimensional space(i=1). Generalization is done as follows: $(\forall X)(\forall Z)(\forall Y)$
  1. $(cont(X, Y) \vee cont(Y, X) \vee inter(Y, X) \vee inter(X, Y)) \Rightarrow (\psi_i(X, Y, Z) = \psi_i(X, Y))$
  2. $(\forall Z)((disc(X, Z) \wedge disc(Y, Z)) \vee (disc(Z, X) \wedge disc(Z, Y))) \Rightarrow (\psi_i(X, Y, Z) = \psi_i(X, Y))$

  If none of the two premises is valid, generalization fails.
- **high dimensions** The definition of $\psi_i(X, Y, Z)$ $i \in \{1, .., n\}$ is done as follow: $\psi_i(X, Y, Z) = \psi_i(X, Y)$ if, and only if, one of the following expressions is checked:
  1. $(\exists j)(j \neq i)DISC(x_j, z_j)$ or
  2. $(R(x_i, z_i) \wedge R(y_i, z_i) \wedge R(x_i, y_i))) = ok$ in the table 1.

  Otherwise generalization fails.

As for the operator $\psi_{i_1,..,i_k}(X, Y)$:

$$\psi_{i_1,...,i_k}(X, Y, Z) = \circ_{m=i_1}^{m=i_k} \psi_m(X, Y, Z)$$

Table 1 gives the list of the possible situations which can occur between the examples X, Y and Z, knowing that X and Y describe the same concept and Z a different concept. The table 1 indicates when generalization is possible, i.e. $\psi_i(x, y, z) = \psi_i(x, y)$, by the word: ok, when generalization is not possible, i.e. no modification is brought to the base of the examples, by the word $\neg ok$ and when the relation between X, Y and Z is inconsistency by the word 0.
The first column indicates relations $R(X, Z)$ and $R(Y, Z)$, respectively. The first line indicates the relation $R(X, Y)$. The relations $cont$ and eq are not represented in the first line because the generalization process make no modification for X.

Proposition 2. $(\forall i)(R(x_i, z_i)R(y_i, z_i)R(x_i, y_i) = ok) \rightarrow \psi(X, Y, Z) = \psi(X, Y)$

Proposition 3. $\psi(X, Y, Z) = GX$ *such that:*

1. $GX = \circ_{i=1}^{i=n} \psi_i(X, Y)$
2. $(\forall i)$ $\psi_i(\circ_{j=1}^{j=i-1} \psi_j(X, Y), Y, Z) = \psi_i(\circ_{j=1}^{j=i-1} \psi_j(X, Y), Y)$

Let us consider $S = (Z_1, \ldots Z_l)$ the ordered set of the counterexamples for the concept described by the examples X and Y.

$$\psi(X, Y, S) = \circ_{i=1}^{i=l} \psi(X, Y, Z_i)$$

Table 1. Generalization of an example X compared to Y knowing the counterexample Z.

	disc	inter	$cont^{-1}$	$inter^{-1}$	$disc^{-1}$
disc disc	ok	ok	ok	ok	ok
disc int	ok	ok	ok	0	0
disc cont	ok	ok	ok	0	0
disc eq	ok	0	0	0	0
disc $cont^{-1}$	$\neg ok$	0	0	0	0
disc $inter^{-1}$	¬ok	0	0	0	0
disc $disc^{-1}$	¬ok	0	0	0	0
inter disc	0	0	0	ok	ok
inter int	0	ok	ok	ok	0
inter cont	0	ok	ok	0	0
inter eq	0	ok	0	0	0
inter $cont^{-1}$	$\neg ok$	0	0	0	0
inter $inter^{-1}$	¬ok	ok	0	0	0
inter $disc^{-1}$	¬ok	0	0	0	0
cont disc	0	0	0	ok	ok
cont int	0	0	0	ok	0
cont cont	0	ok	ok	ok	0
cont eq	0	0	0	0	0
cont $cont^{-1}$	0	0	0	0	0
cont $inter^{-1}$	0	ok	0	0	0
cont $disc^{-1}$	ok	ok	0	0	0

