Discovering Plausible Explanations of Carcinogenecity in Chemical Compounds

Eva Armengol

IIIA - Artificial Intelligence Research Institute,
CSIC - Spanish Council for Scientific Research,
Campus UAB, 08193 Bellaterra, Catalonia (Spain)
eva@iiia.csic.es

Abstract. The goal of predictive toxicology is the automatic construction of carcinogenecity models. Most common artificial intelligence techniques used to construct these models are inductive learning methods. In a previous work we presented an approach that uses lazy learning methods for solving the problem of predicting carcinogenecity. Lazy learning methods solve new problems based on their similarity to already solved problems. Nevertheless, a weakness of these kind of methods is that sometimes the result is not completely understandable by the user. In this paper we propose an explanation scheme for a concrete lazy learning method. This scheme is particularly interesting to justify the predictions about the carcinogenesis of chemical compounds.

1 Introduction

During the seventies Europe and the United States respectively started long term programs with the aim of developing toxicology chemical databases. The idea was to establish standardized experimental protocols allowing to determine the carcinogenecity of chemical compounds. In particular, the American National Toxicology Program (NTP) established two protocols to be performed on rodents: a short-term protocol (90 days) and a long-term protocol (2 years). To develop both protocols is necessary to sacrify a lot of animals and sometimes the results are not clearly conclusive concerning to carcinogenecity. Moreover, even in the situation of clear carcinogenic activity of a chemical compounds on rodents, there is no certainty that the results may be extrapolable to humans.

The use of computational models applied to toxicology could contribute to reduce the cost of experimental procedures. In particular, artificial intelligence methods such as knowledge discovery and machine learning can be used for building models of carcinogenecity (see [13]). The construction of such models is called *predictive toxicology*. From the machine learning point of view, the goal of the predictive toxicology is a classification task, i.e. toxic compounds are classified as belonging to the positive class and non-toxic compounds are classified as belonging to the negative class.

Most of machine learning approaches use representations of chemical compoudns based on *structure-activity relationship (SAR)* descriptions since there

P. Perner (Ed.): MLDM 2007, LNAI 4571, pp. 756–769, 2007.

are easily obtained from commercial drug design tools ([14], www.accelrys.com/chem/, www.disat.inimib.it/chm/Dragon.htm). Concerning the classification of chemical compounds, a widely used technique to build carcinogenecity models is *inductive logic programming (ILP)*. The main idea of ILP is to induce general descriptions satisfied by a set of examples represented using logical predicates. In these approaches (for instance see [9]), compounds are represented as sets of predicates relating the atoms of the molecule and they also include information about the chemical compunds (such as molecular weight, charge, etc). Nevertheless, due to the wide variety of chemical compounds, the use of inductive learning methods for building a general model of carcinogenesis is very difficult.

In [6] we proposed the use of lazy learning methods, instead of inductive learning methods, to classify chemical compounds. The main difference among both kinds of approaches is that inductive learning methods build a model and then they use it to classify new chemical compounds. Instead, lazy approaches do not build any model, but given a new problem they try to classify it based on both its features and the similarity of that problem with already known problems. This represents an advantage because lazy learning methods are not aware of the variability of the problems but they only focus on the features of the new problem. Concerning to the toxicology domain, since chemical compounds have high variability, inductive learning methods produce models with rules that are too general. Instead, a lazy learning method only focuses on the features of the new chemical compound to assess the similarity of that compound with others compounds with known carcinogenic activity.

A weakness of lazy learning methods is the way they are able to explain the result to the user. The most common way used by case-based reasoning systems to explain the result is to show the user all the cases that support the classification of a new problem. This kind of explanation seems appropriate when domain objects are not too complicated, however when domain objects have a complicated structure the user is not able to detect similarities among the cases. McSherry [18] argues that the most similar case could be a good explanation but it also may have features that could act as arguments against that case and, therefore, against the classification that it proposes. For this reason, McSherry proposes that the explanation of a case-based reasoning system has to explicitly distinguish between the case features supporting a classification and the case features against it. In that way, the user could decide about the final solution of the problem. A related idea proposed in [17] is to use the differences among cases to support the user in understanding why some cases do not satisfy some requirements.

