

# Automated Aggregator — Rewriting with the Counting Aggregate

Michael Dingess and Miroslaw Truszczynski\*

Department of Computer Science, University of Kentucky, United States

[madingess@yahoo.com](mailto:madingess@yahoo.com), [mirek@cs.uky.edu](mailto:mirek@cs.uky.edu)

Answer set programming is a leading declarative constraint programming paradigm with wide use for complex knowledge-intensive applications. Modern answer set programming languages support many equivalent ways to model constraints and specifications in a program. However, so far answer set programming has failed to develop systematic methodologies for building representations that would uniformly lend well to automated processing. This suggests that encoding selection, in the same way as algorithm selection and portfolio solving, may be a viable direction for improving performance of answer-set solving. The necessary precondition is automating the process of generating possible alternative encodings. Here we present an automated rewriting system, the *Automated Aggregator* or *AAgg*, that given a non-ground logic program, produces a family of equivalent programs with complementary performance when run under modern answer set programming solvers. We demonstrate this behavior through experimental analysis and propose the system's use in automated answer set programming solver selection tools.

## 1 Introduction

Developers of answer set programming (ASP) solutions often face situations where individual constraints of a problem or even the problem as a whole can be expressed in several syntactically different but semantically equivalent ways. Picking the right representation is crucial to designing these solutions because, given an instance, certain representations perform better (often much better) than others when processed with modern ASP grounders and solvers. However, techniques for selecting a particular representation are often ad hoc and tailored to the needs of the particular application, and require significant programming expertise from the programmer.

About a decade ago, Gebser et al. presented a set of "rules-of-thumb" used by their team in manual tuning of ASP solutions [6]. These rules include suggestions on program rewritings that often result in substantial performance gains. This was verified experimentally by Gebser et al., with all rewritings used in their experiments generated manually [6]. Buddenhagen and Lierler studied the impact of these rewritings on an ASP-based natural language processor and reported orders of magnitude gains in memory and time consumption as a result of some program transformations they executed manually [2]. Because of these promising results, researchers proposed to automate the task of program rewriting. Bichler et al. investigated rewritings of long rules guided by the tree decomposition of graphs built of program rules (a form of join optimization) [1]. Hippen and Lierler proposed a method of applying projection to rewrite rules based on estimates of the size of the ground program [8].

These projects demonstrated that automated program rewritings may lead to programs performing better than the original ones when run under current ASP solvers such as *gringo/clasp* [5]. However, the effects of rewritings are not uniformly the same. In fact, depending on the actual instance they are run

---

\*Research supported by National Science Foundation grant 1707371

with, rewritten programs may perform worse than the original ones. This is a problem because a non-uniform behavior makes the process of selecting the *uniformly best* encoding ill defined. Nevertheless, automated rewritings potentially can significantly improve the state-of-the-art of ASP solving. Namely, the non-uniform behavior of programs obtained by rewriting opens a possibility of using families of equivalent alternative programs in algorithm selection and portfolio solving [13, 14, 9]. In fact, a recent study of suite of six encodings of the Hamiltonian cycle problem shows that one can train performance prediction models to select, given an instance, a program from the suite to run on that instance, guaranteeing a much better overall performance on large sets of instances than that of each of the six programs alone [12].

In this work, we focused on rewritings in which rules using counting based on explicit naming of a required number of objects are rewritten with the use of a *counting* aggregate. We designed a software, *Automated Aggregator* (AAgg), for automating such rewritings into several equivalent forms. We studied the software’s applicability and effectiveness. In our experiments, we applied the software to programs submitted to past ASP Competitions. We found that while many of them already included aggregate expressions and AAgg did not detect any rules to which it could apply, in several cases, it was applicable! For those cases, we studied the performance of the original program and the rewritings produced by AAgg. The results showed that depending on an instance, rewritings produced by our software often performed better than the original programs. In other words, the family of encodings generated by AAgg (the original program and its rewritings) showed a complementary performance on both the instances used in the ASP Competitions and on instances which we generate ourselves. These results show that AAgg can be used as a tool for generating collections of encodings to be used in algorithm selection and portfolio solving. A systematic experimental verification of this claim will be the subject of a future work.

## 2 Aggregate Equivalence Rewriting

In this section we describe the *aggregate equivalence* rewriting, its input and output forms. Currently, we support one input form and three output forms. The input form is a rule that expresses a constraint that there are  $b$  different objects with a certain property by explicitly introducing  $b$  variables to name these  $b$  objects. The output forms model the same property by relying, in some way, on the counting aggregate. For each of the rewritings we establish its correctness and experimentally study its performance. The correctness follows from Theorems A.3 and A.5, presented and proved in the appendix.

### 2.1 Preliminaries

We follow the *ASP-Core-2 Input Language Format* [3]. We consider rules of the form  $head \leftarrow body$ . The *head* may consist of a single literal or be empty, the latter representing a contradiction, making the rule a *constraint*. The *body* may contain one or more literals or be empty, which constitutes a *fact*. *Literals* are composed of an atom, which may be preceded by *not*. *Negative* literals include *not*; *positive* literals do not. *Atoms* have the form  $p(t_1, \dots, t_k)$ , where  $p$  is a predicate symbol of arity  $k$  and each  $t_i$  is a term, that is, a constant, variable, or expression of the form  $f(t_1, \dots, t_k)$  where  $f$  is a function symbol of arity  $k > 0$  and each  $t_i$  is a term.<sup>1</sup>

Atoms may also take the form of an aggregate expression. In this work, we focus on *counting*

