HAL
open science

A phase transition-based perspective on multiple instance kernels

Romaric Gaudel, Michèle Sebag, Antoine Cornuéjols

To cite this version:

Romaric Gaudel, Michèle Sebag, Antoine Cornuéjols. A phase transition-based perspective on multiple instance kernels. Plate-forme AFIA 2007: Conférence francophone sur l'Apprentissage automatique, Jul 2007, Grenoble, France. 10.1007/978-3-540-78469-2_14 . hal-01197534

HAL Id: hal-01197534

https://hal.science/hal-01197534v1

Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Phase Transition-Based Perspective on Multiple Instance Kernels

Romaric Gaudel[1,2], Michèle Sebag[1], and Antoine Cornuéjols[3]

[1] CNRS − INRIA − Univ. Paris-Sud, F-91405 Orsay, France
{romaric.gaudel, michele.sebag}@lri.fr
[2] École Normale Supérieure de Cachan
[3] AgroParisTech − INRA, F-75005 Paris, France
antoine.cornuejols@agroparistech.fr

Abstract. This paper is concerned with Relational Support Vector Machines, at the intersection of Support Vector Machines (SVM) and Inductive Logic Programming or Relational Learning. The so-called phase transition framework, originally developed for constraint satisfaction problems, has been extended to relational learning and it has provided relevant insights into the limitations and difficulties thereof. The goal of this paper is to examine relational SVMs and specifically Multiple Instance (MI) Kernels along the phase transition framework. A relaxation of the MI-SVM problem formalized as a linear programming problem (LPP) is defined and we show that the LPP satisfiability rate induces a lower bound on the MI-SVM generalization error. An extensive experimental study shows the existence of a critical region, where both LPP unsatisfiability and MI-SVM error rates are high. An interpretation for these results is proposed.
Key words: Phase Transition, Multiple Instance Problems, Relational Learning, Relational Kernels, Support Vector Machines.

1 Introduction

This paper is concerned with Relational Support Vector Machines, at the intersection of Support Vector Machines (SVM) [20] and Inductive Logic Programming or Relational Learning [18]. After the so-called kernel trick, the extension of SVMs to relational representations relies on the design of specific kernels (see [8, 10] among many others). Relational kernels thus achieve a particular type of propositionalization [14], mapping every relational example onto a propositional space defined after the training examples. However, relational representations intrinsically embed combinatorial issues; for instance the Plotkin's θ-subsumption test used as relational covering test is equivalent to a Constraint Satisfaction Problem (CSP) [11]. The fact that relational learning involves the resolution of CSPs as a core routine has far-fetched consequences besides exponential (worst-case) complexity, referred to as the Phase Transition (PT) paradigm (more on this in section 2).

The question investigated in this paper is whether relational SVMs overcome the limitations of relational learners related to the PT [3]. Specifically, the

study focuses on the Multiple Instance (MI) setting [9], for which several SVM approaches have been proposed [10, 8, 16, 15]. This paper presents two contributions. Firstly, a relaxation of the MI-SVM problem is introduced and formalized as a Linear Programming Problem (LPP); we show that the LPP satisfiability rate derives a lower bound on the generalization error of the MI-SVM. Secondly, a principled experimental study is conducted, based on a set of order parameters; these experiments show the existence of a critical region, conditioned by the value of order parameters, where both LPP unsatisfiability and MI-SVM error rates are high.

The paper is organized as follows. For the sake of self-containedness, the Phase Transition framework is briefly introduced in Section 2 together with the Multiple Instance setting. Section 3 defines a relaxed formalization of the MI-SVM expressed as a LPP, and establishes a relation between the MI-SVM generalization error and the LPP satisfiability rate. Section 4 reports on the experimental study and discusses the results. The paper concludes with some perspectives for further research.

2 State of the Art

It is widely acknowledged that there is a huge gap between the empirical and the worst case complexity analysis for CSPs [4]. This remark led to developing the so-called *phase transition paradigm* (PT) [12], which considers the satisfiability and the resolution complexity of CSP instances as random variables depending on order parameters of the problem instance (e.g. constraint density and tightness).

The phase transition paradigm has been transported to relational machine learning and inductive logic programming (ILP) by [11], and was shown to be instrumental in discovering and analyzing some limitations of relational learning [3] or grammatical inference [19] algorithms, such as the existence of a failure region for existing relational learners [3].

