Incorporating Molecule's Stereisomerism within the Machine Learning Framework

Pierre-Anthony Grenier, Luc Brun, Didier Villemin

Incorporating molecule's stereisomerism within the machine learning framework

Pierre-Anthony Grenier†, Luc Brun†, and Didier Villemin‡

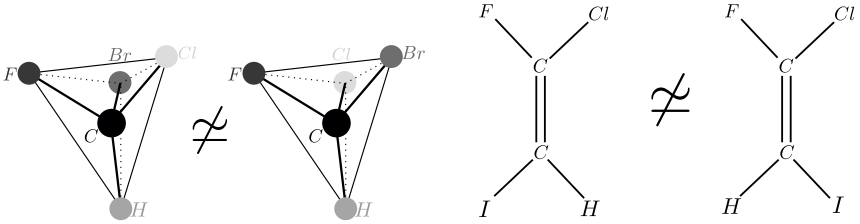
†GREYC UMR CNRS 6072, ‡LCMT UMR CNRS 6507,
Caen, France
{pierre-anthony.grenier,luc.brun,didier.villemin}@ensicaen.fr,

Abstract. An important field of chemoinformatics consists in the prediction of molecule's properties, and within this field, graph kernels constitute a powerful framework thanks to their ability to combine a natural encoding of molecules by graphs, with classical statistical tools. Unfortunately some molecules encoded by a same graph and differing only by the three dimensional orientation of their atoms in space have different properties. Such molecules are called stereoisomers. These latter properties can not be predicted by usual graph methods which do not encode stereoisomerism. In this paper we propose to encode the stereoisomerism property of each atom of a molecule by a local subgraph. A kernel between bags of such subgraphs provides a similarity measure incorporating stereoisomerism properties. We then propose two extensions of this kernel incorporating in each sub graph information about its surroundings.

1 Introduction

A molecular graph is a graph $G = (V, E, \mu, \nu)$, where each node $v \in V$ encodes an atom and each edge $e \in E$ a bond between two atoms. The labelling functions μ and ν associate to each vertex and each edge a label encoding respectively the nature of the atom (carbon, oxygen,...) and the type of the bond (single, double, triple or aromatic). However, those graphs have a limitation: they do not encode the spatial configuration of atoms. Some molecules, called stereoisomers, are associated to a same molecular graph but differ by the relative positioning of their atoms.

Most of stereoisomers are characterized by the three dimensional orientation of the direct neighbors of a single atom or two connected atoms. We can consider for example, a carbon atom, with four neighbors, each of them located on a summit of a tetrahedron. If we permute two of the atoms, we obtain a different spatial configuration and hence an alternative stereoisomer (Figure 1(a)). An atom is called a stereocenter if a permutation of two atoms belonging to its neighborhood produces a different stereoisomer. We should stress here that, to a large extend, stereoisomerism is independent of a particular embedding of a molecule. Indeed, in Figure 1(a), any particular embedding keeping the same relative positioning of atoms H, Cl, Br and F according to the central carbon atom C, would correspond to a same stereoisomer. In the same way, two connected atoms form



(a) Two different spatial configurations of the neighbors of a carbon

(b) Two different spatial configurations of two carbons linked by a double bond.

Fig. 1. Two types of stereocenters.

a stereocenter if a permutation of the positions of two atoms belonging to the union of their neighborhoods produces a different stereoisomer (Figure 1(b)). According to chemical experts [9], within molecules currently used in chemistry, 98% of stereocenters correspond either to carbons with four neighbors, called asymmetric carbon (Figure 1(a)) or to couples of two carbons adjacent through a double bond (Figure 1(b)). We thus restrict the present paper to such cases.

Graph kernels [10, 6], provide a measure of similarity between graphs. Under the assumption that a kernel k is symmetric and definite positive, the value $k(G, G')$, where G and G' encode two graphs, corresponds to a scalar product between two vectors $\Psi(G)$ and $\Psi(G')$ in an Hilbert space. This latter property allows us to combine graph kernels with usual machine learning methods such as SVM or kernel ridge regression by using the well known kernel trick, which consists in replacing the scalar product between $\Psi(G)$ and $\Psi(G')$ by $k(G, G')$ in these algorithms.

