# Merging of Biomedical Decision Diagrams

Christoph Redl

May 31, 2010

### **Outline**

- 1 Motivation
- 2 Decision Diagrams and Task formally defined
- 3 Implementation and Usage of the Tool
- 4 Case Study
- 5 Summary

### Motivation

### **Decision Diagrams**

Important means for decision making

### Motivation

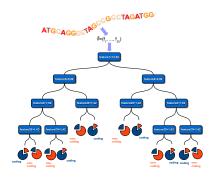
#### **Decision Diagrams**

- Important means for decision making
- Intuitively understandable
- Not only for knowledge engineers

### Motivation

#### **Decision Diagrams**

- Important means for decision making
- Intuitively understandable
- Not only for knowledge engineers
- Common in clinical guidelines



### **Applications**

Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case
- Psychology: diagnosis of personality disorders or mental diseases

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case
- Psychology: diagnosis of personality disorders or mental diseases
- Molecular biology: DNA classification

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case
- Psychology: diagnosis of personality disorders or mental diseases
- Molecular biology: DNA classification
- Applications also in other fields (e.g., economy)

### **Applications**

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case
- Psychology: diagnosis of personality disorders or mental diseases
- Molecular biology: DNA classification
- Applications also in other fields (e.g., economy)

Sometimes we have more than one diagram for a certain purpose

### **Applications**

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case
- Psychology: diagnosis of personality disorders or mental diseases
- Molecular biology: DNA classification
- Applications also in other fields (e.g., economy)

Sometimes we have more than one diagram for a certain purpose

#### Reasons

different opinions

### **Applications**

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case
- Psychology: diagnosis of personality disorders or mental diseases
- Molecular biology: DNA classification
- Applications also in other fields (e.g., economy)

Sometimes we have more than one diagram for a certain purpose

#### Reasons

- different opinions
- randomized machine-learning algorithms

### **Applications**

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case
- Psychology: diagnosis of personality disorders or mental diseases
- Molecular biology: DNA classification
- Applications also in other fields (e.g., economy)

Sometimes we have more than one diagram for a certain purpose

#### Reasons

- different opinions
- randomized machine-learning algorithms
- several diagrams trained in parallel

### **Applications**

- Tumor staginge.g., TNM system(Tumor, Nodes, Metastasis)
- Quantification of the severity of a medical case
- Psychology: diagnosis of personality disorders or mental diseases
- Molecular biology: DNA classification
- Applications also in other fields (e.g., economy)

Sometimes we have more than one diagram for a certain purpose

#### Reasons

- different opinions
- randomized machine-learning algorithms
- several diagrams trained in parallel

**Task:** Incorporate them into one

#### **Decision Procedures**

- Leave source diagrams unchanged
- Classification by the following procedure:
  - 1. Classify a case by each diagram
  - 2. Aggregate the results (e.g., majority voting)

#### **Decision Procedures**

- Leave source diagrams unchanged
- Classification by the following procedure:
  - 1. Classify a case by each diagram
  - 2. Aggregate the results (e.g., majority voting)

#### But

- This does not give us a single diagram
- We lose the property of intuitive understandability

#### **Decision Procedures**

- Leave source diagrams unchanged
- Classification by the following procedure:
  - 1. Classify a case by each diagram
  - 2. Aggregate the results (e.g., majority voting)

#### But

- This does not give us a single diagram
- We lose the property of intuitive understandability
- ⇒ Inappropriate for clinical practice

#### Rule-based systems

- Conversion: diagram to production rules
- One rule for each leaf node

#### Rule-based systems

- Conversion: diagram to production rules
- One rule for each leaf node

#### But

- Not human-readable
- Overhead

### **Therefore**

### New approach

Merge several decision diagrams

- into a standalone one
- without referring to original training data
- without the overhead of translating them into rule-based systems

#### Definition

Let  $\mathcal D$  be a domain and  $\mathcal C$  a set of classes.

#### Definition

Let  $\mathcal{D}$  be a domain and  $\mathcal{C}$  a set of classes.