---

<sup>1</sup>AAgg also accepts other term expressions following the definition used by *clingo* [5].

aggregates, that is, expressions of the form:

$$s_1 \prec_1 \#count\{\mathbf{t}_1 : \mathbf{L}_1; \dots; \mathbf{t}_n : \mathbf{L}_n\} \prec_2 s_2 \quad (1)$$

In (1),  $\mathbf{t}_i$  and  $\mathbf{L}_i$  form an *aggregate element*, which is a non-empty tuple of terms and literals, respectively. The *count* operation simply counts the number of *unique* term tuples  $\mathbf{t}_i$  whose corresponding condition  $\mathbf{L}_i$  holds. The result of the count function is compared by the *comparison predicates*  $\prec_1$  and  $\prec_2$  to the terms  $s_1$  and  $s_2$ . These comparison predicates may be one of  $\{<, \leq, =, \neq\}$ . One or both of these comparisons can be omitted [5].

## 2.2 Input Forms

The aggregate equivalence rewriting takes as input rules of the form:

$$H \leftarrow \bigwedge_{1 \leq i \leq b} F(X_i, \mathbf{Y}), \bigwedge_{1 \leq i < j \leq b} X_i \neq X_j, G. \quad (2)$$

where

- $H$  is the head of a rule ( $H$  may be empty making the rule a constraint)
- $F$  is a predicate of arity  $1 + |\mathbf{Y}|$
- $X_1, \dots, X_b$  are variables, all in the same position in  $F$
- $\mathbf{Y}$  is a comma-separated list of variables, identical in variables and variable positions for all  $F$  in the rule
- $G$  is the remaining body of the rule, possibly empty,

and the following hold true:

- $b \geq 2$
- $H, G$ , and  $\mathbf{Y}$  have no occurrences of variables  $X_1, \dots, X_b$
- The terms  $\bigwedge_{1 \leq i < j \leq b} X_i \neq X_j$  may instead be a continuous chain of comparisons:

$$\bigwedge_{1 \leq i \leq b-1} X_i < X_{i+1} \text{ or } \bigwedge_{1 \leq i \leq b-1} X_i > X_{i+1}.$$

Note that  $X_i$ 's need not be in the first position in  $F$ , so long as they are all in the same position in  $F$  and the other variables in  $F$  (if any) are identical in all occurrences of  $F$  in the rule. Additionally, some other forms logically equivalent to  $\bigwedge_{1 \leq i < j \leq b} X_i \neq X_j$  are also acceptable. For instance, the condition  $X_1 \neq X_2$  may be expressed as  $X_1 + a \neq X_2 + a$ , for some integer  $a$ .

## 2.3 Output Forms

The form of the output depends on the size of  $\mathbf{Y}$ . When  $|\mathbf{Y}| = 0$ , the output form is:

$$H \leftarrow b \leq \#count\{X : F(X)\}, G. \quad (3)$$

where

- $H, b, F$ , and  $G$  are as above
- Aggregate element  $X : F(X)$  follows the form **term : literal** as defined above. The prior,  $X$ , is a tuple of one term, which in this case is a variable. The second,  $F(X)$ , is a literal.

When  $|\mathbf{Y}| > 0$ , we perform *projection* to project out the variable  $X$  from  $F$ . This gives us an output form consisting of two rules:

$$\begin{aligned} H &\leftarrow b \leq \#\text{count}\{X : F(X, \mathbf{Y})\}, F'(\mathbf{Y}), G. \\ F'(\mathbf{Y}) &\leftarrow F(X, \mathbf{Y}). \end{aligned} \quad (4)$$

where

- $H, b, F, \mathbf{Y}$ , and  $G$  are as above
- Aggregate element  $X : F(X, \mathbf{Y})$  follows the form **term : literal** as above
- $F'$  is a new predicate symbol of arity equal to the size of  $\mathbf{Y}$ , that is, equal to the arity of  $F$  minus one; introducing  $F'$  ensures that variables in  $\mathbf{Y}$  are universally quantified.

The correctness of this rewriting, follows from the results presented in the appendix (Theorem A.3). Specifically, we show there that if a program is obtained from another program by rewriting one of its rules in the way described above, then both programs have the same answer sets (modulo atoms  $F'(\mathbf{y})$ , if  $F'$  is introduced).

### 2.3.1 Alternative Output Forms

Two alternative, logically equivalent output forms are also available, each derived from the output form presented above. First, we observe that  $b \leq F$  is logically equivalent to the negation of  $F < b$ . We can then restate the original literal as follows:

$$\text{not } \#\text{count}\{X : F(X, \mathbf{Y})\} < b. \quad (5)$$

Second, we note that the input language we consider permits integer-only arithmetic. Consequently, the expression  $\text{not } a < b$  for some integers  $a$  and  $b$  is equivalent to the conjunctive expression:

$$\neg(a = -\infty) \wedge \neg(a = -\infty + 1) \wedge \dots \wedge \neg(a = b - 2) \wedge \neg(a = b - 1).$$

Additionally, the result of the count never returns a negative number. Therefore, we can restate the aggregate literal in this alternative output form as a conjunction of aggregate literals having the form:

$$\begin{aligned} \text{not } \#\text{count}\{X : F(X, \mathbf{Y})\} &= 0, \\ \text{not } \#\text{count}\{X : F(X, \mathbf{Y})\} &= 1, \\ &\dots, \\ \text{not } \#\text{count}\{X : F(X, \mathbf{Y})\} &= b - 1. \end{aligned} \quad (6)$$

Note that, due to the precise semantics of logic programs, the equivalence of these two alternative logic forms relies on additional assumptions about the input program (it has to be splittable) and the rule to be rewritten. Theorem A.5 provides conditions under which the rewriting is guaranteed to be correct. These conditions are checked by our software and only when they hold, the software proceeds with rewriting into an alternative form (5) or (6), as selected by the user.