Up to now, only few methods have attempted to incorporate stereoisomerism within the graph kernel framework. Brown et al. [2] have proposed to incorporate this information through an extension of the tree-pattern kernel [10]. One drawback of this method is that, patterns which encode stereo information, and patterns which do not, are combined without any weighting in the final kernel value. So for a property only related to stereoisomerism, patterns that do not encode stereo information may be assimilated to noise which deteriorates the prediction. Grenier et al. [8] have introduced the minimal subtree which characterizes a stereocenter within an acyclic molecule. They also proposed a kernel based on this minimal subtree, which takes into account stereoisomerism. This kernel is however restricted to acyclic graphs.

Based on [8], we present in Section 2 an encoding of molecules distinguishing stereoisomers. Section 3 present the construction of a subgraph, which allows to characterizes locally a stereocenter. Then in Section 4, we use this subgraph to propose new graph kernels valid for cyclic as well as acyclic molecules, thus overcoming the main limitation of [8]. We finally present in Section 5 results obtained using those kernels and compare these results with state of the art methods.

2 Ordered Graph and Stereo Vertices

The spatial configuration of the neighbors of each atom may be encoded through an ordering of its neighborhood. For example, considering the left part of Figure 1(a), and looking at the central carbon from the hydrogen atom (H), the sequence of remaining neighbors of the carbon: Cl, Br and F may be considered as lying on a plane and are encountered clockwise. Thus, this spatial configuration is encoded by the sequence H, Cl, Br, F and the sequence H, Br, Cl, F encodes the second configuration.

In order to encode this information in a graph, we introduce the notion of ordered graph. An ordered graph $G = (V, E, \mu, \nu, ord)$ is a molecular graph $G_m = (V, E, \mu, \nu)$ together with a function $ord : V \to V^*$ which maps each vertex to an ordered list of its neighbors. Two ordered graphs G and G' are isomorphic ($G \underset{o}{\simeq} G'$) if there exists an isomorphism f between their respective molecular graphs G_m and G'_m such that $ord'(f(v)) = (f(v_1) \dots f(v_n))$ with $ord(v) = (v_1 \dots v_n)$ (where $N(v) = \{v_1, \dots, v_n\}$ denotes the neighborhood of v). In this case f is called an ordered isomorphism between G and G'.

However, different ordered graphs may encode a same molecule. In the example of the left part of Figure 1(a), if we look to the central carbon from a different neighbor, we can obtain a different sequence, for example F, Br, Cl, H, that represents the same configuration but now considered from the atom F. We thus have to define an equivalence relationship between ordered graphs, such that two ordered graphs are equivalent if they represent a same configuration.

To do so, we introduce the notion of re-ordering function σ, which associates to each vertex $v \in V$ of degree n a permutation $\sigma(v)$ on $\{1, \dots, n\}$, which allows to re-order its neighborhood. The graph with re-ordered neighborhoods $\sigma(G)$ is obtained by mapping for each vertex v its order $ord(v) = v_1 \dots . v_n$ onto the sequence $v_{\sigma(v)(1)} \dots . v_{\sigma(v)(n)}$ where $\sigma(v)$ is the permutation applied on v.

In order to define a permutation $\sigma(v)$ for each vertex of a graph, we first introduce the notion of potential asymmetric carbon which corresponds to a carbon with four neighbors. Such a vertex corresponds to a stereocenter if one permutation of two of its neighbors provides a different stereoisomer (Section 1). Permutations associated to a potential asymmetric carbon correspond to all even permutations of its four neighbors [11]. For a double bond between two carbons, permutations associated to each carbon of the double bound must have a same parity. Finally, for any vertex which does not correspond to a potential asymmetric carbon nor to a carbon of a double bond, we do not search to characterize its spatial configuration. So these vertices are associated to all possible permutations of their neighbors.

