Learning Constraints and Optimization Criteria

Samuel Kolb
KU Leuven, Leuven, Belgium
samuel.kolb@cs.kuleuven.be

Abstract

While there exist several approaches in the constraint programming community to learn a constraint theory, few of them have considered the learning of constraint optimization problems. To alleviate this situation, we introduce an initial approach to learning first-order weighted MAX-SAT theories. It employs inductive logic programming techniques to learn a set of first-order clauses and then uses preference learning techniques to learn the weights of the clauses. In order to learn these weighted clauses, the clausal optimization system uses examples of possible worlds and a set of preferences that state which examples are preferred over other ones. The technique is also empirically evaluated on a number of examples.

Introduction

An important class of declarative models is concerned with (discrete) constrained optimization problems. These combine constraint satisfaction problems with an optimization function that specifies which solutions are optimal. While these models are powerful and elegant, they are often also hard to obtain, certainly for non-experts (Puget 2004). This has motivated several researchers to investigate whether such models can be learned.

Learning constraints has been addressed in constraint programming (e.g. (Beldiceanu and Simonis 2012), (Bessiere et al. 2013)) and inductive logic programming (e.g. (De Raedt and Dehaspe 1997)). Recent efforts (Lallouet et al. 2010) have attempted to combine these two fields. Furthermore, (Campigotto, Passerini, and Battiti 2011)) have learned propositional weighted MAX-SAT theories from examples. Also, work in statistical relational learning (SRL) on learning the structure of Markov Logic networks (MLN) or other SRL models (Kok and Domingos 2005) is relevant, as MAP inference can be used to query for the most likely state.

We use weighted first-order logical theories to represent constraint optimization problems. The weights capture the optimization function, and (first-order) weighted MAX-SAT can be used to determine the *best* solutions. Furthermore, we learn such models by combining inductive logic programming (ILP) principles (such as clausal discovery (De Raedt and Dehaspe 1997)) with preference learning. We use ILP to learn the clausal theories, and preference learning to determine the weights. Therefore, users provide examples of possible worlds and a set of preferences that specify which examples are preferred over other ones. This setting extends traditional constraint learning and ILP approaches with the ability to learn soft constraints and integrates constraint satisfaction and optimization. It also extends the approach of (Campigotto, Passerini, and Battiti 2011) as first-order theories are learned and preferences are used rather than examples with target values. Moreover, it differs from the SRL and MLN approaches in that one learns from preferences rather than from examples.

This paper is organized as follows. Section 2 reviews background and related work. In Section 3 we introduce the formalism and the setting, in Section 4 the algorithm and the clausal optimization system to solve this problem, and in Section 5 we report on some experiments. Finally, Section 6 concludes this paper.

Background

In this section we provide a brief overview of the important concepts and relevant literature.

Clauses We use (function free) first-order clauses; which we now introduce. An atom $p(t_1, t_2, ..., t_n)$ consists of a predicate symbol p that is applied to n terms t_i. Terms are either constants or variables. In this paper constants will be written in uppercase ($John$) and variables in lowercase (x). A literal can be an atom a (positive literal) or a negated atom $\neg a$ (negative literal). Clauses are disjunctions of literals $l_1 \vee l_2 \vee ... \vee l_k$ and are assumed to be universally quantified. A clause can be expressed using a body and a head $head \leftarrow body$. Hereby, positive literals are grouped in the head and negative literals are grouped in the body. Sometimes clauses are also written as sets of literals $\{l_1, l_2, ..., l_k\}$.

Atoms are grounded if they contain no variables. A Herbrand interpretation is a set of ground atoms. All atoms in the interpretation are assumed to be true and

all other possible ground atoms are assumed to be false. $\mathcal{I} \models c$ is used to denote that interpretation $\mathcal{I}$ satisfies the clause c.