Resuming the above studies, this paper investigates the PT phenomenon in the Multiple Instance Learning setting introduced by Dietterich et al. [9], which is viewed as intermediate between relational and propositional settings. Formally, a MI example $\mathbf{x}$ is a bag of (propositional) instances noted $x^{(1)}, \ldots, x^{(N)}$.

In the original MI setting, referred to in the following as *linear*, an example is labelled positive iff it includes at least one instance satisfying some target concept C:

$$pos(\mathbf{x}) \text{ iff } \exists\, i \in 1 \ldots N \ s.t.\ C(x^{(i)})$$

However, in some contexts such as image categorization, [5] pointed out that the example label might depend on the properties of several instances; along the same lines, several alternative formalizations were proposed by [21] and the remainder of the paper will consider the so-called *presence-based* setting, with:

$$pos(\mathbf{x}) \text{ iff } \forall\, j = 1 \ldots m,\ \exists\, i_j \in 1 \ldots N \ s.t.\ C_j(x^{(i_j)})$$

Many approaches have been developed to address MI problems, including specific algorithms focussing on *linear* MI [17, 22], relational algorithms [6, 2],

and specific Support Vector Machine (SVM) approaches [10, 8, 16, 15]. Assuming the reader's familiarity with SVMs [20] and restricting ourselves to standard bag kernels in this paper, MI-kernels K are constructed on the top of propositional kernels k. Formally, letting $\mathbf{x} = (x^{(1)}, \ldots x^{(N)})$ and $\mathbf{x}' = (x'^{(1)}, \ldots x'^{(N')})$ denote two examples, standard MI-kernels are defined as:

$$K(\mathbf{x}, \mathbf{x}') = f(\mathbf{x}).f(\mathbf{x}') \sum_{k=1}^{N} \sum_{\ell=1}^{N'} k(x^{(k)}, x'^{(\ell)}) \tag{1}$$

where $f(\mathbf{x})$ corresponds to a normalization term, e.g. $f(\mathbf{x}) = 1$ or $1/N$ or $1/\sqrt{K(\mathbf{x}, \mathbf{x})}$.

MI-SVMs have obtained good results on *linear* MI problems [10], and also in application domains which rather belong to the *presence-based* setting, such as image categorization [15] or chemometry [16].

Still, by construction standard MI-kernels consider the average similarity among the example instances. The question examined in this paper is to which extent this average information is sufficient to reconstruct existential concepts involved in *presence-based* MI problems.

3 Overview

This section introduces a relaxation of MI-SVM problems in terms of Linear Programming problems, which will be exploited to analyze MI-SVM along the phase transition framework.

3.1 When MI Learning meets Linear Programming

In order to investigate the performance of an algorithm within the PT framework, a standard procedure is to generate artificial problems after the selected order parameters (see below), where each problem is made of a training set $\mathcal{L} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_\ell, y_\ell)\}$ and a test set $\mathcal{T} = \{(\mathbf{x'}_1, y'_1), \ldots, (\mathbf{x'}_t, y'_t)\}$, and to compute the error on the test set of the hypothesis learned from the training set. The test error, averaged over a sample of artificial problems generated after some order parameter values, indeed measures the competence of the algorithm conditionally to these parameter values [3].

A different approach is followed in the present paper, for the following reason. Our goal is to examine how kernel tricks can be used to alleviate the specific difficulties of relational learning; in relational terms, the question is about the quality of the propositionalization achieved through relational kernels. In other words, the focus is on the competence of the representation (the capacity of the hypothesis search space defined after the MI kernel) as opposed to, the competence of a particular algorithm (the average quality of the hypotheses learned by this algorithm in this search space).

Accordingly, while the proposed methodology is still based on the generation of artificial problems, it focuses on the kernel-based propositionalization of the

MI examples. Formally, to each training set $\mathcal{L}$ is associated the propositional representation $\mathcal{R}_{\mathcal{L}}$, characterizing every MI example $\mathbf{x}$ as the ℓ-dimensional real-valued vector defined as $\Psi_{\mathcal{L}}(\mathbf{x}) = (K(\mathbf{x}_1, \mathbf{x}), \ldots, K(\mathbf{x}_\ell, \mathbf{x}))$.