Our approach is based on generating an explanation scheme from the similarities among a problem and a set of cases. As the approaches of McSherry and McCarthy et al. [17], the explanation scheme of our approach is also oriented to the user. The difference of our approach with that of McSherry is that we explain the result using a set of similar cases whereas McSherry explains it using both similarities and differences among the most similar case compared to the problem at hand. An interesting part of the explanation scheme we propose is

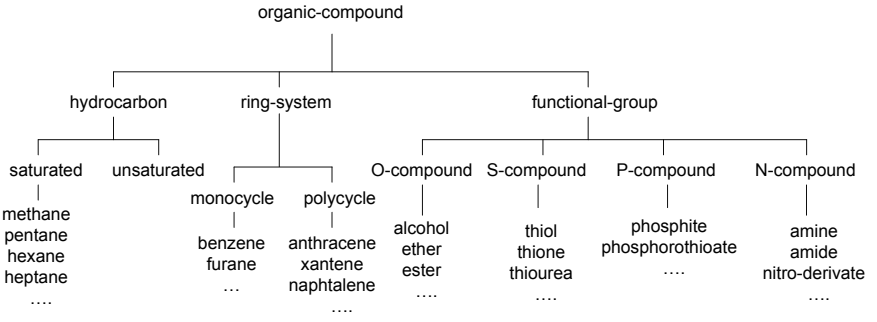


Fig. 1. Partial view of the chemical ontology

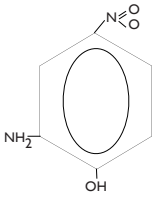
that it allows the user to focus on relevant aspects that differentiate carcinogen compounds from non carcinogen compounds.

The structure of the paper is the following. In the next section we briefly describe the formalism of feature terms, the representation we use to describe chemical compounds. Then in Section 3 we introduce LID, the lazy learning method we use to classify chemical compounds and that handles objects represented as feature terms. In Section 4 we introduce the anti-unification concept in which is based the explanation scheme described in Section 5. We end up with some related works and conclusions.

2 Representation of the Chemical Compounds Using Feature Terms

Current approaches using artificial intelligence techniques applied to chemistry use representations inherited from existing tools. These tools describe chemical compounds with a set of structure-activity relationship (SAR) descriptors because they were developed mainly for the task of drug design. In [6] we proposed the use of a representation of chemical compounds based on the *chemical ontology* given by the IUPAC nomenclature (www.chem.qmul.ac.uk/ iupac/). The IUPAC chemical nomenclature is a standard form to describe the (organic and inorganic) molecules from their chemical structure. From our point of view, a formal representation using the IUPAC nomenclature could be very useful since it allows a direct description of the chemical structure, in a way very familiar to the chemist. Our point is that, using the standard nomenclature, the name of a molecule provides enough information to graphically represent its structure. Actually, we represent a compound as a structure with substructures using the chemical ontology that is implicit in the nomenclature of the compounds. Figure 1 shows part of the chemical ontology we have used to represent the compounds in the Toxicology data set.

The implementation of our approach has been done using the *feature terms* formalism [2]. Feature terms is a kind of *relational representation*, i.e. an object



$$TR\text{-}339 = \left[\begin{array}{ll} \underline{benzene} & \\ \mathsf{radical\text{-}set} \doteq & \begin{array}{l} alcohol \\ amine \\ nitro-derivate \end{array} \\ \mathsf{p\text{-}radicals} \doteq & \left[\begin{array}{l} \underline{position\text{-}radical} \\ \mathsf{position} \doteq one \\ \mathsf{radicals} \doteq alcohol \end{array}\right] \\ & \left[\begin{array}{l} \underline{position\text{-}radical} \\ \mathsf{position} \doteq two \\ \mathsf{radicals} \doteq amine \end{array}\right] \\ & \left[\begin{array}{l} \underline{position\text{-}radical} \\ \mathsf{position} \doteq four \\ \mathsf{radicals} \doteq nitro-derivate \end{array}\right] \end{array}\right]$$

Fig. 2. Representation of TR-339, *2-amino-4-nitrophenol*, with feature terms

is described by its parts and the relationhs among these parts. The intuition behind a feature term is that it can be described as a labelled graph where nodes are objects and links are the features describing the objects. An object, as well as the values of the features of that object belong to a *sort*. A *sort* is described by a set of features, where each feature represents a relation of this sort with another sort. Sorts are related among them by partial order $\preceq$ (see 4.1) that induces a hierarchy of sorts/subsorts relating the concepts of a domain. Thus, the chemical ontology shown in Fig. 1 can be viewed as a sort/subsort hierarchy relating the chemical concepts describing the molecular structure of a chemical compound.

Let us illustrate with an exemple how chemical compounds are represented using feature terms. Figure 2 shows the molecular structure of the chemical compound TR-339, called *2-amino-4-nitrophenol*, and its representation using feature terms. Chemical compound TR-339 is represented by a feature term with *root* TR-339 of sort *benzene* described by two features: radical-set and p-radicals. The value of the feature radical-set is the set $\{alcohol, amine, nitro\text{-}derivate\}$. The value of the feature p-radicals is a set whose elements are of sort *position-radical*. In turn, the sort *position-radical* is described using two features: radicals and position. Values of radicals are those of the feature radical-set meaning the position where the radical is placed. TR-339 has the radical *alcohol* placed in position *one*, the radical *amine* in position *two* and the radical *nitro-derivate* in position *four*. Note that this information has been directly extracted from the chemical name of the compound following the nomenclature rules.