Optimization The Boolean Satisfiability Problem (SAT) attempts to determine whether a propositional formula can be satisfied. Formulas can be rewritten into Conjunctive Normal Form (CNF), which consists of a conjunction of clauses (disjunctions). An extension of this problem is the Maximum Satisfiability Problem (MAX-SAT), which is an optimization problem and attempts to maximize the number of satisfied clauses. By assigning positive integer weights to clauses and maximizing the sum of the weights of satisfied clauses, one obtains the weighted MAX-SAT problem.

In related work (Campigotto, Passerini, and Battiti 2011) weighted MAX-SAT problems are induced by learning propositional clauses and weights using examples and absolute scores. (Change this section) Inspired by their results, our research tries to learn first-order logic clauses and weights using examples and rankings. The resulting optimization problem can be seen as an extension of weighted MAX-SAT in which the clauses are in first-order logic.

Learning constraints Constraint learning attempts to automatically identify constraints in one or multiple examples. This task is hard because the search space of constraints is typically very large and, unlike in typical machine learning problems, there are usually only few examples to learn from.

There are different approaches to accomplishing this task. The problem of having little examples can be alleviated by interactively generating data. Both the systems Conacq2 (Bessiere et al. 2007) en QuAcq (Bessiere et al. 2013) generate (partial) examples and query the user about their validity.

ModelSeeker (Beldiceanu and Simonis 2012) structures the problem variables in different ways. The system then uses a large catalog of global constraints to find the ones that hold within this structure. This approach allows for constraint learning from a few examples.

As mentioned earlier, the research of (Lallouet et al. 2010) used ILP techniques to learn constraints. In their approach, the search space of constraints is explored in a bidirectional manner, primarily using negative examples.

Clausal Discovery (Dissolve) The clausal discovery algorithm (De Raedt and Dehaspe 1997) attempts to learn first-order logic clauses from positive examples. Using a refinement operator, the search space is traversed by incrementally generalizing clauses until they cover all examples, starting from the empty clause ($\square$). The result set consists of the most specific clauses covering all examples.

Problem Statement

We want to learn constraints and optimization criteria based on positive examples and user preferences. The goal of this task is to simplify the process of generating formal representations, therefore, any input should be reasonably easy for a user to provide.

Constraint learning

Given a set of examples and a limit t, the goal is to find maximally specific clauses that are satisfied by at least t of the given examples.

Examples are Herbrand interpretations. They consist of set of constants, the domain, and all ground atoms over these constants that are true. The learned clauses are domain-independent and contain only variables.

Learning optimization criteria

Given a set of examples $\mathcal{E}$ and a set of preferences $\mathcal{P}$ over these examples, the goal is to learn soft constraints and weights such that the order described by these optimization criteria maximally corresponds with $\mathcal{P}$.

Every preference $p \in \mathcal{P}$ describes a relative order $e_1 \succ e_2 \succ ... \succ e_n$ over a subset of examples $\{e_1, e_2, ..., e_n\} \subseteq \mathcal{E}$. A preference $e_1 \succ e_2$ indicates that out of examples e_1 and e_2 the user prefers e_1.

The soft constraints consist of clauses $\mathcal{C}$ that are satisfied by some but not all of the examples. They can be identified using a system that implements the constraint learning task. The weighted clauses $\mathcal{W}$ are tuples $\mathbb{R} \times \mathcal{C}$ that can be used to calculate a score for any example e:

$$score(e, \mathcal{W}) = \sum_{(w,c) \in \mathcal{W}} w \cdot \mathbf{1}_{e \models c} \tag{1}$$

Example 1 (Moving)**.** Consider the scenario of moving to a new city and choosing areas for housing, work and school. The choices are captured by the predicates *live_in*, *work_in* and *school_in*. Additionally, the predicates *cheap* and *low_crime* are used to designate cheap and low crime areas, respectively.