By construction [20], any MI-SVM hypothesis h is expressed as a linear hypothesis in $\mathcal{R}_{\mathcal{L}}$, $h(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i.y_i.K(\mathbf{x}_i, \mathbf{x}) + \beta$, subject to ℓ inequality constraints:

$$\forall i = 1 \ldots \ell \qquad \alpha_i \geq 0 \tag{2}$$

Let $\mathcal{T}$ denote a t-example dataset propositionalized after $\mathcal{R}_{\mathcal{L}}$; the existence of a separating hyperplane for $\mathcal{T}$ is formalized as a set of t inequality constraints:

$$\forall j = 1 \ldots t \qquad y'_j.h(\mathbf{x}'_j) = y'_j.\left(\sum_{i=1}^{N} \alpha_i.y_i.K(\mathbf{x}_i, \mathbf{x}'_j) + \beta\right) \geq 1 \tag{3}$$

Let $\mathrm{Q}(\mathcal{L}, \mathcal{T})$ be defined as the set of inequality constraints (2) and (3). $\mathrm{Q}(\mathcal{L}, \mathcal{T})$ admits a solution iff the MI-SVM propositionalization defined from $\mathcal{L}$ has the capacity to separate the examples in $\mathcal{T}$. Note that the linear programming problem[4] (LPP) $\mathrm{Q}(\mathcal{L}, \mathcal{T})$ is much easier than the standard learning problem of whether the hypothesis actually learned from $\mathcal{L}$ will correctly classify $\mathcal{T}$. $\mathrm{Q}(\mathcal{L}, \mathcal{T})$ is an easier problem as it explicitly exploits the labels of the test examples (i.e., cheats) in order to find the $\ell + 1$ coefficients α_i and β; further, it can select *a posteriori* some of the SVM hyper-parameters, e.g. the error cost C.

The central argument of the paper is: $\mathrm{Q}(\mathcal{L}, \mathcal{T})$ gives deep insights into the quality of the propositionalization based on the kernel trick. Formally we show that the probability for $\mathrm{Q}(\mathcal{L}, \mathcal{T})$ to admit a solution, referred to as LPP satisfiability rate, induces a lower bound on the MI-SVM generalization error.

Proposition

Within a MI-SVM setting, let $\mathcal{L}$ be a training set of size ℓ, $\mathcal{R}_{\mathcal{L}}$ the associated kernel-based propositionalization, and $p_{\mathcal{L}}$ the generalization error of the optimal linear classifier $h^*_{\mathcal{L}}$ defined on $\mathcal{R}_{\mathcal{L}}$. Let $\mathbb{E}_\ell[p_{\mathcal{L}}]$ denote the expectation of $p_{\mathcal{L}}$ conditionally to $|\mathcal{L}| = \ell$.

Let a MI-SVM problem be defined as a pair of example sets $(\mathcal{L}, \mathcal{T})$. Considering a sequence of R independent MI-SVM problems $(\mathcal{L}_i, \mathcal{T}_i)$ such that the size of $\mathcal{L}_i$ (respectively $\mathcal{T}_i$) is ℓ (resp. t), let $\varepsilon_R(\ell, t)$ denote the fraction of LPPs $\mathrm{Q}(\mathcal{L}_i, \mathcal{T}_i)$ that are satisfiable. Then for any $\eta > 0$, with probability at least $1 - exp(-2\eta^2 R)$,

$$\mathbb{E}_\ell[p_{\mathcal{L}}] \geq 1 - (\varepsilon_R(\ell, t) + \eta)^{\frac{1}{t}}. \tag{4}$$

Proof

Given $\mathcal{L}$, $h^*_{\mathcal{L}}$ and $p_{\mathcal{L}}$ as above, the probability for a t example set $\mathcal{T}$ to include no example misclassified by $h^*_{\mathcal{L}}$ is $(1 - p_{\mathcal{L}})^t$.

[4] Actually, this problem should rather be viewed as a constraint satisfaction problem on continuous variables, as it does not involve any optimization objective; the only point is whether the set of linear inequalities admits a solution.