A *leaf* of a feature term is defined as a feature whose value is a (set of) feature term without features. For instance, leaf features of TR-339 are the following: {radical-set, position, radicals, position, radicals, position, radicals}. Notice that there is a leaf position and also a leaf radicals for each value of p-radicals.

```
Function LID (p, S_D, D, C)
    S_D := Discriminatory-set (D)
    if stopping-condition(S_D)
        then return class(S_D)
        else f_d := Select-leaf (p, S_D, C)
             D' := Add-path(π(root(p), f_d), D)
             S_D' := Discriminatory-set (D', S_D)
             LID (S_D', p, D', C)
    end-if
end-function
```

Fig. 3. The LID algorithm. p is the problem to be solved, D is the similitude term, S_D is the discriminatory set of D, C is the set of solution classes, $class(S_D)$ is the class $C_i \in C$ to which all elements in S_D belong.

A *path* $\Pi(root, f)$ is the sequence of features leading from the root of the feature term to the feature f. Paths of TR-339 from the root to the leaves are the following:

- TR-339.radical-set with value the set $\{alcohol, amine, nitro\text{-}derivate\}$
- TR-339.p-radicals.position with value *one*
- TR-339.p-radicals.radicals with value *alcohol*
- TR-339.p-radicals.position with value *two*
- TR-339.p-radicals.radicals with value *amine*
- TR-339.p-radicals.position with value *four*
- TR-339.p-radicals.radicals with value *nitro-derivate*

3 Lazy Induction of Descriptions

Lazy Induction of Descriptions (LID) is a lazy learning method for classification tasks. LID determines which are the most relevant features of a problem and searches in a case base for cases sharing these relevant features. The problem is classified when LID finds a set of relevant features shared by a subset of cases all o them belonging to a same class. We call *similitude term* the feature term formed by these relevant features and *discriminatory set* the set of cases satisfying the similitude term. A first version of LID was introduced in [3] to assess the risk of complications in diabetic patients. In order to assess the carcinogenecity of chemical compounds, the LID algorithm has been modified to cope with some situation that, although general, they do not occur in the diabetes domain.

Given a problem p, the LID algorithm (Fig. 3) initializes D as a feature term such that $sort(D) = sort(p)$, with no features and with the discriminatory set S_D initialized to the set of cases satisfying D. For the Toxicology domain we set $C = \{positive, negative\}$.

Let D be the current similitude term, the first step is to form the set S_D with all the cases satisfying the similitude term D. When the stopping condition of LID is not satisfied, the next step is to select a leaf for specializing D.

The specialization of a similitude term D is achieved by adding features to it. Given a set F_l of features candidate to specialize D, the next step of LID is the selection of a leaf feature $f_d \in F_l$ to specialize the similitude term D [1]. Selecting the most discriminatory leaf feature in the set F_l is heuristically done using the López de Mántaras' distance (LM) [16] over the features in F_l. LM measures the distance among two partitions and LID uses it to compare each partition P_j induced by a feature f_j with the correct partition P_c. The *correct partition* has two sets: one with the examples belonging to a solution class C_i and the other containing the cases not in C_i. Each feature $f_j \in F_l$ induces in the discriminatory set a partition P_j with two sets, one with the cases where f_j takes the same value than p and the other with the rest. Given two features f_i and f_j inducing respectively partitions P_i and P_j, we say that f_i is *more discriminatory* than f_j iff $LM(P_i, P_c) < LM(P_j, P_c)$. This means that the partition P_i induced by f_i is closer to the correct partition than the partition P_j induced by f_j. LID selects the most discriminatory feature to specialize D.

Let us call f_d the most discriminatory feature in F_l. The specialization of D defines a new similitude term D' by adding to D the sequence of features appearing in the path $\Pi(root(p), f_d)$. After adding the path Π to D, the new similitude term $D' = D + \Pi$ subsumes a subset of cases in S_D, namely $S_{D'}$.

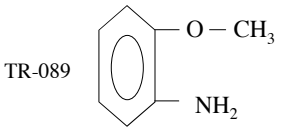
Next, LID is recursively called with the discriminatory set $S_{D'}$ and the similitude term D'. The recursive call of LID has $S_{D'}$ instead of S_D because the cases that are not subsumed by D' will not be subsumed by any further specialization. The process of specialization reduces the discriminatory set at each step, therefore we get a sequence $S_D^n \subseteq S_D^{n-1} \subseteq \ldots \subseteq S_D^0$.

LID has three possible stopping situations: 1) all the cases in the discriminatory set belong to the same solution class, 2) there is no feature allowing to specialize the similitude term, and 3) there are no cases subsumed by the similitude term.