Assume the following first-order weighted MAX-SAT model $\mathcal{M}$ describes the optimization problem to solve:

$$(w_1, c_1) = (0.50, low\_crime(a) \leftarrow live\_in(a))$$
$$(w_2, c_2) = (0.25, school\_in(a) \leftarrow work\_in(a))$$
$$(w_3, c_3) = (1.00, low\_crime(a) \leftarrow school\_in(a))$$
$$(w_4, c_4) = (-1.00, false \leftarrow live\_in(a) \wedge cheap(a))$$

Let examples e_1, e_2 and e_3 be interpretations that reason over the same city. Specifically, let all examples share constants (areas) A_1, A_2 and A_3 and ground atoms $cheap(A_1)$, $cheap(A_3)$, $low\_crime(A_1)$ and $low\_crime(A_2)$. Additionally, they specify:

$$e_1 : live\_in(A_1), work\_in(A_2), school\_in(A_2)$$
$$e_2 : live\_in(A_2), work\_in(A_2), school\_in(A_3)$$
$$e_3 : live\_in(A_2), work\_in(A_2), school\_in(A_1)$$

The score of an example for the model $\mathcal{M}$ can be calculated using (1). Example e_1 is covered by clauses c_1, c_2 and c_3, therefore, $score(e_1, \mathcal{M}) = 0.5+0.25+1+0 = 1.75$. Similarly, $score(e_2, \mathcal{M}) = -1$ and $score(e_3, \mathcal{M}) = 0$ can be computer. Based on these scores, the model $\mathcal{M}$ would prefer e_1 over e_2 ($e_1 \succ e_2$), e_1 over e_3 ($e_1 \succ e_3$) and e_2 over e_3 ($e_2 \succ e_3$). The aim of our research is to learn a model like $\mathcal{M}$, given such examples and preferences.

Apart from learning optimization criteria, this paper presents an approach to use these optimization criteria to find an optimal solution in practice.

Approach

In this work, we use a two-step approach to learn first-order weighted MAX-SAT. First, clause learning is used to find soft constraints in the given examples. Then, user-provided preference information is used to learn weights for these constraints.

Input

(Dissolve, partially or entirely) The input consists of global definitions that specify the types and predicates that are used to describe examples and a set of examples $\mathcal{E}$. (Explain: what is type information?) Type information relates directly to the problem domain and is usually easy to provide. Including types also improves the accuracy and efficiency of the learned clauses.

Examples are interpretations, they specify the constants in their domain and all predicates that are true. Predicates can also be generated by background knowledge, in which case they are not included in the example.

When learning optimization criteria, preference information is also provided by the user. Each preference is of the form $e_1 \succ e_2 \succ ... \succ e_n$ with $e_i \in \mathcal{E}$. Preferences describe the relative position of a few examples at a time and are typically easier to provide than absolute scores for all examples.

Example 2. (Rewrite) Consider the moving problem (example 2). A model could, for example, use a type *Area* and predicate definitions low_crime(*Area*), cheap(*Area*), school_in(*Area*), work_in(*Area*) and live_in(*Area*). Examples would then specify various areas A_i and describe the conditions (e.g. cheap(A_1), low_crime(A_2)) as well as the users choices (e.g. school_in(A_2), work_in(A_2), live_in(A_1)).

Clausal Discovery

The clause learning system is based on clausal discovery (De Raedt and Dehaspe 1997). Algorithm 1 illustrates how clausal discovery works on a high level. Note the double use of the $\models$ operator.(Review this once yet) On the one hand, it is used in $e \models c$ to denote that clause c is true in the example (interpretation) e and c is said to cover e. This test is used to compute if a clause covers enough examples. On the other hand, in $\mathcal{T} \models c$ it signifies that $\mathcal{T}$ logically entails c. The algorithm uses this test to remove redundant clauses. For both cases the IDP knowledge base system (De Pooter, Wittocx, and Denecker 2013; Wittocx, Mariën, and Denecker 2008) is used to compute $\models$. The algorithm can find both soft constraints ($t < \#\mathcal{E}$) or hard constraints ($t = \#\mathcal{E}$).