It is straightforward to see that if $\mathcal{T}$ does not contain examples that are misclassified by $h^*_{\mathcal{L}}$, $\mathrm{Q}(\mathcal{L},\mathcal{T})$ is satisfiable. Therefore the probability for $\mathrm{Q}(\mathcal{L},\mathcal{T})$ to be satisfiable conditionally to $\mathcal{L}$ is greater than $(1-p_{\mathcal{L}})^t$:

$$\mathbb{E}_{|\mathcal{T}|=t}[\ \mathrm{Q}(\mathcal{L},\mathcal{T}) \text{ satisfiable}] \geq (1-p_{\mathcal{L}})^t$$

Taking the expectation of the above w.r.t. $|\mathcal{L}|=\ell$, it comes:

$$\mathbb{E}_{|\mathcal{T}|=t,\ |\mathcal{L}|=\ell}[\ \mathrm{Q}(\mathcal{L},\mathcal{T}) \text{ satisfiable}] \geq \mathbb{E}_{|\mathcal{L}|=\ell}[(1-p_{\mathcal{L}})^t] \geq (1-\mathbb{E}_\ell[p_{\mathcal{L}}])^t \qquad (5)$$

where the right inequality follows from Jensen's inequality. Next step is to bound the left term from its empirical estimate $\varepsilon_R(\ell,t)$, using Hoeffding's bound. With probability at least $1-exp(-2\eta^2 R)$,

$$\mathbb{E}_{|\mathcal{T}|=t,\ |\mathcal{L}|=\ell}[\ \mathrm{Q}(\mathcal{L},\mathcal{T}) \text{ satisfiable}] < \varepsilon_R(\ell,t)+\eta \qquad (6)$$

From (5) and (6) it comes that with probability at least $1-exp(-2\eta^2 R)$

$$(1-\mathbb{E}_\ell[p_{\mathcal{L}}])^t \leq \varepsilon_R(\ell,t)+\eta$$

which concludes the proof. □

This theoretical result allows us to draw conclusions about the quality (generalization error) of the MI-SVM framework, based on the experimental satisfiability rate of the linear programming problem $\mathrm{Q}(\mathcal{L},\mathcal{T})$.

3.2 Order Parameters and Experimental Setting

The satisfiability of $\mathrm{Q}(\mathcal{L},\mathcal{T})$ is systematically investigated following the PT paradigm [3], based on the definition of order parameters. These order parameters, summarized in Table 1 together with their range of variation in the experiments, intend to characterize the key complexity factors in a MI-SVM problem, related to the instances, the examples, and the target concept.

d	Dimension of the instance space $\mathcal{X}=[0,1]^d$	30
m	Number of sub-concepts in the target concept	30
ε	Coverage of a sub-concept $=\varepsilon^d$	.15
ℓ	Number of training examples	60 (30 +, 30 −)
t	Number of test examples	200 (100 +, 100 −)
N, N'	Number of instances in pos./neg. example	100
n	Number of relevant instances per positive example	30...100
n'	Number of relevant instances per negative example	0...100
nm	Number of sub-concepts not satisfied by neg. examples	10,20,25

Table 1. Order parameters for the MI LPP, and range of variation in the experiments

Instance space $\mathcal{X}$ is set to $[0,1]^d$; unless specified otherwise, any instance x is uniformly drawn in $\mathcal{X}$. We denote $\mathcal{B}_\varepsilon(x)$ the ε-radius ball centered on x w.r.t. L_∞

norm. The target concept involves m sub-concepts C_i ; $C_i(x)$ holds iff x belongs to $B_\varepsilon(z_i)$, where z_i is a uniformly drawn instance. For $m > 1$ (resp. $m = 1$) such a target concept follows the *presence-based* (resp. *linear*) MI setting (section 2), Positive (respectively negative) examples include N (resp. N') instances. An instance is said to be *relevant* if it satisfies some sub-concept. An example is said to satisfy a sub-concept if it includes an instance satisfying this sub-concept. Positive (respectively negative) examples involve n (resp. n') relevant instances. Any negative example fails to satisfy exactly nm (for near-miss) sub-concepts. Naturally, $n \geq m$ and $nm \geq 1$.

For each order parameter setting, 40 pairs (training set $\mathcal{L}$, test set $\mathcal{T}$) are built, made of an equal number of positive and negative iid examples; each example involves the required number of relevant instances, uniformly drawn in some $B_\varepsilon(z_i)$, and other instances uniformly drawn in $\mathcal{X}$, conditionally to parameters N and n for positive examples (resp., N', n' and nm for negative examples). Set $\mathcal{T}$ is propositionalized after $\mathcal{R}_\mathcal{L}$, using Gaussian instance kernels with parameter $\sigma = 1$; the bag kernel uses the number of example instances as normalising function (eq. 1).