The set of re-ordering functions, transforming an ordered graph into another one representing the same configuration is called a valid family of re-ordering functions Σ [7]. We say that it exists an equivalent ordered isomorphism f between G and G' according to Σ if it exists $\sigma \in \Sigma$ such that f is an ordered isomorphism between $\sigma(G)$ and G' ($\sigma(G) \underset{o}{\simeq} G'$). The equivalent order relationship defines an equivalence relationship [7] and two different stereoisomers are

encoded by non equivalent ordered graphs. We denote by $\text{IsomEqOrd}(G, G')$ the set of equivalent ordered isomorphism between G and G'.

Potential asymmetric carbons, and double bonds between carbons, are not necessarily stereocenters. For example if the label of vertex Br of Figure 1(a) is replaced by Cl, both left and right molecules of Figure 1(a) would be identical. In the same way, if the label of the vertex F in Figure 1(b) is replaced by Cl, the left and right molecules of this figure also become identical. For those cases, any permutation in the ordered list of the carbons would lead to an equivalent ordered graph. We thus define a stereo vertex as a vertex for which any permutation of two of its neighbors produces a non-equivalent ordered graph:

Definition 1 (Stereo vertex)**.** Let $G = (V, E, \mu, \nu, ord)$ be an ordered graph. A vertex $v \in V$ is called a stereo vertex iff:

$$\forall (i,j) \ \in \{1, \ldots, |N(v)|\}^2, i \neq j, \nexists f \in \text{IsomEqOrd}(G, \tau^v_{i,j}(G)) \text{ with } f(v) = v. \tag{1}$$

where $\tau^v_{i,j}(G)$ corresponds to an ordered graph deduced from G by permuting nodes of index i and j in $ord(v)$.

3 Minimal Stereo SubGraph

Definition 1 is based on the whole graph G to test if a vertex v is a stereo vertex. However, given a stereo vertex s, one can observe that on some configurations, the removal of some vertices far from s should not change its stereo property. In order to obtain a more local characterization of a stereo vertex, we should thus determine a vertex induced subgraph H of G, including s, large enough to characterize the stereo property of s (i.e. $\forall (i,j) \in \{1, \ldots, |N(s)|\}^2, i \neq j, \nexists f \in \text{IsomEqOrd}(H, \tau^s_{i,j}(H))$ with $f(s) = s$), but sufficiently small to encode only the relevant information characterizing the stereo vertex s. Such a subgraph is called a minimal stereo subgraph of s.

We now present an heuristic, used to compute a minimal stereo subgraph of a stereo vertex. We focus our attention on asymmetric carbons. Let H be a subgraph of G containing a stereo vertex s corresponding to an asymmetric carbon. We say that the stereo property of s is not captured by H if (Definition 1):

$$\exists (i,j) \in \{1, \ldots, |N(s)|\}^2, i \neq j, \exists f \in \text{IsomEqOrd}(H, \tau^s_{i,j}(H)) \text{ with } f(s) = s \tag{2}$$

To define a minimal stereo subgraph of s, we consider a finite sequence $(H^k_s)^n_{k=1}$ of vertex induced subgraphs of G. The first element of this sequence H^1_s is the smaller vertex induced subgraph for which we can test (2) :

$$V(H^1_s) = \{s\} \cup N(s)$$

where $V(H^1_s)$ and $N(s)$ denote respectively the set of vertices of H^1_s and the set of neighbors of s in G.

If the current vertex induced subgraph H_s^k does not capture the stereo property of s, we know by (2), that it exists some isomorphisms f of equivalent ordered graphs between H_s^k and $\tau_{i,j}^s(H_s^k)$ with $i \neq j$ and $f(s) = s$. Let us consider such an isomorphism f. By definition of equivalent ordered isomorphism, it exists $\sigma \in \Sigma$ such that f is an ordered isomorphism between H_s^k and $\sigma\left(\tau_{i,j}^s(H_s^k)\right)$. By definition of ordered isomorphisms, and since $f(s) = s$, we have:

$$\forall l \in \{1, \ldots, |N(s)|\}, f(v_l) = v_{\sigma(s) \circ \tau_{i,j}^s(l)}.$$

with $ord(s) = v_1, \ldots, v_n$.