### 3 The Automated Aggregator System

We now present the *Automated Aggregator*<sup>2</sup> (AAGg) software system for performing the Aggregate Equivalence rewriting. The software provides an automated way to detect rules within a given program following the input format (2) and rewrite those rules into an equivalent output format (3/4), (5), or (6).

#### 3.1 Usage

The software relies on the *clingo* Python module provided by the Potassco suite [7]. The module is written in Python 2.7. As such, Python 2.7 is required to run the *Automated Aggregator* system. Installation information is provided in the software’s README file. The *Automated Aggregator* is invoked as follows:

```
python aagg/main.py
    [-h,--help] [-o,--output FILENAME] [--no-rewrite]
    [--no-prompt] [--use-anonymous-variable]
    [--aggregate-form ID] [-d,--debug] [-r,--run-clingo]
    [encoding_1 encoding_2 ...]
```

The `-h` flag lists the help options. The encoding(s) are the filename(s) of the input encoding(s), and the output is the desired name for the output file. If no output filename is given, one is generated based on the first input filename given. When a candidate rule is discovered, the user is shown the proposed rewriting and prompted for confirmation. If the `--no-rewrite` flag is given, no prompts are given and no rewriting is performed. If the `--no-prompt` flag is given, no prompts are given and rewriting is performed where possible. The ID supplied to the `--aggregate-form` argument informs the program which aggregate form to use when performing rewrites: its values 1, 2, and 3 correspond to aggregate forms (3/4), (5), and (6), respectively. The `-d` debug flag directs the application to operate with verbosity, printing details during the rewriting candidate discovery process and printing some statistics after the application’s conclusion. The `--r run-clingo` flag directs the application to run the resulting program through *clingo* after any rewritings are performed.

Finally, the `--use-anonymous-variable` flag indicates an additional modification of the output form to be performed. It uses the anonymous variable ‘`_`’ in place of the variable `X` in the output forms listed above, with some additional modifications of the rule to ensure the transformation is correct. Specifically, we replace the aggregate element `X : F(X, Y)` with `F(_ , Y) : F(_ , Y)` rather than with `_ : F(_ , Y)`, because the latter is not a valid *gringo* syntax. We mention this option for completeness sake, since it is available in our implementation. However, we found that the programs generated when using and when not using the anonymous variable are identical after grounding, so we neither discuss it further nor use these rewritings in our experiments.

By default all boolean flags are disabled and the aggregate-form ID is set to 1 indicating output form (3). At least one input encoding filename must be specified.

#### 3.2 Methodology

The methodology used for discovering whether a rule is a candidate for the aggregate equivalence rewriting is as follows. The given logic program(s) are parsed by the *clingo* module, generating an abstract

---

<sup>2</sup>The system and all encodings, test instances, and driver programs can be found online at:  
[https://drive.google.com/drive/folders/1lqRsy9HGIDHvyX\\_Pvk8zKt1-xvbwcAp?usp=sharing](https://drive.google.com/drive/folders/1lqRsy9HGIDHvyX_Pvk8zKt1-xvbwcAp?usp=sharing)

syntax tree for each rule. Each such tree is passed to a *transformer* class for preprocessing. After preprocessing, some information is gathered from the program as a whole; specifically, predicate dependencies are determined, which in turn determine when output forms (5) and (6) are appropriate for a given rule (see the appendix for more details). Rules are then passed individually to an *equivalence transformer* class for processing. After processing, and if the requested rewriting is possible and confirmed by the user, the rewritten form of the rule is returned. Otherwise the original rule is returned. All returned rules, rewritten or not, are collected and output to the desired output file location. Optionally, the resulting program is then run using *clingo*.

When a rule is passed to the equivalence transformer for processing, it first undergoes a process of exploration, which traverses the rule's abstract syntax tree, recording comparison literals between two variables as well as other pertinent information found along the way. These comparison literals are scrutinized to determine whether a subset of the comparisons follows the form given in (2) or some equivalent format as detailed in the section 2.2. The process also identifies those variables that play the role of  $X_1, \dots, X_b$  as in (2). The rule is then analyzed to determine whether there are  $b$  occurrences of some positive literal of predicate  $F$  with the arguments  $X_i$  and  $\mathbf{Y}$ , where each  $X_i$ ,  $1 \leq i \leq b$ , exists at least once in the set of occurrences, and  $\mathbf{Y}$  is the same for each occurrence and contains no variables  $X_1, \dots, X_b$ . Let us call this set of  $b$  literals combined with the corresponding comparisons following the form given in (2) or equivalent, the rule's *counting literals*. Similarly, we denote the variables playing the role of the variables  $X_i$  as given in (2) as the rule's *counting variables*.

After gathering counting literals and counting variables, the equivalence transformer verifies that the counting variables are not used within literals anywhere in the rule excluding literals within the set of counting literals. If this verification fails or if any of the constraints for the counting literals cannot be satisfied or if no such set of counting variables can be obtained, then the rule is not fit for rewriting. As a result, no rewriting is performed on the rule and the original rule is returned.