Then a *classification function* c (*classify*) has the signature  $c: \mathcal{D} \to \mathcal{C}$ 

#### Definition

Let  $\mathcal{D}$  be a domain and  $\mathcal{C}$  a set of classes.

Then a classification function c (classify) has the signature  $c:\mathcal{D}\to\mathcal{C}$  and can also be represented as decision diagram

$$D = \langle V, E, l_c, l_e, \mathcal{M}^Q \rangle,$$

where

V ... set of nodes and

 $E \subseteq V \times V \dots$  set of directed edges (s.t. (V, E) is acyclic and D has a unique root node)

#### Definition

Let  $\mathcal{D}$  be a domain and  $\mathcal{C}$  a set of classes.

Then a classification function c ( $\underline{c}$ lassify) has the signature  $c:\mathcal{D}\to\mathcal{C}$  and can also be represented as decision diagram

$$D = \langle V, E, l_c, l_e, \mathcal{M}^Q \rangle,$$

where

V ... set of nodes and

 $E \subseteq V \times V$  ... set of directed edges (s.t. (V, E) is acyclic and D has a unique root node)

$$l_c: Leafs(V) \rightarrow \mathcal{C}$$

#### Definition

Let  $\mathcal{D}$  be a domain and  $\mathcal{C}$  a set of classes.

Then a classification function c (classify) has the signature  $c:\mathcal{D}\to\mathcal{C}$  and can also be represented as decision diagram

$$D = \langle V, E, l_c, l_e, \mathcal{M}^Q \rangle,$$

#### where

V ... set of nodes and

 $E \subseteq V \times V \dots$  set of directed edges (s.t. (V,E) is acyclic and D has a unique root node)

 $l_c: Leafs(V) \rightarrow \mathcal{C}$ 

 $l_e: E o \Sigma^{\mathcal Q}$  where  $\Sigma^{\mathcal Q}$  is the set of syntactically correct expressions in  $\mathcal Q$ 

#### Definition

Let  $\mathcal{D}$  be a domain and  $\mathcal{C}$  a set of classes.

Then a classification function c (classify) has the signature  $c:\mathcal{D}\to\mathcal{C}$  and can also be represented as decision diagram

$$D = \langle V, E, l_c, l_e, \mathcal{M}^Q \rangle,$$

#### where

V ... set of nodes and

 $E \subseteq V \times V$  ... set of directed edges (s.t. (V, E) is acyclic and D has a unique root node)

 $l_c: Leafs(V) \rightarrow \mathcal{C}$ 

 $l_e: E \to \Sigma^Q$  where  $\Sigma^Q$  is the set of syntactically correct expressions in Q  $\mathcal{M}^Q$ : The meaning function of a query language Q is of kind

$$\mathcal{M}^{\mathcal{Q}}:\Sigma^{\mathcal{Q}} o 2^{\mathcal{D}}$$

Assume that *F* is a first-order like query language.

### Example

Let 
$$\mathcal{D} = \{1, 2, \dots, 15\}, \mathcal{C} = \{\textit{prime}, \textit{not prime}\},$$

Assume that *F* is a first-order like query language.

Example

Let 
$$\mathcal{D} = \{1, 2, \dots, 15\}, \mathcal{C} = \{\textit{prime}, \textit{not prime}\},$$
 and

$$D_F = \left\langle \{r, p, n\}, \{(r, p), (r, n)\}, l_c, l_e, \mathcal{M}^F \right\rangle$$

with

$$l_c(p) = prime, l_c(n) = not prime$$

Assume that F is a first-order like query language.

Example

Let 
$$\mathcal{D} = \{1, 2, \dots, 15\}, \mathcal{C} = \{\textit{prime}, \textit{not prime}\},$$
 and

$$D_F = \left\langle \{r, p, n\}, \{(r, p), (r, n)\}, l_c, l_e, \mathcal{M}^F \right\rangle$$

with

$$l_c(p) = prime, l_c(n) = not prime$$

and

$$l_e((r,p)) = f_1 = \nexists x, y : x > 1 \land y > 1 \land x \cdot y = z$$
  
 $l_e((r,n)) = f_2 = \exists x, y : x > 1 \land y > 1 \land x \cdot y = z$ 

Assume that F is a first-order like query language.