As $\sigma(s)$ is an even permutation, $\sigma(s) \circ \tau_{i,j}^s$ is an odd one. Hence it exists l in $\{1, \ldots, |N(s)|\}$ such that $l \neq \sigma(s) \circ \tau_{i,j}^s(l)$ and we have $f(v_l) \neq v_l$ and $f^{(2)}(v_l) \neq f(v_l)$. In other words, any equivalent ordered isomorphism corresponding to equation (2) maps at least two vertices in the neighborhood of s in H_s^k onto a different vertex in the same neighborhood. Let us denote by $\mathcal{E}_f^k$ the set of vertices of H_s^k connected to s by a path whose all vertices are mapped onto other vertices by f:

$$\mathcal{E}_f^k = \{v \in V(H_s^k) \mid \exists c = (v_0, \ldots, v_q) \in H_s^k \text{ with } v_0 = s \text{ and } v_q = v \; s.t. \; \forall r \in \{1, \ldots, q\}, f(v_r) \neq v_r\} \quad (3)$$

For any equivalent ordered isomorphism f satisfying (2), the set $\mathcal{E}_f^k$ is not empty since it contains at least 2 vertices. A vertex v belongs to $\mathcal{E}_f^k$ if neither its label nor its neighborhood in H_s^k allows to differentiate it from $f(v)$. The basic idea of our algorithm consists in enforcing constraints on each $v \in \mathcal{E}_f^k$ at iteration $k+1$ by adding to H_s^k the neighborhood in G of all vertices belonging to $\mathcal{E}_f^k$. This last set is denoted by $N(\mathcal{E}_f^k)$. The set of vertices of the vertex induced subgraph H_s^{k+1} is thus defined by:

$$V(H_s^{k+1}) = V(H_s^k) \cup \bigcup_{f \in \mathcal{F}_s^k} N(\mathcal{E}_f^k) \quad (4)$$

where $\mathcal{F}_s^k$ denotes all equivalent ordered isomorphisms satisfying (2).

Since $f \in \mathcal{F}_s^k$ implies that $\mathcal{E}_f^k$ is not empty, adding iteratively constraints on the existence of vertices in $\mathcal{E}_f^k$ removes f from $\mathcal{F}_s^k$. The algorithm stops when the set $\mathcal{F}_s^k$ becomes empty. Note that such a condition must be satisfied since s is a stereocenter and hence the whole molecule does not satisfy (2).

The intermediate vertex induced subgraphs found by our algorithm are illustrated in Fig. 2. Note that at iteration 2, it exists an equivalent ordered isomorphism $f \in \mathcal{F}_C^2$ mapping the path CCO (bottom right of the figure) onto the same path located on the top right part of Fig 2. In this case $\mathcal{E}_f^2$ contains the three carbons of these two paths and both oxygen atoms. The oxygen atoms belong to $\mathcal{E}_f^2$ since their neighborhoods in H_C^2 does not allow to differentiate them (Fig. 2). At iteration 3, the neighborhood in G of these oxygen atoms are added to H_C^3, hence adding N and Br which allow to differentiate both paths and thus removes the equivalent ordered isomorphism f from $\mathcal{F}_C^3$.