Algorithm 1 The clausal discovery algorithm

```
Given: Examples E and threshold t
Q ← {□}, T ← {}
while #Q > 0 do
    c ← next(Q)
    if #{e ∈ E | e ⊨ c} ≥ t then
        if ¬(T ⊨ c) then
            T = T ∪ c
    else
        Q ← Q ∪ ρ(c)
T ← prune(T)
return T
```

Starting with the empty clause (□), clauses are refined until they cover enough examples. If a clause covers enough examples it is added to the result set, provided it is not entailed by other clauses already in the result set. In the refinement step, a clause $c = \{l_1, l_2, ..., l_k\}$ is extended (refined) by adding a new literal l. The new clause, $c' = c \cup \{l\}$ is more general than c ($c' \models c$) and potentially covers more examples. By adding just one literal the refinement operator ρ generates the set of clauses that are minimally more general. In order to do this efficiently, a list of atoms is pre-calculated, given a maximal number of variables that can occur within a clause. Within the body and head of a clause, atoms may only be added in a specific order to avoid generating redundant clauses.

A few additional measures are used to enable the implementation to be more efficient. Most importantly, there are two syntactic restrictions on the form of the clauses. Clauses must be connected (i.e. new atoms must always contain a variable that has already occurred) and range-restricted (i.e. no new variables may be introduced in the head of the clause). Moreover, Object Identity is used to specify that variables with different names must denote different objects. Therefore, a clause c with variables $\mathcal{V}$ is extended to the clause c_{OI} by adding literals $x_i \neq x_j$, for $x_i, x_j \in \mathcal{V}, i \neq j$. Finally, the implementation natively supports symmetric predicates for which the order of the terms do not matter

To avoid extensive coverage ($e \models c$) and entailment ($\mathcal{T} \models c$) calculations, two additional tests are used. The subset-test rejects clauses that are supersets of a clause that covers the same or all examples and has been accepted earlier. Because clauses can be formulated in multiple ways, a representative-test is used to remove clauses that are not in canonical form. (Not clear, needs more explanation, or should be removed.)

Table 1: Clause coverage

Example	Covered by	Feature vector
e_1	c_1	$\mathbf{f}_{e_1} = (1,0,0)$
e_2	c_2	$\mathbf{f}_{e_2} = (0,1,0)$
e_3	c_1, c_2, c_3	$\mathbf{f}_{e_3} = (1,1,1)$

Optimization

In order to find weighted soft constraints, soft constraints $\mathcal{C}$ for the given examples are learned by using the constraint learning system with a (low) threshold. The example are then translated to Boolean feature vectors of length $\#\mathcal{C}$ by introducing a feature for every clause $c_i \in \mathcal{C}$. For an example e, the i-th value of its feature vector $\mathbf{f}_e$ is calculated as $\mathbf{f}_e(i) = \mathbf{1}_{e \models c_i}$. Next, existing software is used to learn the linear function $\sum_i w_i \cdot c_i$ over these features, based on user provided preferences. This function corresponds with the scoring function (1)) introduced in Section . It can be used to calculate scores for (unseen) examples and by maximizing the score function an optimal solution can be found.

Example 3. (Rewrite) Consider examples e_1, e_2, e_3, rankings $e_1 \succ e_2, e_2 \succ e_3$ and clauses c_1, c_2, c_3. If the clauses cover examples according to table 1, then, for example, the function $(1 \cdot c_1) + (0 \cdot c_2) + (-2 \cdot c_3)$ perfectly models the given rankings. A new example that is covered by c_1 and c_2, would be assigned a score of 1, according to this function. This score has no value in the absolute sense, it can only be used to compare it to other examples ranked by the same function. In this example clause c_1 represents a desirable property, clause c_2 is ignored because it does not influence the ranking and clause c_3 represents an undesirable property.