In the other case, if the verification succeeds (the counting literals and the counting variables satisfy all the required constraints), then we proceed as follows. If the requested output form is (3/4), then we check whether  $|\mathbf{Y}| > 0$  and rewrite the rule accordingly using (3) or (4). If the requested output form is (5) or (6), then the predicate dependency condition related to splitting must be verified (see the appendix for details). If it holds, then we check whether  $|\mathbf{Y}| > 0$  and rewrite the rule into the desired form (5) or (6), adding the second of the two rules in (4) when  $|\mathbf{Y}| > 0$ .

Note that the user is first prompted for confirmation for each rewritten rule before the rule is added to the final program. If the user denies a rewriting, then the original rule is used.

### 3.3 Limitations

Here we discuss the limitations of the *Automated Aggregator* in its current form.

1. The rewriting is one-directional. The software will only rewrite rules from the form (2) into rules of the forms (3/4), (5), and (6). As it stands, the system will not rewrite rules given in any of the forms (3/4), (5), or (6) into rules of any other form.
2. In the cases when multiple rewritings are possible for a single rule, only one rewriting will be detected and performed. To illustrate, if the form given in (2) occurs twice in one rule over a disjoint set of variables and predicates, where both sets of counting literals consider the other set as part of  $G$  and all conditions hold, then only the set of counting literals with the highest number of counting variables will be used for rewriting. If both sets contain the same number of counting

Table 1: Problem Domains of ASP Competition Encodings with Rules for Rewriting

ASP Competition Year	Problem Domain of Encodings
2009	Golomb Ruler & Wire Routing
2014	Steiner Tree
2015	Steiner Tree & Graceful Graphs

variables, then one is chosen based on the order in which the comparisons using those variables occur in the rule.

3. There are some cases in which the software will not recognize a valid rewriting when one exists. These cases mostly lie in the many obscure ways of representing a chain of comparisons equivalent to that in (2). However, there are no known cases in which an incorrect rewriting will be proposed when no valid rewriting is possible. More details are given in the software’s README document.

## 4 Experimental Analysis

The *Automated Aggregator* system has been made available for download online.<sup>3</sup> Encodings with which the application was tested, their corresponding instances, and (Python) scripts for driving such tests, are included there too.

### 4.1 Automated Aggregator in Practice

The *Automated Aggregator* system was applied to logic programs in *gringo* syntax submitted to the 2009, 2014, and 2015 Answer Set Programming Competitions. Of the 58 encodings given to the application, five contained rules which were candidates for the rewriting described. Table 1 lists the encoding problem names and the ASP Competitions for which they were developed.

Additionally, to provide an example of the functionality of *AAgg*, the system was applied to an encoding for the Hamiltonian Cycle problem. The original program is shown in Figure 1. The rewritten program, as output by *AAgg*, is shown in Figure 2. We see that the first two constraints in the program are both rewritten with the necessary projection performed as in (4). Experimental results for these encodings when applied to a generated set of hard instances are given in the following section.

### 4.2 Experimental Results

Results were gathered by systematically grounding and solving each instance-encoding pair within families of encodings for each problem type. The data sets of instances we used are available together with the *AAgg* tool (see the url listed earlier).<sup>4</sup> Each encoding in each family of encodings was run for each instance. The total grounding plus solving time of instance-encoding pairs were recorded and compared.

The precise encodings used as input to the *AAgg* software for gathering results are listed in Table 2. Two output encodings are generated for each input encoding and together they form the encoding family for that problem domain. The two output forms used were those shown in equations (3/4) and

<sup>3</sup>[https://drive.google.com/drive/folders/1lqRs9HGIDHvyX\\_Pvkc8zKt1-xvbwcAp?usp=sharing](https://drive.google.com/drive/folders/1lqRs9HGIDHvyX_Pvkc8zKt1-xvbwcAp?usp=sharing)

<sup>4</sup>We thank Daniel Houston and Liu Liu for providing us with the data sets of instances for the Latin Square and the Hamiltonian Cycle problems, respectively. The data sets for the remaining problems were generated by software tools we developed for the purpose. These tools and descriptions of instance sets used in experiments can be found at the *AAgg* site.

---

```

node(X) :- edge(X,Y).
node(X) :- edge(Y,X).

{ hc(X,Y) } :- edge(X,Y).
:- hc(X,Y), hc(X,Z), Y!=Z.
:- hc(X,Y), hc(Z,Y), X!=Z.

reach(X,Y) :- hc(X,Y).
reach(X,Y) :- hc(X,Z), reach(Z,Y).

:- node(X), node(Y), not reach(X,Y).

```

Figure 1: Example Hamiltonian Cycle problem encoding. Original version.

---

```

node(X) :- edge(X,Y).
node(X) :- edge(Y,X).

{ hc(X,Y) } :- edge(X,Y).
:- 2 <= #count{ Y : hc(X,Y) }, hc_project_Z(X).
hc_project_Z(X) :- hc(X,Y).
:- 2 <= #count { X : hc(X,Y) }, hc_project_Z1(Y).
hc_project_Z1(Y) :- hc(X,Y).

reach(X,Y) :- hc(X,Y).
reach(X,Y) :- hc(X,Z), reach(Z,Y).

:- node(X), node(Y), not reach(X,Y).

```

Figure 2: Example Hamiltonian Cycle problem encoding. Rewritten version.

Table 2: Sources of Encodings Used for Experimental Results

Source of Encoding	Problem Domain of Encodings
ASP Competition 2009	Wire Routing
ASP Competition 2015	Steiner Tree & Graceful Graphs
Home-Brewed	Latin Squares& Hamiltonian Cycle <sup>a</sup>

<sup>a</sup>We are thankful to Daniel Houston and Liu Liu for supplying the Latin Squares and Hamiltonian Cycle instance sets, respectively.