Example

Let 
$$\mathcal{D} = \{1, 2, \dots, 15\}, \mathcal{C} = \{\textit{prime}, \textit{not prime}\},$$
 and

$$D_F = \left\langle \{r, p, n\}, \{(r, p), (r, n)\}, l_c, l_e, \mathcal{M}^F \right\rangle$$

with

$$l_c(p) = prime, l_c(n) = not prime$$

and

$$l_e((r,p)) = f_1 = \nexists x, y : x > 1 \land y > 1 \land x \cdot y = z$$
  
 $l_e((r,n)) = f_2 = \exists x, y : x > 1 \land y > 1 \land x \cdot y = z$ 

with the meaning function  $\mathcal{M}^F(f) = \{n \in \mathcal{D} : I_n(f) = true\}$  where  $I_n$  is the first-order interpretation  $I = \{z \leftarrow n\}$ .

Assume that *F* is a first-order like query language.

Example

Let 
$$\mathcal{D} = \{1, 2, \dots, 15\}, \mathcal{C} = \{\textit{prime}, \textit{not prime}\},$$
 and

$$D_F = \left\langle \{r, p, n\}, \{(r, p), (r, n)\}, l_c, l_e, \mathcal{M}^F \right\rangle$$

with

$$l_c(p) = prime, l_c(n) = not prime$$

and

$$l_e((r,p)) = f_1 = \nexists x, y : x > 1 \land y > 1 \land x \cdot y = z$$
  
$$l_e((r,n)) = f_2 = \exists x, y : x > 1 \land y > 1 \land x \cdot y = z$$

with the meaning function  $\mathcal{M}^F(f) = \{n \in \mathcal{D} : I_n(f) = true\}$  where  $I_n$  is the first-order interpretation  $I = \{z \leftarrow n\}$ .

**Example evaluation:** We want to classify 7 and start in root node r.

Assume that *F* is a first-order like query language.

Example

Let  $\mathcal{D} = \{1, 2, \dots, 15\}, \mathcal{C} = \{\textit{prime}, \textit{not prime}\}, \text{ and }$ 

$$D_F = \left\langle \{r, p, n\}, \{(r, p), (r, n)\}, l_c, l_e, \mathcal{M}^F \right\rangle$$

with

$$l_c(p) = prime, l_c(n) = not prime$$

and

$$l_e((r,p)) = f_1 = \nexists x, y : x > 1 \land y > 1 \land x \cdot y = z$$
  
$$l_e((r,n)) = f_2 = \exists x, y : x > 1 \land y > 1 \land x \cdot y = z$$

with the meaning function  $\mathcal{M}^F(f) = \{n \in \mathcal{D} : I_n(f) = true\}$  where  $I_n$  is the first-order interpretation  $I = \{z \leftarrow n\}$ .

**Example evaluation:** We want to classify 7 and start in root node r.  $7 \in \mathcal{M}^F(f_1)$  but  $7 \notin \mathcal{M}^F(f_2)$ , therefore we choose edge (r,p) and end up in node p.  $l_c(p) = prime$ .

### **Task Definition**

#### Definition

The set of all decision diagrams over domain  $\mathcal D$  and classes  $\mathcal C$  is denoted as  $\mathbb T_{\mathcal D,\mathcal C}.$ 

### **Task Definition**

#### Definition

The set of all decision diagrams over domain  $\mathcal D$  and classes  $\mathcal C$  is denoted as  $\mathbb T_{\mathcal D,\mathcal C}.$ 

#### Definition

An *n*-ary decision diagram merging operator

$$\circ^n: \underbrace{ \exists_{\mathcal{D},\mathcal{C}} \times \exists_{\mathcal{D},\mathcal{C}} \times \cdots \times \exists_{\mathcal{D},\mathcal{C}}}_{\textit{n times}} \to \exists_{\mathcal{D},\mathcal{C}}$$

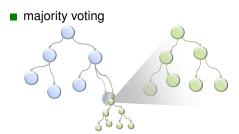
maps n input classifiers (over  $\mathcal{D}$  and  $\mathcal{C}$ ) onto a new diagram.