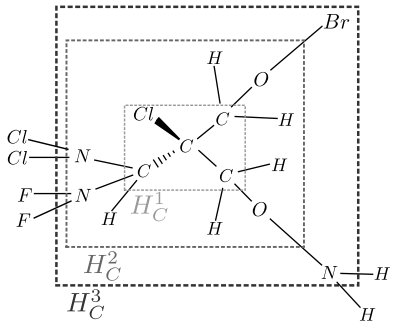


Fig. 2. An asymmetric carbon and its associated sequence $(H_C^k)_{k=1}^3$

4 Stereo Kernel and Extensions

4.1 Stereo Kernel

Given an ordered graph G, we can associate a minimal stereo subgraph to each of its stereocenter. A same stereo subgraph may be present more than once in a given molecule, we thus need to associate a unique code to each such subgraph in order to enumerate efficiently the eventual multiple occurences of a stereo subgraph within a molecule. To do so, we use [13], which associates to each molecule a unique code which allows to test the existence of an equivalent ordered isomorphism between two stereo subgraphs, unlike [1] which allows to find efficiently all isomorphisms between two graphs. We can thus compute the set of minimal stereo subgraphs $\mathcal{H}(G)$ together with the spectrum $spec(G) = (f_H(G))_{H\in\mathcal{H}(G)}$ which encodes the frequency $f_H(G)$ of each $H \in \mathcal{H}(G)$. The set $\mathcal{H}(G)$ and the spectrum *spec* provide a characterisation of each stereo center of G and hence describe the stereoisomerism of G.

The comparison of the spectrum of two ordered graphs, is then used to define a kernel between two molecules taking into account the stereoisomerism:

$$k(G, G') = \sum_{H\in\mathcal{H}(G)\cap\mathcal{H}(G')} K(f_H(G), f_H(G')). \tag{5}$$

where K denotes a kernel between real values (e.g. Gaussian, intersection or polynomial). The choice of a particular kernel, together with its parameters is performed through cross-validation.

4.2 Augmented labels

Cycles are important sub-parts of molecules, and thus two atoms with identical label could have different influence if one of them is included in a cycle. Thus it can be useful, during the computation of a minimal stereo subgraph, to consider two atoms with a same label, but not included in a same number of cycles, as different.

To do so, we first compute the set of relevant cycles of each ordered graph. Relevant cycles are defined as cycles of a graph which can not be deduced from shorter cycles [12]. We can associate to each vertex v, the number n_v of relevant cycles to which it belongs. Then, for an ordered graph G, this information is added to the label of each of its vertex v ($\mu_A(v) = \mu(v).n_v$) to obtain a new ordered graph G_A. The method described in Section 3 is then applied on the ordered graph G_A. We thus obtain a different set of minimal stereo subgraphs $\mathcal{H}(G_A)$, composed of smaller stereo subgraphs where nodes encode a more global information. As in Section 4.1 we define from this set of subgraphs a spectrum encoding the frequency of each subgraph, and define a kernel between graphs by comparing those spectrum:

$$k(G, G') = \sum_{H \in \mathcal{H}(G_A) \cap \mathcal{H}(G'_A)} K(f_H(G), f_H(G')). \tag{6}$$

4.3 Expanded Subgraphs

Equations 5 and 6 allow to compare two molecules through the distribution of their stereo subgraphs. However those kernels are based on a binary similarity measure between configurations: either two stereo subgraphs are different, and thus the configurations encoded by those subgraphs are dissimilar, or the subgraphs are identical and thus the configurations are similar. By adding information about the adjacency relationships between these stereo subgraphs and the remaining part of the molecule, we can obtain a finer measure of similarity between configurations around stereocenters.

To take into account the adjacency relationships between a stereo subgraph H_s and its surrounding, we consider larger vertex induced subgraphs than H_s. Let H be a subgraphs of G, the neighborhood $N(H)$ of H is the set of vertices of $G - H$ which are neighbors of a vertex of H:

$$N(H) = \{v \in V(G) - V(H) \mid \exists (u, v) \in E \ s.t \ u \in V(H)\}$$

The set of vertex induced subgraphs obtained by adding k of its neighbors to H_s can be used to construct a kernel between graphs. We can also consider subgraphs where vertex located farther from the stereo subgraph than its direct neighborhood are added. However we have to limit the number of vertices we add, in order to keep a local information. Moreover the number of subgraphs increases quickly with the number of added vertices. Indeed, C_N^k subgraphs can be constructed by adding k vertices of $N(H_s)$ to H_s, with $N = |N(H_s)|$. We thus, have to determine a number of vertex to add, which is large enough to characterize the adjacency relationships between a stereo subgraph and the remaining part of a molecule, but also sufficiently small to keep a local information. In our experiment, we have considered subgraphs obtained by adding up to three different neighbors of H_s and those obtained by adding one neighbor v of H_s, and one neighbor of v not included in the neighborhood of H_s.