Table 3: Result Statistics by Problem Domain

Encoding	Wins	Exclusive Wins	Wins by 20%	Wins by 50%
Wire Routing Input Encoding	122 (58.4%)	34 (27.8%)	115 (92.0%)	97 (77.6%)
	50 (24.0%)	26 (52.0%)	48 (87.3%)	38 (69.1%)
	37 (17.7%)	13 (35.1%)	39 (90.7%)	23 (53.5%)
Steiner Tree Input Encoding	130 (33.9%)	0 (0%)	1 (0.8%)	0 (0%)
	123 (32.1%)	0 (0%)	4 (3.3%)	0 (0%)
	130 (33.9%)	0 (0%)	6 (4.6%)	0 (0%)
Graceful Graphs Input Encoding	212 (37.3%)	62 (29.2%)	182 (85.8%)	128 (60.4%)
	97 (17.1%)	23 (23.7%)	82 (84.5%)	58 (59.8%)
	259 (45.6%)	51 (19.7%)	229 (88.4%)	169 (63.7%)
Latin Squares Input Encoding	5611 (77.7%)	5 (0.1%)	4435 (79.0%)	2003 (35.7%)
	1432 (19.8%)	0 (0%)	655 (45.7%)	86 (6.0%)
	176 (2.4%)	0 (0%)	64 (36.4%)	7 (4.0%)
Hamiltonian Cycle Input Encoding	72 (28.7%)	47 (65.3%)	59 (81.9%)	34 (47.2%)
	102 (40.6%)	69 (67.7%)	79 (77.5%)	52 (51.0%)
	77 (30.7%)	45 (58.4%)	64 (83.1%)	44 (57.1%)

(6); previous experiments showed extreme similarity between forms (3/4) and (5), so (5) was left out to preserve machine time. The machine used for testing contained an Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz with 16GB RAM.

Result statistics are shown in Table 3. Data is grouped by problem domain. The first line of each grouping shows statistics for the input encoding. The second line shows statistics for the encoding output by the *AAGG* software using the first line as input and the output rule form (3/4) as the selected output form. The third and final line of each grouping shows statistics for the encoding output by the *AAGG* software again using the first line as input but now selecting the output form (6) as the chosen output form.

A win is when an encoding grounds and solves for an instance in the shortest amount of time as compared with the other two encoding in its grouping. The percentage value beside the win count is the proportion of instances which that encoding won relative to the number of instances in the instance set for which at least one of the three encodings terminated within the total time limit set for grounding and solving. This time limit was set to 200 seconds in all experiments, except those with encodings of the Hamiltonian Cycle problem that used a timeout value of 400 seconds due to the relative difficulty of the instance set (in all problem domains except for the Hamiltonian Cycle domain, the number of instances for which no encoding computed an answer within the time limit was very small; for the Hamiltonian

Cycle problem, even with the increased time limit, it was significant).

An exclusive win is when the encoding is the only encoding in its grouping to find a solution (or determine there is no solution) for an instance while both of the other two encodings failed to do so within the time limit. The column shows the number and the percentage of wins which were exclusive wins for the encoding. The next column shows the number and the percentage of wins by a margin of at least 20% (this is when the best encoding runs at least 20% faster than the second one; in case of exclusive wins, we count a win as by at least 20%, if it is faster by at least 20% than the time limit used). The data in the last column shows the numbers and percentages of wins by at least 50% (it is to be interpreted similarly to the data in the previous column).

Results indicate that in some cases, the rewriting can provide complementary performance of encodings. Specifically, the results for the Wire Routing, Graceful Graphs, and Hamiltonian Cycle problems support the claim. This is shown first by the fact that for each problem, each of the encodings scores a significant proportion of wins, and that among those wins a significant proportion are exclusive wins, and an even greater proportion (about 50% or more) are wins by at least 50%. This means that about half of the times when an encoding outperforms the two other encodings, it outperforms both by a factor of at least two.

The Steiner Tree results indicate that sometimes the rewriting produces little to no effect at all. While each encoding for the problem registers a similar proportion of wins, there are very few instances (just eleven out of 283) when the best encoding outperforms the other two by 20% or more, and no instances when the best encoding would outperform the other two by 50% or more. The Latin Squares results are mixed in the sense that for the most part the original encoding works best. However, one of the rewritings registers almost 20% of wins. Moreover, about 45% of those wins are by 20% or more and 6% by 50% or more.

In summary, we see that the rewriting can provide programs that perform complementary to the original. This complementarity is perhaps somewhat surprising, because aggregates are assumed to lead to better performance. Our results show that the picture is more complicated and whether rewriting with aggregates yields better performance is instance-dependent. It is also interesting to note that for some encodings replacing simple counting, like that used in the Latin Square encodings (no two identical elements in a row or column), with the *count* aggregates does not lead to substantial improvements. Finally, for some problems (in our experiments for the Steiner Tree problem) where complementary behavior does emerge, the differences in performance are relatively small.

## 5 Future Work

As detailed in Section 3.3, the Automated Aggregator software can be improved. Extending the software so that it rewrites rules by eliminating the *counting* aggregate by inverting the current rewriting seems to be a potentially most beneficial direction. Indeed, our results show that introducing aggregates does not guarantee uniformly improved performance. This gives reason to think that removing the *counting* aggregate, that is, rewriting it in an explicit way, has a potential of generating encodings that on many instances may perform better. Just like the present version of the AAgg system, this could yield collections of encodings of complementary strengths. Moreover, this form of rewriting would be quite widely applicable, as the *counting* aggregate is commonly used.