	disc	inter	$cont^{-1}$	$inter^{-1}$	$disc^{-1}$
$cont^{-1}$ disc	0	0	0	0	¬ok
$cont^{-1}$ int	0	0	ok	ok	¬ok
$cont^{-1}$ cont	0	0	ok	0	0
$cont^{-1}$ eq	0	0	ok	0	0
$cont^{-1}$ $cont^{-1}$	$\neg ok$	ok	ok	ok	¬ok
$cont^{-1}$ $inter^{-1}$	¬ok	ok	ok	0	0
$cont^{-1}$ $disc^{-1}$	¬ok	ok	0	0	0
$inter^{-1}$ disc	0	0	0	0	¬ok
$inter^{-1}$ int	0	0	0	ok	¬ok
$inter^{-1}$ cont	0	0	ok	ok	0
$inter^{-1}$ eq	0	0	0	ok	0
$inter^{-1}$ $cont^{-1}$	0	0	0	ok	¬ok
$inter^{-1}$ $inter^{-1}$	0	ok	ok	ok	0
$inter^{-1}$ $disc^{-1}$	ok	ok	0	0	0
$disc^{-1}$ disc	0	0	0	0	¬ok
$disc^{-1}$ int	0	0	0	0	¬ok
$disc^{-1}$ cont	0	0	ok	ok	ok
$disc^{-1}$ eq	0	0	0	0	ok
$disc^{-1}$ $cont^{-1}$	0	0	0	0	¬ok
$disc^{-1}$ $inter^{-1}$	0	0	ok	ok	ok
$disc^{-1}$ $disc^{-1}$	ok	ok	ok	ok	ok

5 Learning Algorithms

This section presents two training algorithms: a naive learning algorithm (NLAGI) and an optimized learning algorithm (OLAGI) which reduce the training time.

Algorithm 1 *algorithm NLAGI*

Input : $S = (E_1, ..., E_m)$ *and* $C = (C_1, ..., C_n)$
begin

```
Result = ∅ ;
for (i=1 to n) do generalized_i=false ;
for (i=1 to n) do {
  for (j=i+1 to n) do {
    if (ψ(C_i, C_j, S) = ψ(C_i, C_j))
    { generalized_i=true ; generalized_j=true ;
      C = C + ψ(C_i, C_j); n++ ;
      generalized_n=false;
    }
  }
  C = C − {C_i} ;
  if (generalized_i=false){result+=C_i ;}
}
```

EndAlgorithm

Algorithm OLAGI exploits the lattice structure between interval relations [Lig96] to define a partial order $\preceq$ between the training base of examples. In fact, examples are ordered sequentially in the database. To obtain this total ordering starting from the partial ordering, we take into account the fact that the operator ψ is noncommutative.

Figure 4(a) shows the lattice between interval relations. Figure 4(b) uses this result to define a lattice between our defined relations. We call the obtained lattice: an increased lattice. The dotted lines extract a total order between the relations by using ψ previous defining properties. In the increased lattice, $X \preceq Y$ if there is a path from X to Y.

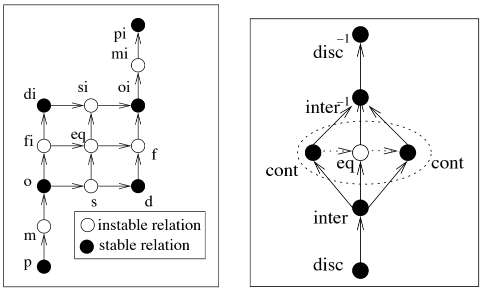

Fig. 4. (a) Simplified lattice (b) Lattice of interval relations

The extension of this relation lattice between intervals to a relation lattice between generalized intervals uses a simplified version of the generalized lattice defined in [BCdCO98].