For each minimal stereo subgraph, we have a set of subgraphs which encodes its adjacency relationships with other parts of the molecule. As in section 4.1, we

associate to those subgraphs a unique code. By adding those subgraphs to the set of minimal stereo subgraphs, we obtain a new set of subgraphs $\mathcal{H}^{\mathcal{E}}(G)$, for which we can associate a spectrum which encodes the frequency of each subgraphs $H \in \mathcal{H}^{\mathcal{E}}(G)$. We thus define a kernel between ordered graphs, which takes into account stereoisomerism, and the adjacency relationships of each stereo subgraphs with its surrounding:

$$k(G, G') = \sum_{H \in \mathcal{H}^{\mathcal{E}}(G) \cap \mathcal{H}^{\mathcal{E}}(G')} K(f_H(G), f_H(G')). \tag{7}$$

5 Experiments

Our first experiment is based on a dataset composed of all the stereoisomers of the perindoprilate [3]. As this molecule has 5 stereocenters, the dataset is composed of $2^5 = 32$ molecules. In this dataset, we try to predict if a molecule inhibit the angiotensin-converting enzyme (ACE). The dataset is split into a training set of 23 molecules, and a test set of 9 molecules, as in [3].

Table 1 shows the results obtained by our kernels and the adaptation of the tree pattern kernel to stereoisomerism [2]. All these kernels are combined with the standard SVM method [4] to classify molecules. As all molecules in the dataset are stereoisomers of each other, methods which do not include stereoisomerism information [10, 6] are unable to differentiate any molecule of this dataset and are consequently unable to predict the considered property. Moreover, information not related to stereoisomerism included in kernel [2] consists of the same patterns for all molecules. This leads to add a constant shift to all values of the kernel and hence does not deteriorate the prediction for this dataset. Two stereocenters of the molecules of the dataset have a same minimal stereo subgraph, however one of them contains vertices belonging to a cycle. The stereo kernel (Section 4.1) is not able to distinguish these stereocenters, and hence misclassified molecules containing these stereocenters. By using augmented labels (Section 4.2), these two stereocenters are distinguished, and this distinction allows us to reach a prediction accuracy of 100%. The expanded subgraph (Section 4.3) may also help to differentiate the two stereocenters, but for this dataset, one molecule of the testset remains misclassified due to the same stereocenters which are not sufficiently discriminated by this kernel.

Table 1. Classification of the ACE inhibitory activity of perindopirilates stereoisomers

Method	Accuracy Trainset %	Accuracy Testset %
Stereo Kernel (Section 4.1)	91.3	88.9
Stereo Kernel + Augmented Labels (Section 4.2)	100	100
Stereo Kernel + Expanded subgraph (Section 4.3)	100	88.9
Tree patterns Kernel with stereo information [2]	100	100

The second dataset is a dataset of synthetic vitamin D derivatives, used in [2]. This dataset is composed of 69 molecules containing cycles, with an average of 9 stereocenters per molecule. This dataset is associated to a regression problem, which consists in predicting the biological activities of each molecules. Each kernel is test by using it with the standard SVM regression method [5].

After normalizing the values of the dataset, the standard deviation of the biological activities is equal to 0.258. To choose the different parameters and estimate the performance of each kernel on this dataset we use a nested cross-validation. The outer cross-validation is a leave-one-out procedure, used to compute an error for each molecule of the dataset. For each fold, we use another leave-one-out procedure on the remaining molecules, to compute a validation error. Parameters which provide the lowest root mean squared error on the validation are selected. We obtain for each molecule an error, and report in Table 2, the mean of this distribution of errors together with the confidence interval at 95% of this distribution.