Expanding the software to detect and perform multiple aggregate equivalence rewritings on a single rule and to detect more obscure forms of representations are two other directions for improvement. While necessary to ensure the software has a possibly broad scope of applicability, we do not expect

these extensions to have a major practical impact due to low frequency with which such less intuitive and convoluted forms of modeling are found in programs.

Automated rewriting has been studied before. The PROJECTOR system [8] and the *lpopt* system [1] are two notable examples. That earlier work sought to develop rewritings improving the performance over the original ones. That goal is in general difficult to meet; both systems were shown to offer gains, but the rewritten encodings are not always performing better. Our goal was different. We aimed at rewritings of varying relative performance depending on input instances. With the AAgg system we showed that even a rather simple rewriting consisting of introducing the *count* aggregate often leads to families of encodings with complementary strengths (areas of superior performance). This opens a way for using machine learning to develop models in support of effective encoding selection or encoding portfolio solving [12]. In such work, to generate promising collections of encodings, one could use the AAgg system, with extensions mentioned above, but also other program rewriting software such as PROJECTOR and *lpopt* which, as noted, while they do yield good encodings, often better than the original one, they do not perform uniformly better.

## References

- [1] Manuel Bichler, Michael Morak & Stefan Woltran (2016): *lpopt: A Rule Optimization Tool for Answer Set Programming*. In Manuel V. Hermenegildo & Pedro López-García, editors: *Proceedings of the 26th International Symposium on Logic-Based Program Synthesis and Transformation, LOPSTR 2016, Selected Papers, Lecture Notes in Computer Science 10184*, Springer, pp. 114–130, doi:10.1007/978-3-319-63139-4\_7.
- [2] Matthew Buddenhagen & Yuliya Lierler (2015): *Performance Tuning in Answer Set Programming*. In Francesco Calimeri, Giovambattista Ianni & Miroslaw Truszczynski, editors: *Proceedings of the 13th International Conference on Logic Programming and Nonmonotonic Reasoning, LPNMR 2015, Lecture Notes in Computer Science 9345*, Springer, pp. 186–198, doi:10.1007/978-3-319-23264-5\_17.
- [3] Francesco Calimeri, Wolfgang Faber, Martin Gebser, Giovambattista Ianni, Roland Kaminski, Thomas Krennwallner, Nicola Leone, Marco Maratea, Francesco Ricca & Torsten Schaub (2020): *ASP-Core-2 Input Language Format*. *Theory Pract. Log. Program.* 20(2), pp. 294–309, doi:10.1017/S1471068419000450.
- [4] Paolo Ferraris, Joohyung Lee, Vladimir Lifschitz & Ravi Palla (2009): *Symmetric Splitting in the General Theory of Stable Models*. In Craig Boutilier, editor: *Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI 2009*, pp. 797–803. Available at <http://ijcai.org/Proceedings/09/Papers/137.pdf>.
- [5] M. Gebser, R. Kaminski, B. Kaufmann, M. Lindauer, M. Ostrowski, J. Romero, T. Schaub & S. Thiele (2015): *Potassco User Guide*. Available at <https://github.com/potassco/guide/releases/>.
- [6] Martin Gebser, Roland Kaminski, Benjamin Kaufmann & Torsten Schaub (2011): *Challenges in Answer Set Solving*. In Marcello Balduccini & Tran Cao Son, editors: *Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning - Essays Dedicated to Michael Gelfond on the Occasion of His 65th Birthday, Lecture Notes in Computer Science 6565*, Springer, pp. 74–90, doi:10.1007/978-3-642-20832-4\_6.
- [7] Martin Gebser, Roland Kaminski, Benjamin Kaufmann & Torsten Schaub (2019): *Multi-shot ASP solving with clingo*. *Theory Pract. Log. Program.* 19(1), pp. 27–82, doi:10.1017/S1471068418000054.
- [8] Nicholas Hippen & Yuliya Lierler (2019): *Automatic Program Rewriting in Non-Ground Answer Set Programs*. In José Júlio Alferes & Moa Johansson, editors: *Proceedings of the 21th International Symposium on Practical Aspects of Declarative Languages, PADL 2019, Lecture Notes in Computer Science 11372*, Springer, pp. 19–36, doi:10.1007/978-3-030-05998-9\_2.
- [9] Holger H. Hoos, Marius Lindauer & Torsten Schaub (2014): *claspfolio 2: Advances in Algorithm Selection for Answer Set Programming*. *Theory Pract. Log. Program.* 14(4-5), pp. 569–585, doi:10.1017/S1471068414000210.