The contribution of this property is to make the training local. Indeed, by ordering the examples of training (cost $O(n^2)$, N is the number of examples in the database), we obtain the following result:
Let us consider $X_1, \ldots, X_m$ the sequence of training examples of training such that:
$(\forall i)(\nexists j)(i < j)$ *and* $X_j \preceq X_i$

Property 2 $(\forall i)(\forall j > i)$ *such that* $X_i, X_{i+1}, \ldots, X_j$ *describe the same concept, for all* Z *in the sequence of learning examples describing another concept than* X_i*,*

$$\psi_{t,1 \leq t \leq n}(X_{k,i \leq k \leq j}, X_{l,i \leq l \leq j}, Z) = \psi_t(X_k, X_l)$$

and more generally : $\psi(X_{k,i \leq k \leq j}, X_{l,i \leq l \leq j}, Z) = \psi(X_k, X_l)$

The OLAGI learning algorithm uses this propriety to reduce learning time.

6 Conclusion

In our considered application sequences of alarms are generated by the telecommunication network. Each sequence describes a particular state of the system. We have proposed a technique to represent these sequences by generalized intervals each one describing a generalized example of a breakdown situation. This technique uses temporal CSP and translates some known results to propose an interesting generalization algorithm. In this paper we describe our contribution but the developed framework include a large part of related work (see section 2). The learning techniques are used more and more for the telecommunication management network. This article presents a formalism to learn temporal patterns with interval.

References

[All83] J.F. Allen. Maintaining knowledge about temporal intervals. *Journal of the Association for Computing Machinery*, 26:832–843, 1983.

[ALS97] P. Auer, P.M. Long, and A. Srinivasan. Approximating hyperrectangles: Learning and pseudo-random sets. In *ACM. Symposium on the Theory of Computing*, pages 314–323, 1997.

[BA56] J.S. Bruner and G.A. Auster. A study of thinking. In *New York: John Wiley*, 1956.

[BCdCO98] P. Balbiani, J.-F Condotta, L. Fariñas del Cerro, and A. Osmani. Reasoning about generalized intervals. In *LNAI 1480*, pages 50–61. Springer, AIMSA-98, 1998.

[Hop01] F. Hoppner. Discovery of temporal patterns -learning rules about the qualitative behaviour of time series. In *LNAI 2168 PKDD*, pages 192–203. Springer-Verlag Berlin, 2001.

[Koc95] H.U. Koc. *Classification with overlapping feature intervals*. Phd thesis, Bilkent University (department of computer engineering and information science), 1995.

[Lad86] P.B. Ladkin. Time representation: A taxonomy of interval relations. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 360–366, 1986.

[Lan96] P. Langley. *Elements of machine learning*. Morgan Kaufmann Publishers, 1996.

[Lig91] G. Ligozat. On generalized interval calculi. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 234–240, 1991.

[Lig96] G. Ligozat. A new proof of tractability for ORD-HORN relations. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 395–401, 1996.

[Mit77] T.M. Mitchell. Version spaces: A candidate elimination approach to rule learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 305–310, 1977.

[OL] A. Osmani and F. Lévy. Generation d'une base d'apprentissage pour l'apprentissage de pannes dans un reseau de télécommunications. In *(RFIA'2000)*.

[Osm99a] A. Osmani. *Diagnostic à base de modèles pour la supervision de pannes dans un réseau de télécommunications & raisonnement sur les intervalles*. PhD thesis, Laboratoire LIPN, Institut Galilée, Université de Paris XIII, 1999.

[Osm99b] A. Osmani. Modeling and simulating breakdown situations in telecommunication networks. In *LNAI 1611*, pages 698–707, IEA/AIE-99, 1999.

[Sal91] S. Salzberg. A nearest hyperrectangle learning method. In *Machine Learning*, volume 6(3), pages 251–276, 1991.

[Tho87] C. Thornton. Hypercuboid formation behavior of two learning algorithms. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 301–313, 1987.

[WD95] Dietrich Wettschereck and Thomas G. Dietterich. An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms. *Machine Learning*, 19(1):5–27, 1995.

[Win75] P.H. Winston. Learning structural descriptions from examples. In *In P.H. Winston (ed.). The psychology of computer vision, optADDRESS =*, 1975.

[WK91] S.M. Weiss and C.A. Kulikowski. *Computer System that learn*. San Mateo CA: Morgan Kaufmann Publishers, INC, 1991.