3.3 Goal of the Experiments

The paper goal is to see whether the MI-SVM framework overcomes the specific difficulties of relational learning, and whether a phase transition phenomenon occurs. The first goal of the experiments is to assess the satisfiability of the LPP; it is expected that the problem is satisfiable, i.e. positive and negative test examples can be discriminated, as far as their number of relevant instances are sufficiently different ($n <> n'$); the question thus is whether the diagonal region $n = n'$ is a critical region, and if it is the case, what its width is. This goal is achieved by measuring the LPP satisfiability, averaged over 40 problems $(\mathcal{L}_i, \mathcal{T}_i)$ independently generated for each order parameter setting.

The second goal is to assess the actual relation between the LPP satisfiability and the MI-SVM generalization error, in other words the relevance of the proposed approach. Indeed the lower bound on the MI-SVM generalization error based on the satisfiability does not say much as only $R = 40$ problems are considered per order parameter setting for computational feasibility. It thus remains to see whether the critical LPP region is also critical from a MI-SVM point of view, i.e. if it is a region where the standard test error is high too. This goal is classically achieved by learning a MI-SVM hypothesis from $\mathcal{L}_i$, measuring its error on $\mathcal{T}_i$, and averaging the test error over all problems generated for each order parameter setting.

4 Experiments

This section reports on the extensive experimental study conducted after the order parameters (Table 1). In total, 30,000 artificial MI-SVM problems have been considered. Let us first summarize the lessons learned before detailing and discussing the results.

4.1 Summary of the results

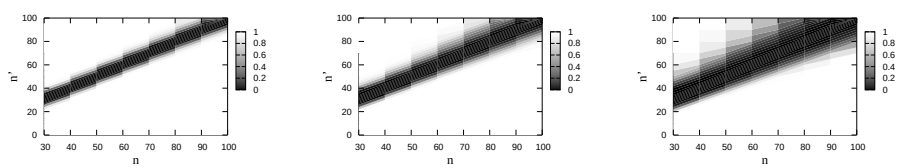
Firstly, the existence of an unsatisfiable region is experimentally demonstrated (Fig. 1). As expected, the unsatisfiable region corresponds to "truly relational" problems, e.g. when no distinction can be made between positive and negative examples based on their number of relevant instances ($n' = n$). Surprisingly, the width of the unsatisfiable region increases as parameter nm increases, i.e. when few sub-concepts are satisfied by a negative example. An interpretation for these findings is proposed in section 4.2.

Secondly, the unsatisfiable region is also a critical region from a MI-SVM learning viewpoint, which confirms the practical relevance of the lower bound established in section 3.1. The learning accuracy decreases smoothly but significantly while the satisfiability rate abruptly goes to 0 (Fig. 3); in the unsatisfiable region, the average test error is circa 40%.

4.2 LPP Satisfiability Landscape

Each LPP has been solved using the GGLPK package, with an average resolution cost of 16 seconds (on PC Pentium IV, 3.0 Ghz).

The average satisfiability computed for each order parameter setting mostly depends on the number n and n' of relevant instances in positive and negative examples. For the sake of readability, the satisfiability is thus graphically displayed in the (n, n') plane; the color of pixel (x, y) is black (respectively white) if all LPP with $(n = x, n' = y)$ are unsatisfiable (resp. satisfiable). Fig. 1 shows the unsatisfiable black region, centered on the diagonal $n = n'$.



(a) $nm = 10$ (b) $nm = 20$ (c) $nm = 25$

Fig. 1. LPP satisfiability versus n and n', averaged over 40 runs, for various values of the number nm of sub-concepts not satisfied by a negative example. All other order parameter values are as in Table 1.

These results are explained from the distribution of the examples in the kernel-based propositional space. Fig. 2 illustrates this distribution in a propositionalized plane where the two attributes are derived from a positive and a negative training example. Let the instance kernel be the Gaussian kernel[5]. Let

[5] The interpretation only considers the Gaussian case; however complementary experiments done with polynomial kernels lead to similar LPP unsatisfiability landscape.