- [10] Yuliya Lierler (2019): *Strong Equivalence and Program's Structure in Arguing Essential Equivalence Between First-Order Logic Programs*. In José Júlio Alferes & Moa Johansson, editors: *Proceedings of 21th International Symposium on Practical Aspects of Declarative Languages, PADL 2019, Lecture Notes in Computer Science 11372*, Springer, pp. 1–18, doi:10.1007/978-3-030-05998-9\_1.
- [11] Vladimir Lifschitz & Hudson Turner (1994): *Splitting a Logic Program*. In Pascal Van Hentenryck, editor: *Proceedings of the 11th International Conference on Logic Programming, ICLP 1994*, MIT Press, pp. 23–37.
- [12] Liu Liu & Miroslaw Truszcynski (2019): *Encoding Selection for Solving Hamiltonian Cycle Problems with ASP*. In Bart Bogaerts, Esra Erdem, Paul Fodor, Andrea Formisano, Giovambattista Ianni, Daniela Inclezan, Germán Vidal, Alicia Villanueva, Marina De Vos & Fangkai Yang, editors: *Proceedings of the 35th International Conference on Logic Programming, ICLP 2019, Technical Communications, EPTCS 306*, pp. 302–308, doi:10.4204/EPTCS.306.35.
- [13] John R. Rice (1976): *The Algorithm Selection Problem*. *Advances in Computers* 15, pp. 65–118, doi:10.1016/S0065-2458(08)60520-3.
- [14] Lin Xu, Frank Hutter, Holger H. Hoos & Kevin Leyton-Brown (2008): *SATzilla: Portfolio-based Algorithm Selection for SAT*. *J. Artif. Intell. Res.* 32, pp. 565–606, doi:10.1613/jair.2490.

## A Correctness of Rewritings

Let us consider the following program rule ( $H$  may be  $\perp$ ):

$$H \leftarrow \bigwedge_{1 \leq i \leq b} F(X_i, \mathbf{Z}, \mathbf{Z}'), Q(\mathbf{X}), G(\mathbf{Z}', \mathbf{Z}''). \quad (7)$$

where  $F$  is a predicate,  $\mathbf{X}$  a tuple of variables  $X_1, \dots, X_b$ ,  $Q(\mathbf{X})$  is a list of literals over variables  $X_1, \dots, X_b$ , and  $\mathbf{Z}, \mathbf{Z}'$  and  $\mathbf{Z}''$  are three pairwise disjoint tuples of variables, and disjoint with  $\mathbf{X}$ , and  $H$  contains none of  $X_i$ .

Let  $P$  be a program containing rule (7) and let  $P'$  be the program obtained from  $P$  by replacing that rule with the two rules

$$\begin{aligned} H \leftarrow \bigwedge_{1 \leq i \leq b} F(X_i, \mathbf{Z}, \mathbf{Z}'), Q(\mathbf{X}), G(\mathbf{Z}', \mathbf{Z}''), F'(\mathbf{Z}). \\ F'(\mathbf{Z}) \leftarrow F(X, \mathbf{Z}, \mathbf{Z}'). \end{aligned} \quad (8)$$

where  $F'$  is a predicate not occurring in  $P$ .

**Theorem A.1.** *The programs  $P$  and  $P'$  have the same answer sets modulo ground atoms of the form  $F'(\mathbf{z})$ .*

*Proof.* (Sketch) Consider a ground instance  $r$  of rule (7) and let  $a$  be the first atoms in the body of  $r$  (that is, a ground instance of the atom  $F(X_1, \mathbf{Z}, \mathbf{Z}')$ ). In  $P'$  there are ground rules  $r'$  and  $r''$  obtained from (8) using the same variable instantiation as that used to produce  $r$ . Clearly,  $r$  contributes to the reduct of  $ground(P)$  if and only if  $r'$  and  $r''$  contribute to the reduct of  $ground(P')$ . Moreover, if they do,  $r$  “fires” in the least model computation if and only if  $r'$  fires in the least model computation. It follows that an interpretation  $I$  of  $P$  is an answer set of  $P$  if and only if  $I \cup J$  is an answer set of  $P'$ , where  $J$  consists of all atoms  $F'(\mathbf{z})$  such that  $F(x, \mathbf{z}, \mathbf{z}') \in I$ , for some constant  $x$  and a tuple of constants  $\mathbf{z}'$ .  $\square$

Next, we recall the following theorem proved by Lierler [10].

**Theorem A.2.** *Let  $H$  be an atom ( $H$  may be  $\perp$ ),  $G$  a list of literals,  $X$  a variable,  $\mathbf{Z}$  a tuple of variables, each different from  $X$  and each with at least one occurrence in a literal in  $G$ , and  $F$  a predicate symbol of arity  $1 + |\mathbf{Z}|$ . If  $b$  is an integer, and  $X_1, \dots, X_b$  are variables without any occurrence in  $H$  and  $G$ , then the logic program rule*

$$H \leftarrow \bigwedge_{1 \leq i \leq b} F(X_i, \mathbf{Z}), \bigwedge_{1 \leq i < j \leq b} X_i \neq X_j, G. \quad (9)$$

*is strongly equivalent to the logic program rule*

$$H \leftarrow b \leq \#\text{count}\{X : F(X, \mathbf{Z})\}, G. \quad (10)$$

*where  $\wedge$  is used to represent a sequence of expressions separated by commas.*

Combining the two results leads to the following result that proves the correctness of the first rewriting implemented by our tool *AAgg*.