SVM^{rank} (Joachims 2006) was chosen to find the scoring function since it uses a linear model and offers a robust and efficient implementation. The weights that it assigns to the features can be used directly as weights for the optimization criteria. While this paper uses SVM^{rank} other linear ranking systems such as Coordinate Ascent (Metzler and Croft 2007) could also be used.

Optimal Solution

Solvers, such as IDP (more on IDP, can be used as ASP solver?), can use clauses directly as hard constraints to generate a solution. There is no solver that can use the chosen optimization criteria directly. Weighted MAX-SAT solvers use propositional clauses and only allow for positive weights.

This optimization task can be solved in IDP, using inductive definitions, aggregates and minimization. The only limitation is that the current version only supports integer values. Therefore, the weights of the clauses are divided by the smallest absolute weight, multiplied by a constant and rounded to the closest integer if necessary.

In order to model the optimization problem in IDP, every clause c_i with variables $v_1, ..., v_n$ is represented by a number i. For every clause a predicate $t(i)$ is added to capture the truth value of the clause. A function $cost(i)$ specifies the cost of not satisfying the clause, which is equal to the weight of the clause.

$$t(i) \Leftrightarrow \forall v_1, ..., v_n : c_i.$$
$$cost(i) = w_i.$$

Using t and $cost$, a function $actual(i) = \mathbf{1}_{\neg t(i)} \cdot cost(i)$ is then defined inductively. This function is used in the optimization criterion $\sum_i actual(i)$ which should be minimized.

Evaluation

In our experiments we aim to measure the accuracy and efficiency of the learning systems for constraints and optimization criteria. In order to account for non-determinism, experiments measuring execution times or optimization scores were repeated eight times and their result were averaged. (First step evaluation constraint learning, then optimization)

Constraints

We used four problems to evaluate constraint learning: (clarify hard constraints)

Map Coloring Countries are assigned colors and neighboring countries may not have the same color. Two examples, each containing three correctly colored countries, are given.

Sudoku A single, solved 4×4 sudoku is given.

Elevator Three examples have been generated, two of which respect the underlying soft constraint.

Co-housing Contains four hard constraints and five examples have been generated that respect the constraints.

(more elaborate descriptions of these?) (link to thesis)

Question 1 (Accuracy)**.** *Are the essential constraints discovered and what influence do different parameters have on the accuracy?*

For all problems the essential constraints are found. Often additional constraints were found that describe some structure in the problems. For example, for the map coloring problem one of the learned constraint states that countries are never their own neighbor. These kind of constraints may help a constraint solver to work more efficiently.

The learning process is parameterized by the maximal amount of variables and literals allowed per clause. If these limits are too large, constraints are found that over-fit the training data. These constraints are too specific and will exclude valid solutions that are not in the training set. On the other hand, if the chosen limits are too small, the necessary constraints will not be found.

Table 2: Execution times overview

Omitted	Problem	Average time (s)
Nothing (baseline)	Map coloring	1.581 (± 0.117)
	Sudoku	4.787 (± 0.062)
	Elevator	3.182 (± 0.073)
	Co-housing	25.903 (± 0.446)
Range (Remove) restriction	Map coloring	4.629 (± 0.199)
	Sudoku	16.118 (± 0.154)
	Elevator	40.453 (± 0.319)
	Co-housing	207.768 (± 0.330)
Connected clauses	Map coloring	1.589 (± 0.110)
	Sudoku	7.068 (± 0.150)
	Elevator	6.157 (± 0.114)
	Co-housing	103.633 (± 0.131)

Over-fitting can be addressed by removing constraints that exclude valid solutions (manually or automatically) as well as providing more (or larger) training examples. Negative examples can be used to detect under-fitting, indicating that the parameter values are too small.

Question 2 (Efficiency)**.** *(review) How fast is the clause learning system and what is the effect of various design decisions and input on the execution time?*

Table 2 shows the execution times for several experiments. It shows that smaller problems can be solved efficiently and shows that removing the syntactical restrictions exponentially increase the search space.