In a previous version of LID [3] there only the stopping conditions 1) and 2) were considered. Now, in the Toxicology domain we have introduced a third stopping condition: the similitude term does not subsumes any case. Let us explain condition 3 in more detail. The similitude term is a feature term of the same sort than p, and the sort of p is any sort of the ontology of organic compounds (Fig. 1). Nevertheless, it is possible than there is no chemical compound of the same sort of p. For instance, let us suppose that *cp* is the compound TR-496 of sort *eicosane*, then the similitude term is a feature term of sort *eicosane*. This means that LID searches in the case base for chemical compounds of sort *eicosane* but there is not any other chemical compound of that sort, therefore $S_D = \emptyset$. In that situation, LID finishes without giving a solution for p.

When the stopping condition 1) is satisfied because all the cases in S_D belong to a same solution class C_i, then p is classified as belonging to C_i. When $S_D = \emptyset$ then LID gives no classification for p, and finally when the discriminatory set contains cases from several classes, then the *majority criteria* is applied, i.e. p is classified as belonging to the class of the majority of cases in S_D.

[1] In fact, the selection of a leaf implies the selection of the path from the root to the leaf.



Π_1 = benzene.radical-set (value = amine)
Π_2 = benzene.radical-set.radical-set (value = methane)
Π_3 = benzene.radical-set.p-radicals.radicals (value = methane)
Π_4 = benzene.positions.radicals (value = {methane, ether})
Π_5 = benzene.positions.distance (value = 1)

Fig. 4. Molecular structure, feature term representation and paths of the chemical compound *TR-089 (resorcinol)*

3.1 Example

In this section we explain the LID algorithm by illustrating the process with the classificacion of the chemical compound *TR-089* (Fig. 4 shows the molecular structure and paths of *TR-089*). The first step is to select a relevant feature, therefore, it is necessary to induce the partitions associated to each feature of *TR-089* and then to compute the distance to the correct partition. Using the LM distance, LID takes the feature radical-set with value *methane*. In such a situation, $D_0 = \Pi_1$ (Fig. 4) and $S_{D_0} = \{$*TR-491, TR-416, TR-414, TR-372, TR-351, TR-223, TR-171, TR-142, TR-128, TR-127, TR-124, TR-120, TR-114, TR-105, TR-084*$\}$ where some compounds are *positive* and some others are *negative* for carcinogenesis. Therefore LID has to specialize D_0 by selecting a relevant feature to be added to it. The next most relevant feature is radical-set with value *amine*. Now the discriminatory set associated to $D_1 = D_0 + \Pi_2$ is $S_{D_1} = \{$*TR-084, TR-105, TR-127, TR-142, TR-171, TR-351, TR-372*$\}$ that still contains both *positive* and *negative* compounds. Therefore a new relevant feature has to be selected. Now the selected feature is distance with value 1. The new similitude term is $D_2 = D_1 + \Pi_5$ and the discriminatory set is $S_{D_2} = \{$*TR-084, TR-127, TR-142, TR-171, TR-372*$\}$ where *TR-171* is the only compound with negative carcinogenecity for male rats. Because LID cannot further specialize D_2, since there are no features able to discriminate the compound *TR-171* from the others, it uses the majority criterion to classify *TR-089* as *positive*.

Notice that in the situation above, the result given by LID after the application of the majority rule seems clear, i.e. TR-089 is positive because *all* the cases assessed as the most similar (except TR-171) are positive. Nevertheless, sometimes such a majority is not so clear. for instance, in the current situation, the user could note that the molecular structures of all the compounds in S_{D_2} are very similar (see Fig. 6) so the question is: why TR-171 is *negative*? In the next section we propose an explanation scheme in order to justify to the user the classifications given by LID.

4 How Results of a Lazy Learning Method Can Be Explained?

Case-based reasoning (CBR) systems predict the solution of a problem based on the similarity between this problem (the *current case*) and already solved

problems (cases). Clearly, the key point is the measure used to assess the similarity among the cases. Since the resulting similarity value is difficult to explain, CBR systems often show the retrieved cases (the set of cases that have been assessed as the most similar to the new problem) to the user as an explanation of the prediction: the solution is predicted *because* the problem was similar to the cases shown. Nevertheless, when the cases have a complex structure, simply showing the most similar cases to the user may not be enough. Our proposal, similar to that introduced in [4] is to show the user a symbolic description (the final similitude term given by LID) that makes explicit what the new problem has in common with the retrieved cases.