**Theorem A.3.** *Let  $P$  be a program containing a rule  $r$  of the form*

$$H \leftarrow \bigwedge_{1 \leq i \leq b} F(X_i, \mathbf{Z}, \mathbf{Z}'), \bigwedge_{1 \leq i < j \leq b} X_i \neq X_j, G(\mathbf{Z}', \mathbf{Z}''). \quad (11)$$

*under the same assumptions about variable tuples  $\mathbf{X}, \mathbf{Z}, \mathbf{Z}'$  and  $\mathbf{Z}''$  as before. If  $\mathbf{Z}$  is empty, let  $P'$  be  $P$ . Otherwise, let  $P'$  be obtained from  $P$  by replacing  $r$  with the rules (8), adjusting the first of them to contain  $\bigwedge_{1 \leq i < j \leq b} X_i \neq X_j$  in place of  $Q(\mathbf{X})$ . Let us call the first of these two rules  $r$  (reusing the name of the original rule). Finally, let  $P''$  be obtained from  $P'$  by replacing  $r$  with the corresponding rule (10). Then, the programs  $P$  and  $P'$  have the same answer sets (in the case, when  $\mathbf{Z}$  is not empty, the same answer sets modulo atoms  $F'(\mathbf{z})$ ).*

*Proof.* Clearly, rule  $r$  is a special case of a rule of the form (7). By A.1, programs  $P$  and  $P'$  have the same answer sets (when  $\mathbf{Z}$  is not empty, the same answer sets modulo atoms  $F'(\mathbf{z})$ ). The rule  $r$  in  $P'$  is of the form (9) required by Theorem A.2. Applying this theorem shows that programs  $P'$  and  $P''$  have the same answer sets and the assertion follows.  $\square$

Once a program has a rule of the form (10), we can often modify it further by exploiting alternative encodings of the aggregate expressions. In particular, under some assumptions about the structure of the program, we can replace a rule (10) by

$$H \leftarrow \text{not } \#\text{count}\{X : F(X, \mathbf{Z})\} < b, G. \quad (12)$$

where we assume that the variable  $X$  is not a variable in  $\mathbf{Z}$ , and that all variables in  $\mathbf{Z}$  appear in  $G$ .

We recall that a partition  $(P_b, P_t)$  of a program  $P$  is a *splitting* of  $P$  if no predicate appearing in the head of a rule from  $P_t$  appears in  $P_b$  [11, 4]. A well-known result on splitting states that answer set of programs that have a splitting can be described in terms of answer sets of programs that form the splitting.

**Theorem A.4.** *Let  $P$  be a logic program and let  $(P_b, P_t)$  be a splitting of  $P$ . For every answer set  $I_b$  of  $P_b$ , every answer set of the program  $P_t \cup I_b$  is an answer set of  $P$ . Conversely, for every answer set  $I$  of  $P$ , there is an answer set  $I_b$  of  $P_b$  such that  $I$  is an answer set of  $P_t \cup I_b$ .*

This result implies that in programs that have a splitting, a rule in  $P_t$  containing in its body an aggregate expression involving only predicates appearing in  $P_b$  can be replaced by a rule in which this aggregate expression is replaced by any of its (classically) equivalent forms. We formally state this result for the case of rules of the form (10).

**Theorem A.5.** *Let  $P$  be a logic program and let  $(P_b, P_t)$  be a splitting of  $P$ . If  $P_t$  contains a rule of the form (10) and  $F$  appears in  $P_b$ , then  $P$  and the program  $P'$  obtained from  $P$  by replacing the rule (10) by the rule (12) have the same answer sets.*

*Proof.* (Sketch) Let  $P'_t$  be the program obtained from  $P_t$  by replacing the rule (10) by the rule (12). It is clear that  $(P_b, P'_t)$  is a splitting of  $P'$ . Let  $I$  be an answer set of  $P$ . By A.4, there is an answer set  $I_b$  of  $P_b$  such that  $I$  is an answer set of  $P_t \cup I_b$ . In particular,  $I$  is an answer set of the program  $I_b \cup \text{ground}(P_t)$ . Let  $Q$  be the program obtained by simplifying the bodies of the rules in  $\text{ground}(P_t)$  as follows. If a conjunct  $c$  in the body of a rule in  $\text{ground}(P_t)$  involves only atoms from the Herbrand base  $HB(P_b)$  of  $P_b$ , we remove  $c$  if  $I_b \models c$ , and we remove the rule, if  $I_b \not\models c$ . Because atoms from  $HB(P_b)$  do not appear in the heads of the rules in  $\text{ground}(P_t)$ ,  $I$  is an answer set of  $Q \cup I_b$ .

We denote by  $Q'$  the program obtained by the same simplification process from  $\text{ground}(P'_t)$ . From the definition of  $P'_t$  it follows that  $Q' = Q$  (indeed, the only difference between  $P_t$  and  $P'_t$  is in the bodies of some rules, in which an aggregate built entirely from the atoms in  $HB(P_b)$  is replaced by a classically equivalent one; thus, both expressions evaluate in the same way under  $I_b$  and the contribution of the corresponding rules to  $Q$  and  $Q'$  in each case is the same). Consequently,  $I$  is an answer set of  $Q' \cup I_b$ . Because atoms from  $HB(P_b)$  do not appear in the heads of the rules in  $\text{ground}(P'_t)$ ,  $I$  is an answer set of  $\text{ground}(P'_t) \cup I_b$  and, because  $(P_b, P'_t)$  is a splitting of  $P'$ , also an answer set of  $P'$ . A similar argument shows that answer sets of  $P'$  are also answer sets of  $P$ .  $\square$

It is clear that the same argument applies to other similar rewritings, for instance, to the one that replaces the rule (9) by the rule

$$H \leftarrow \bigwedge_{0 \leq i < b-1} \text{not } i = \#\text{count}\{X : F(X, \mathbf{Z})\}, G. \quad (13)$$

where, as before, we assume that the variable  $X$  is not a variable in  $\mathbf{Z}$ , and that all variables in  $\mathbf{Z}$  appear in  $G$ .

Theorem A.5 establishes the correctness of the two remaining rewritings implemented by the tool AAgg. The tool is designed to check for splitting and allows the rewriting to take place only if the conditions of A.5 hold.