All efficiency measures (i.e. symmetric predicates, subset test and representative test) have been able to improve the execution time, sometimes by more than 50%. The overhead they introduce is more than compensated by the efficiency gains they cause.

Further experiments show that increasing the number of variables or literals per clause impacts the efficiency. Especially the combination of more variables *and* literals can steeply increase the execution time. Therefore it would be useful to adapt these parameters dynamically.

Adding additional examples only increases the execution time by a constant factor. Since adding more examples can help improve the accuracy, this trade-off is often worthwhile.

Question 3 (Compared to humans)**.** *How do learned constraints compare to human programmed constraints?*

Human programmed theories for map coloring and sudoku are available on the website of the IDP system [1]. These theories usually focus on being compact and contain only the essential constraints. Table 3 shows the results of two experiments that measure the time to find a solution for a new problem. This time is measured for the learned theories as well as for hand made theories. The learned theories are slightly adapted to be able to solve the same problems and remove unfair advantages such as preprocessed input data.

[1] https://dtai.cs.kuleuven.be/software/idp

Table 3: CPU times human vs. learned theory

Problem	Type	Average CPU time (s)
Map coloring	Human	0.968 (±0.023)
	Learned	0.403 (±0.015)
Sudoku	Human	1.453 (±0.018)
	Learned	0.310 (±0.012)

Especially for non-experts, a learning system can be useful to assist them during the modeling process. Additionally, the learning system can function in an automatic setting. These experiments show that learned constraints can be used to solve problems efficiently and even faster than hand programmed constraints for the examined cases.

Optimization

(Preferences, not rankings) The efficiency of the learning system for optimization criteria depends mainly on the efficiency of learning soft constraints. Therefore, the experiments in this section are focused on the accuracy of the optimization criteria and the influence of different factors.

The moving problem with input according to example 2 is used to evaluate the learning of optimization criteria. 18 possible configurations are used as available examples $\mathcal{E}$. Four weighted soft constraints are used to represent the underlying model $\mathcal{M}$ for the users preferences. Two approaches are used for evaluation and differ in the way they construct the training ($\mathcal{E}_{train}$) and test sets ($\mathcal{E}_{test}$). In the first approach, given a parameter p_{train}, the examples $\mathcal{E}$ are randomly partitioned into disjoint train and test sets ($\mathcal{E}_{test}$) such that $\#\mathcal{E}_{train} = p_{train} \cdot \#\mathcal{E}$. The second approach uses $\mathcal{E}_{train} = \mathcal{E}_{test} = \mathcal{E}$.

Preferences are generated by picking two examples e_1, $e_2 \in \mathcal{E}_{train}$ and ordering them according to their scores $score(e_i, \mathcal{M})$. Consider all possible pairwise preferences over $\mathcal{E}_{train}$: $pref(\mathcal{E}_{train})$. Given the fraction p_{pref} of preferences to use, then a random subset $\mathcal{P}_{train} \subset pref(\mathcal{E}_{train})$ of size $p_{pref} \cdot \#\mathcal{P}$ is used for learning. Errors in the input preferences are simulated by flipping preferences (e.g. $e_1 \succ e_2$ becomes $e_1 \prec e_2$). Given the fraction p_{error} of errors to introduce, a random subset $\mathcal{P}_{error} \cup \mathcal{P}_{train}$ of the preferences are flipped, where $\#\mathcal{P}_{error} = p_{error} \cdot \mathcal{P}_{train}$.

To evaluate the learned optimization criteria, all possible pairs of examples in the test set are generated. Both the learned model and the underlying model $\mathcal{M}$ are used to predict the better example for every pair. The accuracy of the learned model is then calculated as the fraction of correctly predicted pairs. Pairs for which $\mathcal{M}$ ranks both examples the same are omitted.