As we already mentioned, LID has three stopping situations for the classification process of a problem p. For the first one, when all the cases in S_D belong to a same solution class, the similitude term is a good explanation since makes explicit the relevant features shared by p and a subset of classes belonging to a class. However, when the second stopping condition holds, p shares relevant features with cases from different solution classes, therefore the similitude term, by its own, is not a good justification of the result. For this resason, we take the *explanation scheme* introduced in [4] to explain results obtained by LID using the majority rule. This scheme is based on the anti-unification concept.

4.1 The Anti-unification Concept

The explanation scheme we propose is based on the concept of *least general generalization (lgg)*, commonly used in Machine Learning. The partial order $\preceq$ among sorts mentioned in Section 2 gives an informational order among sorts since $s_1 \preceq s_2$ (s_2 is a subsort of s_1) means that s_1 provides *less* information than s_2. Using the partial order $\preceq$ we can define the *least upper bound (lub)* of two sorts $lub(s_1, s_2)$ as the most specific super-sort common to both sorts. For instance, Fig. 1 shows the sort hierarchy representing this chemical ontology. The most general sort is *organic-compound* and most specific sorts are the leafs of this hierarchy (e.g. *pentane*, *hexane*, *benzene*, *furane*, etc). Thus, the super-sort of any two sorts of that hierarchy (for instance *benzene* and *furane*) is always *organic-compound*. The anti-unification concept concerns to the most specific sort of two sort, therefore the *lub(benzene, furane)* (Fig. 1) is the sort *monocycle*. Similarly, *lub(benzene, xantene)=ring-system*, and *lub(methane, O-compound) = organic-compound*.

Now, we can define the *least general generalization* or *anti-unification* of a collection of descriptions represented as feature terms (either generalizations or cases) using the relation *more general than* ($\geq_g$) as follows:

- $AU(d_1, ..., d_k) = g$ such that $(g \geq_g d_1) \wedge ... \wedge (g \geq_g d_k)$ and not exists $(g' \geq_g d_1) \wedge ... \wedge (g' \geq_g d_k)$ such that $g >_g g'$

That is to say, g is the most specific generalization of all those generalizations that cover all the descriptions $d_1, ..., d_k$. $AU(d_1, ..., d_n)$ is a feature term described by all the features common to (or shared by) $d_1, ..., d_2$, i.e it describes *all* aspects in which two ore more descriptions are similar.

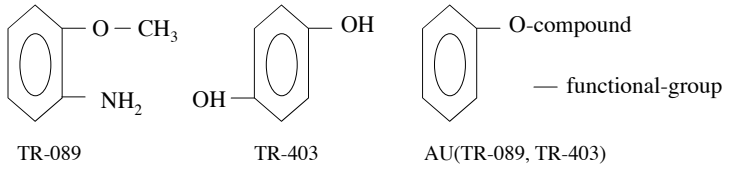


Fig. 5. Graphical representation of both the chemical compound *TR-403* and the anti-unification of *TR-089* and *TR-403*

The anti-unification of the chemical compounds *TR-089* (Fig. 4) and *TR-403* (Fig. 5) is the feature term *AU(TR-089, TR-403)*, shown in Fig. 5. *AU(C-089, C-403)* represents a chemical compound that is a *benzene* with a radical of sort *O-compound* and another radical in a non specified position. See [2] for a more detailed account on feature terms and their anti-unification. In the next section we detail the explanation scheme used to justify the classification of LID.

5 The Explanation Scheme

This section presents the way in which descriptions resulting from the anti-unification of a collection of cases can be used to explain the classification of a new problem in CBR systems. Let S_D the discriminatory set containing cases satisfying the similitude term D given by LID as a result of the classification of a problem p. There are two possible situations: 1) cases in S_D belong to only one class C_i, and 2) cases in S_D belong to several classes.

Concerning the first situation, the similitude term D is a good explanation of why the cases in S_D are similar to p, since it is a description of all that is shared among a subset of cases belonging to a some class C_i and the new problem. Let us to concentrate on the second situation.

Assuming two solution classes C_1 and C_2, let $S_D^1 \subseteq S_D$ be the set of retrieved cases hat belong to a class C_1, and $S_D^2 \subseteq S_D$ the subset of retrieved cases that belong to C_2 ($S_D = S_D^1 \cup S_D^2$). The explanation scheme we proposed in [4] is composed of three descriptions:

- AU^*: the anti-unification of p with all the cases in S_D. This description shows what aspects of the problem are shared by all the retrieved cases, i.e. the k retrieved cases are similar to p because they have in common what is described in AU^*.
- AU^1: the anti-unification of p with the cases in C^1. This description shows what has p in common with the cases in C^1.
- AU^2: the anti-unification of p with the cases in C^2. This description shows what has p in common with the cases in C^2.