### **Task Definition**

Example merging operators

majority voting

#### Example merging operators



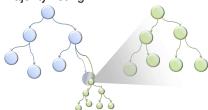
#### Example merging operators

■ majority voting

■ user preferences

#### Example merging operators

majority voting



user preferences

"In doubt, classify it as X rather than Y." (Sometimes one wrong decision is more serious than the other one)

We can make several assumptions about the input diagrams.

We can make several assumptions about the input diagrams.

■ It is a tree or a general acyclic graph?

We can make several assumptions about the input diagrams.

- It is a tree or a general acyclic graph?
- What is the maximum node degree?

We can make several assumptions about the input diagrams.

- It is a tree or a general acyclic graph?
- What is the maximum node degree?
- Is the variable ordering along a path fixed?
- etc.

We can make several assumptions about the input diagrams.

- It is a tree or a general acyclic graph?
- What is the maximum node degree?
- Is the variable ordering along a path fixed?
- etc.

Operators can also assert such properties for their result.

Bad approach: Develop merging operators for the most general task variant, i.e.,

- general decision diagrams
- without any assumptions about node degrees or variable ordering

Bad approach: Develop merging operators for the most general task variant, i.e.,

- general decision diagrams
- without any assumptions about node degrees or variable ordering

This leads to complicated operators

Bad approach: Develop merging operators for the most general task variant, i.e.,

- general decision diagrams
- without any assumptions about node degrees or variable ordering

This leads to complicated operators

Good approach: Reduce complicated diagrams to simple ones, i.e.,

Bad approach: Develop merging operators for the most general task variant, i.e.,

- general decision diagrams
- without any assumptions about node degrees or variable ordering

This leads to complicated operators

Good approach: Reduce complicated diagrams to simple ones, i.e.,

first convert acyclic graphs into trees

Bad approach: Develop merging operators for the most general task variant, i.e.,

- general decision diagrams
- without any assumptions about node degrees or variable ordering

This leads to complicated operators

Good approach: Reduce complicated diagrams to simple ones, i.e.,

- first convert acyclic graphs into trees
- reduce node degrees to make them binary

Bad approach: Develop merging operators for the most general task variant, i.e.,

- general decision diagrams
- without any assumptions about node degrees or variable ordering

This leads to complicated operators

Good approach: Reduce complicated diagrams to simple ones, i.e.,

- first convert acyclic graphs into trees
- reduce node degrees to make them binary
- order the nodes

Therefore we implement two sets of operators

Therefore we implement two sets of operators

Unary conversion operators

for preparing and simplifying the input

Therefore we implement two sets of operators

#### Unary conversion operators

- for preparing and simplifying the input
- eliminating redundancy from the result

Therefore we implement two sets of operators

#### Unary conversion operators

- for preparing and simplifying the input
- eliminating redundancy from the result

#### Merging operators

- $\blacksquare$  *n*-ary merging operators (n > 1) for actual merging task
- they are much simpler if we can make assumptions about the input

- Implemented as plugin for dlvhex
- Uses and extends the mergingplugin

- Implemented as plugin for dlvhex
- Uses and extends the mergingplugin
- Diagrams are given either in a human-readable format like DOT if created manually

- Implemented as plugin for dlvhex
- Uses and extends the mergingplugin
- Diagrams are given either in a human-readable format like DOT if created manually
- ... or in the file format produced by a machine-learning tool if trained automatically

- Implemented as plugin for dlvhex
- Uses and extends the mergingplugin
- Diagrams are given either in a human-readable format like DOT if created manually
- ... or in the file format produced by a machine-learning tool if trained automatically
- Operators are written in C++ (user-defined ones can be added)

- Implemented as plugin for dlvhex
- Uses and extends the mergingplugin
- Diagrams are given either in a human-readable format like DOT if created manually
- ... or in the file format produced by a machine-learning tool if trained automatically
- Operators are written in C++ (user-defined ones can be added)

Decision diagrams are hierarchical: How to handle them with dlvhex?