Question 4 (Accuracy)**.** *How accurately can learned optimization criteria approximate underlying models?*

Figure 1 shows how the scores improve as the amount of examples and the fraction of included preferences in-

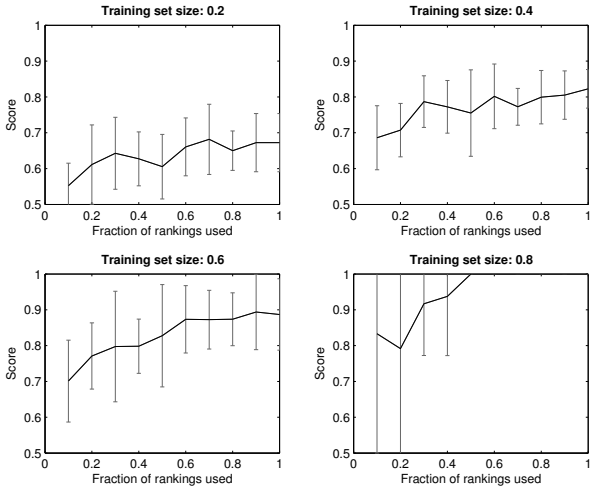

Figure 1: Influence fraction of inequalities

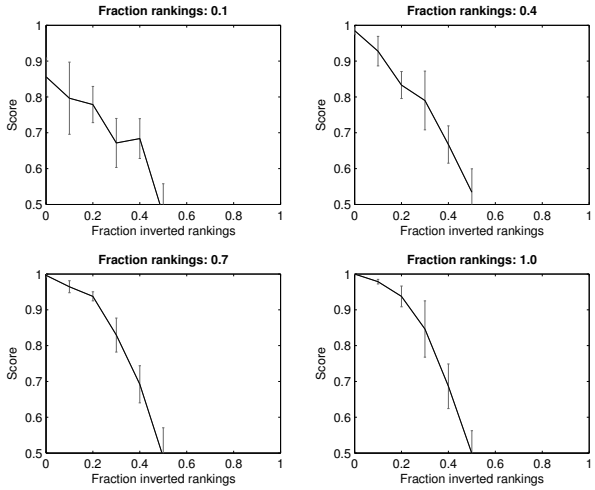

Figure 2: Influence of noise

creases. In all cases, more than half the pairs of examples are correctly predicted and high scores can be obtained even for small datasets. Additional experiments have shown that even for lower scores the learned optimization criteria are often capable of identifying the correct optimal solution. (Refer to thesis)

The standard underlying model can be directly expressed using the soft constraints that can be learned. Even though clausal theories can be very expressive, important restrictions have been placed on the learned clauses in this paper. In order to test the accuracy for models which cannot be directly expressed, a model consisting of two disconnected clauses has been tested as well. While there are limits to the expressibility, the learned optimization criteria were able to obtain similar scores and seem to be robust with respect to the exact formulation.

Question 5 (Noise)**.** *What is the influence of noise on the accuracy?*

The influence of noise is shown in figure 2. Hereby, the second approach of testing is applied, using potentially overlapping train- and test sets. High scores are obtained, even despite significant levels of noise. The figure also shows that providing more rankings improves the robustness of the algorithm, even if the relative amount of noise remains unchanged.

Table 4: Scores for different thresholds (t)

$t = 1$	$t = 2$	$t = 3$	$t = 4$
0.823 (± 0.073)	0.740 (± 0.078)	0.788 (± 0.074)	0.735 (± 0.063)

Question 6 (Threshold)**.** *What is the effect of the soft constraint threshold on the accuracy?*

Table 4 demonstrates that increasing the threshold used for finding soft constraints does not improve the score. This experiment used 40% of the examples as training set and 40% of the available rankings. However, if the size of the problem and examples is increased, a higher threshold will likely be appropriate.

Conclusion

We presented an approach to automatically acquire constraints and optimization criteria from examples and preferences. Implementation for both clause learning and clausal optimization have been provided that accomplish these tasks using first-order logic clauses.