$\bar{k}_C$ and $\bar{k}_U$ respectively denote the expectation of $k(x, x')$ for two instances satisfying the same sub-concept C (resp., uniformly drawn). Considering MI examples $(\mathbf{x}, y)$ and $(\mathbf{x}', y')$, the expectation of $K(\mathbf{x}, \mathbf{x}')$ is thus analytically derived:

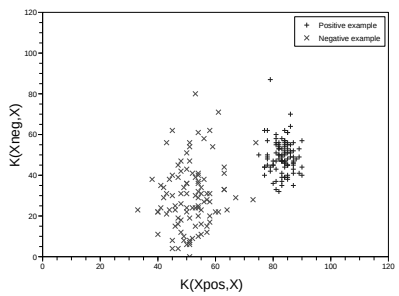


Fig. 2. Distribution of kernel-based propositionalized examples (legend + for positive, $\times$ for negative), with $n = 50,\ n' = 30,\ nm = 10$. First (second) coordinate corresponds to $K(\mathbf{x}, \cdot)$ with $\mathbf{x}$ a positive (negative) training example.

$$\mathbb{E}[K(\mathbf{x}, \mathbf{x}')] = \begin{cases} \frac{1}{m}(\frac{n}{N})^2(\bar{k}_C - \bar{k}_U) + \bar{k}_U & \text{if} \quad y = y' = 1 \\ \frac{1}{m}(\frac{n'}{N'})^2(\bar{k}_C - \bar{k}_U) + \bar{k}_U & \text{if} \quad y = y' = -1 \\ \frac{1}{m}\frac{n}{N}\frac{n'}{N'}(\bar{k}_C - \bar{k}_U) + \bar{k}_C & \text{if} \quad y \neq y' \end{cases}$$

Therefore in the neighborhood of the diagonal region[6] $n = n'$, the distribution of the propositionalized examples hardly depends on their class, adversely affecting the discrimination task.

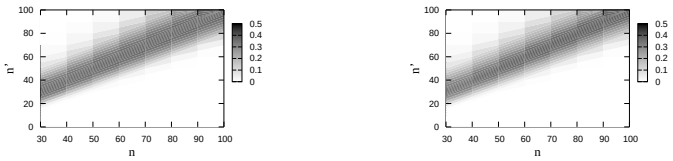
The fact that the width of the unsatisfiable region increases with the number nm of sub-concepts that are not satisfied by negative examples can be explained along the same lines. As nm increases, so does the variance of the distribution of the propositionalized negative examples, thus increasing the overlap between the distribution of positive and negative examples.

4.3 Generalization Error Landscape

As already mentioned, the lower bound given in section 3.1 is poorly informative with respect to the generalization error; an unsatisfiability rate of 100 % over 40 problems only allows us to conclude that the generalization error is greater than 0.8 % with confidence 95%. To estimate the tightness of the bound, the actual generalization error was thus estimated empirically by learning from the training set and measuring the error on the test set, averaged over all problems generated for each order parameter setting. Each MI-SVM problem was solved

[6] Actually, the unsatisfiable region corresponds to $\frac{n}{N} = \frac{n'}{N'}$. For simplicity, the distinction is omitted in the paper as $N = N'$.

using SVMTorch [7] with an average computational cost of 25 seconds (on PC Pentium IV, 3.0 Ghz). For the sake of readability, the error is graphically displayed in the (n, n') plane; the color of pixel (x, y) depicts the average error for $(n = x, n' = y)$; a white pixel stands for no error while a black pixel stands for 50% error (same as random guessing).



(a) $C = 100$ (b) $C = 1{,}000{,}000$

Fig. 3. Generalization error of MI-SVM in the (n, n') plane, estimated from SVMTorch test error averaged on 40 problems, for cost error $C = 10^2$ and 10^6.

Indeed the SVMTorch parameters were not optimized for each problem. Still, experiments done with the cost error C ranging in $10, \ldots, 10^6$ lead to the same general picture, and confirm that the MI-SVM error increases with the LPP unsatisfiability (Fig. 3). While the unsatisfiability rate abruptly goes to 100%, the error rate increases more gently, but significantly; when the unsatisfiability is above 80% the average test error is above 30%.

5 Conclusion and Perspectives

The contribution of the paper is twofold. Firstly, a relaxed formalization of kernel-based learning in terms of linear programming has been defined, and it has been shown that the LPP satisfiability rate induces a lower bound on the generalization error. Contrasting with the mainstream asymptotic framework [20], the presented analysis is relevant for small size datasets, which makes sense indeed in application domains such as chemometry [16].