- Implemented as plugin for dlvhex
- Uses and extends the mergingplugin
- Diagrams are given either in a human-readable format like DOT if created manually
- ... or in the file format produced by a machine-learning tool if trained automatically
- Operators are written in C++ (user-defined ones can be added)

Decision diagrams are hierarchical: How to handle them with dlvhex?

- convert them into a logic program
- use predicates like innernode(X), edge(X, Y, C), etc.

- Implemented as plugin for dlvhex
- Uses and extends the mergingplugin
- Diagrams are given either in a human-readable format like DOT if created manually
- ... or in the file format produced by a machine-learning tool if trained automatically
- Operators are written in C++ (user-defined ones can be added)

#### Decision diagrams are hierarchical: How to handle them with dlvhex?

- convert them into a logic program
- use predicates like innernode(X), edge(X, Y, C), etc.

#### Therefore the plugin consists of

- 1. A tool for conversion of decision diagrams
- 2. Predefined operators

#### **Steps**

■ Convert your diagrams into logic programs using graphconverter

- Convert your diagrams into logic programs using graphconverter
- Define your task "diags.mp" which references the diagrams

- Convert your diagrams into logic programs using graphconverter
- Define your task "diags.mp" which references the diagrams
- 3 Run the merging plan compiler (mpcompiler) on this input

- Convert your diagrams into logic programs using graphconverter
- 2 Define your task "diags.mp" which references the diagrams
- 3 Run the merging plan compiler (mpcompiler) on this input
- 4 Load the result into dlvhex

- Convert your diagrams into logic programs using graphconverter
- 2 Define your task "diags.mp" which references the diagrams
- 3 Run the merging plan compiler (mpcompiler) on this input
- 4 Load the result into dlvhex

- Convert your diagrams into logic programs using graphconverter
- Define your task "diags.mp" which references the diagrams
- 3 Run the merging plan compiler (mpcompiler) on this input
- Load the result into dlvhex
- 5 Convert the result into desired output format

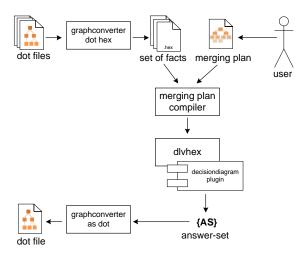
#### **Steps**

- Convert your diagrams into logic programs using graphconverter
- Define your task "diags.mp" which references the diagrams
- 3 Run the merging plan compiler (mpcompiler) on this input
- 4 Load the result into dlvhex
- 5 Convert the result into desired output format

#### Typical call

- Conversion of input diagrams X:
  - \$ graphconverter dot hex < diagX.dot > diagX.hex
- Merging:
  - \$ dlvhex --merging diags.mp > out.as
- Conversion of output diagrams:
  - \$ graphconverter as dot < out.as > out.dot

## Implementation



## Advantages of the Framework

 Framework is considered as an extension to the general belief merging framework

## Advantages of the Framework

- Framework is considered as an extension to the general belief merging framework
- Intention is not to provide high-performance throughput, but an easy-to-use rapid prototyping tool

## Advantages of the Framework

- Framework is considered as an extension to the general belief merging framework
- Intention is not to provide high-performance throughput, but an easy-to-use rapid prototyping tool
- Different operators can be tried out without reimplementing them

# Advantages of the Framework

- Framework is considered as an extension to the general belief merging framework
- Intention is not to provide high-performance throughput, but an easy-to-use rapid prototyping tool
- Different operators can be tried out without reimplementing them
- Input diagrams, training set and merging strategies can be edited quickly without redoing manual remerging

## Advantages of the Framework

- Framework is considered as an extension to the general belief merging framework
- Intention is not to provide high-performance throughput, but an easy-to-use rapid prototyping tool
- Different operators can be tried out without reimplementing them
- Input diagrams, training set and merging strategies can be edited quickly without redoing manual remerging
- When best settings have been found, one can consider application-specific reimplementation due to performance reasons