This explanation scheme supports the user in the understanding of the classification of a problem p. With the explanation scheme we propose, the similarities

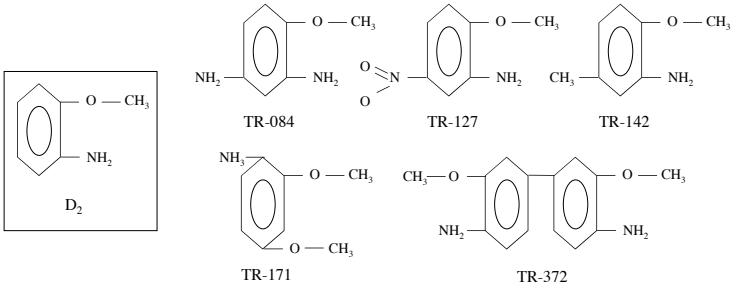


Fig. 6. Graphical representation of the similitude term D_2 and the chemical compounds contained in the discriminatory set S_{D_2} when classifying the compound *TR-089*

among p and the cases of each class are explicitly given to the user, who can decide the final classification of p. This scheme can also be used in situations where more than two classes are present in S_D, our explanation scheme is simply to build one anti-unification description for each one of them. For instance, if cases in the retrieval set belong to 4 classes the explanation scheme consists on the following symbolic descriptions: AU^1, AU^2, AU^3, and AU^4.

When the similitude term AU^* is too general (e.g. most of the features hold the most general sort as value), the meaning is that the cases have low similarity. Conversely, when AU^* is a description with some features holding some specific value, this means that the cases share something more than only the general structure. In this paper instead of AU^* we propose the use of the final similitude term D given by LID. The main difference between AU^* and D is that AU^* shows all the aspects shared by all the retrieved cases whereas D shows the *important* aspects shared by the problem and the retrieved cases, i.e. those aspects considered important to classify the problem.

AU^1 shows the commonalities among the problem p and the retrieved cases belonging to C_1. This allows the user to focus on those aspects that could be relevant to classify p as belonging to C_1. As before, the more specific AU^1 is, the more information it gives for classifying p. Notice that AU^1 could be as general as D; in fact, it is possible that both feature terms are equal. This situation means that p has not too many similar aspects with the cases of C^1. A similar situation may occur with AU^2.

In [7] an example that follows this scheme can be found. Here we illustrate the explanation scheme with the example of the classification of the chemical compound *TR-089* developed in section 3.1. This is an interesting case where the explanation scheme can support the user the search of unclear aspects of the classification of compounds. Figure 6 shows the similitude term D_2 and the discriminatory set $S_{D_2} = \{$*TR-084, TR-127, TR-142, TR-171, TR-372*$\}$ given by LID when classifying the chemical compound *TR-089*. Concerning the carcinogenesis on male rats, S_{D_2} can be partitioned in the following two subsets:

$S^1_{D_2} = \{$*TR-084, TR-127, TR-142, TR-372*$\}$ and $S^1_{D_2} = \{$*TR-171*$\}$, where compounds in $S^1_{D_2}$ are *positive* and the compound in $S^2_{D_2}$ is *negative*. The explanation scheme for chemical compound *TR-089* is the following:

- The similitude term D_2 shows that *TR-089* and the compounds in S_{D_2} have in common that they all have a benzene structure with two radicals at distance 1 among them. One of these radicals is an *ether* that in turn has a radical *methane*. The other radical is an *amine*.
- The description AU^1 is the anti-unification of *TR-089* and the chemical compounds considered as *positive* for male rats. In fact, $AU^1 = D_2$, since all positive compounds share, as before, that they are benzenes with two radicals (an *ether* with a radical *methane* and an *amine*.) with distance 1 among them.
- The description AU^2 is the anti-unification of *TR-089* and *TR-171* that is the unique compound in S_{D_2}, i.e. negative for carcinogenesis. Note that also in that case, $AU^2 = D_2$

From the descriptions AU^1 and AU^2 the user can easily observe the similarities and differences among the compounds in C_1 and those in C_2. In the current example, D_2, AU^1 and AU^2 give the same feature term as explanation, which is quite specific since common radicals have specific sorts (*benzene, ether, methane, amine*), therefore the user can conclude that all the compounds are really very similar. So, the question could be why *TR-171* is negative for carcinogenesis. All compounds in S_D (included *TR-171*) are *aromatic amines* which are highly correlated with carcinogenecity [19,1], therefore, in principle *TR-171* should also be carcinogenic. Because the *TR-171* (*2,4 - Dimethoxyaniniline hydrochloride*) is an *aniline* we performed a search on Internet asking for information about experimental results on anilines. We found from the page of the *International Agency for Research on Cancer (IARC)* that there are defined four categories of chemical compounds according to their potential carcinogenic power on humans. In particular, *anilines* are classified on category 3 corresponding to chemical compounds with inadequate evidence of carcinogenecity in humans or those compounds whose experimental evidence on animals is either inadequate or limited. In fact, the NTP report of experimental results of *TR-171* on rodents (see *long term NTP Study Reports* from web page ntp.niehs.nih.gov/) states that studies began when *2,4 - Dimethoxyaniniline hydrochloride* was suspicious to be the cause of the increment of incidence of bladder cancer among dye manufacturing industry workers. Nevertheless the experimental results on rodents did not provide a convincing evidence of the carcinogenic power of the *2,4 - Dimethoxyaniniline hydrochloride*. This means that for chemical experts, *TR-171* was at first sight a potential carcinogen and despite the experimental results on rodents show no evidence of carcinogenecity, toxic activity on other species could not be discarded.