Secondly, the LPP framework has been used to demonstrate the existence of a phase transition phenomenon for standard MI-SVM kernels; further, the LPP unsatisfiable region corresponds to a critical region from a MI-SVM learning standpoint, where the test error is consistently greater than 30% after an extensive empirical study on artificial problems.

Further research will consider more sophisticated MI-SVM approaches [1, 8], and see whether they also present a phase transition phenomenon in relation with the specific difficulties of *presence-based* MI learning. Another direction perspective is to further investigate the LPP framework, using the satisfiability rate as a criterion for kernel selection, or active learning.

Acknowledgments

The authors thank Olivier Teytaud for fruitful discussions, and gratefully acknowledge the support of the Network of Excellence PASCAL, IST-2002-506778.

References

1. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support Vector Machines for Multiple-Instance Learning. NIPS (2002) 561–568
2. Blockeel, H., Page, D., Srinivasan, A.: Multi-Instance Tree Learning. ICML (2005) 57–64
3. Botta, M., Giordana, A., Saitta, L., Sebag, M.: Relational Learning as Search in a Critical Region. Journal of Machine Learning Research **4** (2003) 431–463
4. Cheeseman, P., Kanefsky, B., Taylor, W.: Where the Really Hard Problems are. IJCAI (1991) 331–337
5. Chen, Y., Wang, J.Z.: Image Categorization by Learning and Reasoning with Regions. Journal of Machine Learning Research **5** (2004) 913–939
6. Chevaleyre, Y., Zucker, J.-D.: Solving Multiple-Instance and Multiple-Part Learning Problems with Decision Trees and Rule Sets. Application to the Mutagenesis Problem. Canadian Conference on Artificial Intelligence (2001) 204–214
7. Collobert, R., Bengio, S., Mariéthoz, J.: Torch: a Modular Machine Learning Software Library. Technical Report IDIAP-RR 02-46 (2002)
8. Cuturi, M., Vert, J.-P.: Semigroup Kernels on Finite Sets. NIPS (2004) 329–336
9. Dietterich, T., Lathrop, R., Lozano-Perez, T.: Solving the Multiple-Instance Problem with Axis-Parallel Rectangles. Artificial Intelligence **89** (1-2) (1997) 31–71
10. Gärtner, T., Flach, P.A., Kowalczyk, A., Smola, A. J.: Multi-Instance Kernels. ICML (2002) 179–186
11. Giordana, A., Saitta, L.: Phase Transitions in Relational Learning. Machine Learning **41** (2000) 217–251
12. Hogg, T., Huberman, B.A., C., Williams, C.P.: Phase Transitions and the Search Problem. Artificial intelligence **81** (1-2) (1996) 1–15
13. Kearns, M., Li, M.: Learning in the Presence of Malicious Errors. SIAM J. Comput. **22** (1993) 807–837
14. Kramer, S., Lavrac, N., Flach, P.: Propositionalization Approaches to Relational Data Mining. S. Dzeroski, N. Lavrac (eds.): Relational data mining (2001) 262–291
15. Kwok, J., Cheung, P.-M.: Marginalized Multi-Instance Kernels. IJCAI (2007) 901–906
16. Mahé, P., Ralaivola, L., Stoven, V., Vert, J.-P.: The Pharmacophore Kernel for Virtual Screening with Support Vector Machines. Journal of Chemical Information and Modeling **46** (2006) 2003–2014
17. Maron, O., Lozano-Pérez, T.: A Framework for Multiple-Instance Learning. NIPS (1997) 570–576
18. Muggleton, S., De Raedt, L.: Inductive Logic Programming: Theory and Methods. Journal of Logic Programming **19** (1994) 629–679
19. Pernot, N., Cornuéjols, A., Sebag, M.: Phase Transitions Within Grammatical Inference. IJCAI (2005) 811–816
20. Vapnik, V.N.: Statistical Learning Theory. Wiley-Interscience (1998)
21. Weidmann, N., Frank, E., Pfahringer, B.: A Two level Learning Method for Generalized Multi-Instance Problems. ECML (2003) 468–479
22. Zhang, Q., Goldman, S.A.: EM-DD: A Improved Multiple-Instance Learning Technique. NIPS (2001) 1073–1080