Greatest errors in Table 2 are obtained by methods [10, 6] which do not include stereo information. The adaptation of the tree pattern kernel to stereoisomerism [2] improves the results over the two previous methods hence showing the insight of adding stereoisomerism information. Our kernel with no extensions obtain results not as good as [2]. For this experiment the modification of label to incorporate information about cycles, decrease our results. However, the addition of information about relationships between minimal stereo subgraphs, and remaining part of the molecules, allow us to obtain better results than [2]. In this case the best results are obtained by considering only subgraphs including one neighbor of H_s. Our methods use a subgraph isomorphism algorithm, but the minimal stereo subgraph are small and thus we have small execution times as we can see in Table 2.

6 Conclusion

The study of stereoisomers constitutes an important subfield of chemistry and thus a major challenge in chemoinformatics. We have proposed in this paper, a graph kernel based on an explicit enumeration of all the stereo subgraphs of a molecule. Each stereo subgraph is associated to a stereo vertex and encodes the

Table 2. Prediction of the biological activity of synthetic vitamin D derivatives.

Method	Mean Error	Confidence interval	Gram's matrices computation (s)
Tree patterns kernel [10]	0.193	± 0.060	230
Treelet kernel [6]	0.207	± 0.064	7
Tree patterns kernel with stereo information [2]	0.138	± 0.043	230
Stereo kernel	0.141	± 0.047	1
Stereo kernel + Augmented Labels (Sec. 4.2)	0.192	± 0.061	3
Stereo kernel + Expanded subgraph (Sec. 4.3)	**0.122**	± 0.041	8

part of the graph which provides the stereo property to this vertex. Based on the notion of stereo subgraphs we propose to describe a molecule by its bag of stereo subgraphs. The similarity between two molecules is then encoded through a graph kernel based on the similarity of both bags. Moreover we propose two extensions of this kernel. One extension consists in adding to the labels of the graphs information about cycles of molecules. The second one consists in considering larger subgraphs encoding relationships between each stereo subgraph and the remaining part of the molecule. Experiments related to stereoisomerism properties demonstrate the relevance of our approach and of its extensions.

Acknowledgements

The authors wish to thank the association CRIHAN for their computing resources.

References

1. V. Bonnici, R. Giugno, A. Pulvirenti, D. Shasha, and A. Ferro. A subgraph isomorphism algorithm and its application to biochemical data. *BMC Bioinformatics*, 14(Suppl 7):S13, 2013.
2. J. Brown, T. Urata, T. Tamura, M. A. Arai, T. Kawabata, and T. Akutsu. Compound analysis via graph kernels incorporating chirality. *Journal of Bioinformatics and Computational Biology*, 8(1):63–81, 2010.
3. J. A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, and R. Rotondo. Atom-based stochastic and non-stochastic 3d-chiral bilinear indices and their applications to central chirality codification. *Journal of Molecular Graphics and Modelling*, 26(1):32–47, 2007.
4. C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
5. H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik. Support vector regression machines. In *NIPS*, pages 155–161, 1996.
6. B. Gaüzère, L. Brun, and D. Villemin. Two New Graphs Kernels in Chemoinformatics. *Pattern Recognition Letters*, 33(15):2038–2047, 2012.
7. P.-A. Grenier, L. Brun, and D. Villemin. Incorporating stereo information within the graph kernel framework. Technical report, CNRS UMR 6072 GREYC, 2013. http://hal.archives-ouvertes.fr/hal-00809066/.
8. P.-A. Grenier, L. Brun, and D. Villemin. Treelet kernel incorporating chiral information. In *GbR in Pattern Recognition*, pages 132–141. Springer, 2013.
9. J. Jacques, A. Collet, and S. Wilen. *Enantiomers, racemates, and resolutions*. Krieger Pub. Co., 1991.
10. P. Mahé and J.-P. Vert. Graph kernels based on tree patterns for molecules. *Machine Learning*, 75(1):3–35, Oct. 2008.
11. M. Petitjean. Chirality in metric spaces. *Symmetry, Culture and Science*, 21:27–36, 2010.
12. P. Vismara. Union of all the minimum cycle bases of a graph. *The Electronic Journal of Combinatorics*, 4(1):73–87, 1997.
13. W. T. Wipke and T. M. Dyott. Stereochemically unique naming algorithm. *Journal of the American Chemical Society*, 96(15):4834–4842, 1974.