6 Related Work

Concerning to the Predictive Toxicology domain, we have proposed 1) a new approach to represent chemical compounds and 2) a lazy approach for solving

the classification task. The most common representation of chemical compounds is using SAR descriptors which represent the compounds from several points of view (structural, physical properties, etc) and they are the basis to build equational models that relate the structure of a chemical compound with its physical-chemical properties. The main difference between the representations based on SAR and our ontological approach is that the former describe the molecular structure of the chemical compounds in an exhaustive way. Instead the representation we propose is more conceptual than SAR in the sense that it directly uses the concepts understood by the chemists.

Some authors use approaches that are not centered on the representation of specific atoms but on molecular structures. For instance, González et al [12] and Deshpande and Karypis [11] represent chemical compounds as labeled graphs, using graph techniques to detect the set of molecular substructures (subgraphs) more frequently occurring in the chemical compounds of the data set. Conceptually, these two approaches are related to ours in that we describe a chemical compound in terms of its radicals (i.e. substructures of the main group).

Concerning the explanation of the solution proposed by a CBR system, there are a lot of possible approaches depending on the kind of explanation we are looking for. Sørmo et al. [20] performed a deep analysis of the different perspectives from whose an explanation can be taken. Related to problem solving tasks, there are two main kinds of explanations that are specially useful: 1) an explanation of how a solution has been reached, and 2) an explanation justifying the result. In this sense, the explanation proposed in this paper justifies the solution proposed by the system. Nevertheless, because part of this explanation scheme (the similitude term) contains the important features that LID used to classify a new problem, the explanation also gives some clues of *how* the system reached the solution.

Most of explanations given by CBR systems are oriented to the user. Sørmo et al. [20], Leake [15] and Cassens [10] also consider that the form of the explanation depends on the user goals. This statement has been proved in the application presented by Bélanger and Martel [8] where the explanations for expert and novice users are completely different. Leake [15] see the process of explanation construction as a form of goal-driven learning where the goals are those facts that need to be explained and the process to achieve them gives the explanation as result. Cassens [10] uses the Activity Theory to systematically analyze how a user evolves in the utilization a system, i.e. how the user model is changing. The idea is that in using a system, the user can change his expectations about it and, in consequence, the explanation of the results would also have to change. In our approach we are considering classification tasks, therefore the user goals are always the same: to classify a new problem. This means that the explanation has to be convincing enough to justify the classification and we assume that the kind of explanation has always the same form, i.e. it does not change along the time. The explanation scheme we have introduced is also oriented to explain the result to the user. Nevertheless, these explanations could also be reused by the system as general domain knowledge as we proposed in [5].

7 Conclusions and Future Work

Lazy learning methods seem to be specially useful on domains such as toxicology, in which object domains are highly variable. Nevertheless, one of the main weakness of the lazy learning methods is how they justify the results to the user. In this paper we have proposed an explanation scheme that supports the user in comparing molecular structures of positive and negative compounds.

The application of that explanation scheme to explain the results of a lazy learning approach to predictive toxicology can be of high utility. Unlike induced leaning methods, a lazy learning method does not build a carcinogenecity model, therefore there is not a clear justification of the result. On the other hand, a chemist needs to focus on both similarities and differences among the molecular structure of chemical compounds. Using our approach, even if it focuses on similarities, the user can easily see the differences among carcinogenic and non carcinogenic compounds. Due to this fact, and because small differences on the molecular structure of compounds may give different carcinogenic activity, the user can revise literature supporting the classification given by the lazy learning method.

As future work we plan to assess the confidence degree of an explanation. This confidence could be assessed taking into account the entropy of the discriminatory set associated to a similitude term. In other words, since LID can finish with a similitude term D satisfied by cases of several classes, a discriminatory set S_D with high entropy means that D is too general, therefore the features included in D, even if considered as relevant, are not actually discriminant. This could be interpreted as a low confidence in the explanation of the classification. Conversely, a discriminatory set with low entropy means that the similitude term D is accurate, therefore the confidence on the classification would be high. This same criteria could be applied to assess the confidence of the parts AU^1 and AU^2 of the explanation scheme.