The constraint learning implementation has been able to learn the relevant hard and soft constraints in all experiments. For each problem, only a small number of examples was given to learn from. The system requires only a minimal amount of information from the user, however, it also allows for the use of expressive background knowledge. The learned constraints are domain independent (review domain), which facilitates the construction of positive examples.

Using the constraint learning implementation, the goal of learning optimization criteria has been accomplished. Optimization criteria can be learned that enable optimal solutions to be found. Even for small datasets and noisy rankings, constraints are found that rank most examples correctly.

Aside from learning formal representations automatically from examples, we have shown how these representations can be used in practice. This also forms an important step to enable the learning of optimization criteria in an interactive setting.

Future work (shorten) There are several opportunities for future work. It would be interesting to adapt the

number of variables and literals in a clause dynamically. This could be accomplished, for example, by using negative examples. Learned clauses must be specific enough to not cover any negative examples. (These restrictions are not really described in the paper, so maybe not that interesting to mention here? Or at least add a sentence to introduce them.)

Additionally, it would be interesting to add interactivity to the learning system for the generation of examples or rankings. Rankings expressing that examples are equally ranked are currently ignored. Whenever provided explicitly, however, it could be interesting to incorporate this information into the algorithm.

As mentioned earlier, the domain independence (review domain) of the clauses has several advantages. In some cases, however, specific objects are inherently present in any problem instantiation. The constraint learning system could be enhanced to include such global constants.

Finally, it would be desirable to improve the implementation in order to tackle larger sized problems. All the experiments were conducted with problems of limited size and the computation time increases rapidly if there are more predicates, variables and literals to be used in clauses.

Acknowledgments

The author thanks his promoters Luc De Raedt en Anton Dries. Samuel Kolb is supported by the Research Foundation - Flanders (FWO).

References

Beldiceanu, N., and Simonis, H. 2012. A model seeker: Extracting global constraint models from positive examples. In *Principles and Practice of Constraint Programming*, 141–157. Springer.

Bessiere, C.; Coletta, R.; O'Sullivan, B.; Paulin, M.; et al. 2007. Query-driven constraint acquisition. In *IJCAI*, 50–55.

Bessiere, C.; Coletta, R.; Hebrard, E.; Katsirelos, G.; Lazaar, N.; Narodytska, N.; Quimper, C.-G.; and Walsh, T. 2013. Constraint acquisition via partial queries. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, 475–481. AAAI Press.

Campigotto, P.; Passerini, A.; and Battiti, R. 2011. Active learning of combinatorial features for interactive optimization. In *Learning and Intelligent Optimization*. Springer. 336–350.

De Pooter, S.; Wittocx, J.; and Denecker, M. 2013. A prototype of a knowledge-based programming environment. In *Applications of Declarative Programming and Knowledge Management*. Springer. 279–286.

De Raedt, L., and Dehaspe, L. 1997. Clausal discovery. *Mach. Learn.* 26(2-3):99–146.

Joachims, T. 2006. Training linear svms in linear time. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 217–226. ACM.

Kok, S., and Domingos, P. 2005. Learning the structure of markov logic networks. In *Proceedings of the 22nd international conference on Machine learning*, 441–448. ACM.

Lallouet, A.; Lopez, M.; Martin, L.; and Vrain, C. 2010. On learning constraint problems. In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, volume 1, 45–52. IEEE.

Metzler, D., and Croft, W. B. 2007. Linear feature-based models for information retrieval. *Information Retrieval* 10(3):257–274.

Puget, J.-F. 2004. Constraint programming next challenge: Simplicity of use. In Wallace, M., ed., *Principles and Practice of Constraint Programming - CP 2004*, volume 3258 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 5–8.

Wittocx, J.; Mariën, M.; and Denecker, M. 2008. The idp system: a model expansion system for an extension of classical logic. In *Proceedings of the 2nd Workshop on Logic and Search*, 153–165.