Acknowledgements. This work has been supported by the MCYT-FEDER Project MID-CBR (TIN2006-15140-C03-01). The author thanks Dr. Lluis Godo for his assistance in the improvement of this paper.

References

1. Ambs, S., Neumann, H.G.: Acute and chronic toxicity of aromatic amines studied in the isolated perfused rat liver. Toxicol. Applied Pharmacol. 139, 186–194 (1996)
2. Armengol, E., Plaza, E.: Bottom-up induction of feature terms. Machine Learning 41(1), 259–294 (2000)
3. Armengol, E., Plaza, E.: Lazy induction of descriptions for relational case-based learning. In: De Reaedt, L., Flach, P. (eds.) ECML-2001. LNCS (LNAI), vol. 2167, pp. 13–24. Springer, Heidelberg (2001)
4. Armengol, E., Plaza, E.: Relational case-based reasoning for carcinogenic activity prediction. Artificial Intelligence Review 20(1–2), 121–141 (2003)

5. Armengol, E., Plaza, E.: Remembering similitude terms in case-based reasoning. In: 3rd Int. Conf. on Machine Learning and Data Mining MLDM-03. LNCS (LNAI), vol. 2734, pp. 121–130. Springer, Heidelberg (2003)
6. Armengol, E., Plaza, E.: An ontological approach to represent molecular structure information. In: Oliveira, J.L., Maojo, V., Martín-Sánchez, F., Pereira, A.S. (eds.) ISBMDA 2005. LNCS (LNBI), vol. 3745, pp. 294–304. Springer, Heidelberg (2005)
7. Armengol, E., Plaza, E.: Using symbolic descriptions to explain similarity on case-based reasoning. In: Radeva, P., Lopez, B., Melendez, J., Vitria, J. (eds.) Artificial Intelligence Research and Develoment, IOS Press, Octubre (2005)
8. Bélanger, M., Martel, J.M.: An automated explanation approach for a decision support system based on mcda. In: Roth-Berghofer, T., Schulz, S. (eds.) Explanation-Aware Computing: Papers from the 2005 Fall Symposium. Technical Report FS-05-04, Menlo Park, California. American Association for Artificial Intelligence, pp. 21–34 (2005)
9. Blockeel, H., Driessens, K., Jacobs, N., Kosala, R., Raeymaekers, S., Ramon, J., Struyf, J., Van Laer, W., Verbaeten, S.: First order models for the predictive toxicology challenge 2001. In: Procs of the Predictive Toxicology Challenge Workshop (2001)
10. Cassens, J.: Knowing what to explain and when. In: Proceedings of the ECCBR 2004 Workshops. Technical Report 142-04. Departamento de Sistemas Informáticos y Programación, Universidad Complutense de Madrid, Madrid, Spain, pp. 97–104 (2004)
11. Deshpande, M., Karypis, G.: Automated approaches for classifying structures. In: Proc. of the 2nd Workshop on Data Mining in Bioinformatics (2002)
12. Gonzalez, J., Holder, L., Cook, D.: Application of graph-based concept learning to the predictive toxicology domain. In: Procs of the Predictive Toxicology Challenge Workshop, Freiburg, Germany (2001)
13. Helma, C., Kramer, S.: A survey of the predictive toxicology challenge 2000-2001. Bioinformatics, pp. 1179–1200 (2003)
14. Katritzky, A.R, Petrukhin, R., Yang, H., Karelson, M.: CODESSA PRO. User's Manual. University of Florida (2002)
15. Leake, D.B.: Issues in goal-driven explanation. In: Proceedings of the AAAI Spring symposium on goal-driven learning, pp. 72–79 (1994)
16. de Mántaras, R.L.: A distance-based attribute selection measure for decision tree induction. Machine Learning 6, 81–92 (1991)
17. McCarthy, K., Reilly, J., McGinty, L., Smyth, B.: Thinking positively - explanatory feedback for conversational recommender systems. In: Procs of the ECCBR 2004 Workshops. TR 142-04. Departamento de Sistemas Informáticos y Programación, Universidad Complutense de Madrid, pp. 115–124 (2004)
18. McSherry, D.: Explanation in recommendation systems. In: Procs of the ECCBR 2004 Workshops. TR 142-04. Departamento de Sistemas Informáticos y Programación, Universidad Complutense de Madrid, pp. 125–134 (2004)
19. Sorensen, R.U.: Allergenicity and toxicity of amines in foods. In: Proceedings of the IFT 2001 Annual Meeting, New Orleans, Louisiana (2001)
20. Sørmo, F., Cassens, J., Aamodt, A.: Explanation in Case-Based Reasoning – Perspectives and Goals. Artificial Intelligence Review 24(2), 109–143 (2005)