#### **Application Scenario**

- DNA classification
- Given a sequence  $S \in \{A, C, G, T\}^*$
- Is it protein-coding or junk DNA?  $\{C, N\}$

#### **Application Scenario**

- DNA classification
- Given a sequence  $S \in \{A, C, G, T\}^*$
- Is it protein-coding or junk DNA? {C,N}

### Approach

- Compute a 20-dimensional feature vector
- Features are statistically motivated (incorporating biological knowledge)

#### **Application Scenario**

- DNA classification
- Given a sequence  $S \in \{A, C, G, T\}^*$
- Is it protein-coding or junk DNA? {C,N}

### **Approach**

- Compute a 20-dimensional feature vector
- Features are statistically motivated (incorporating biological knowledge)

#### Examples

■ Triplet ATG more frequent in coding sequences

#### **Application Scenario**

- DNA classification
- Given a sequence  $S \in \{A, C, G, T\}^*$
- Is it protein-coding or junk DNA? {C,N}

### **Approach**

- Compute a 20-dimensional feature vector
- Features are statistically motivated (incorporating biological knowledge)

#### Examples

- Triplet *ATG* more frequent in coding sequences
- In coding sequences, the base at first codon position is frequently a purine base

#### The role of the merging framework

- Train multiple decision trees for the task
- Merge them afterwards

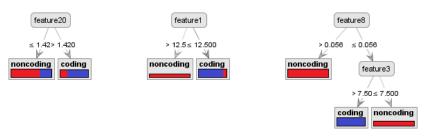
#### The role of the merging framework

- Train multiple decision trees for the task
- Merge them afterwards

### Advantages

- Better accuracy by using different algorithms
- Parallel computing
- Try out different combinations of training algorithms; no need for manual remerging after each changes

Suppose we have 3 trees trained with different algorithms and training sets



Each has accuracy ≈50%

Merging operator in use (Stephen Salzberg)

1 Leafs do not only store classification but also frequency distribution

#### Merging operator in use (Stephen Salzberg)

- Leafs do not only store classification but also frequency distribution
- Insert second tree into each leaf of the first one (iterate for more than two trees)

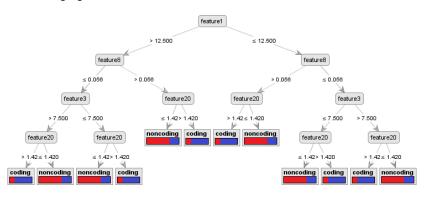
#### Merging operator in use (Stephen Salzberg)

- Leafs do not only store classification but also frequency distribution
- Insert second tree into each leaf of the first one (iterate for more than two trees)
- 3 Weight trees according to size of training set used during training

#### Merging operator in use (Stephen Salzberg)

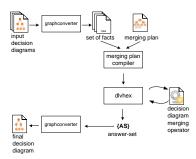
- 1 Leafs do not only store classification but also frequency distribution
- Insert second tree into each leaf of the first one (iterate for more than two trees)
- 3 Weight trees according to size of training set used during training
- Recompute classification for each (new) leaf according to new distribution

#### After merging

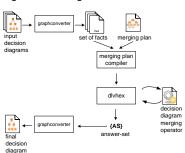


Accuracy  ${\approx}65\%$ 

■ Task: Incorporate several decision diagrams into one



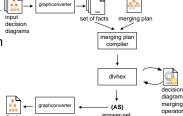
- Task: Incorporate several decision diagrams into one
- Approach:
  - implement a plugin that allows decision diagram handling with dlvhex
  - provide predefined operators



- Task: Incorporate several decision diagrams into one
- Approach:
  - implement a plugin that allows decision diagram handling with dlvhex

decision diagram

- provide predefined operators
- input diagrams ⇒ graphconverter ⇒
  ⇒ dlvhex ⇒ answer set ⇒
  - $\Rightarrow$  graphconverter  $\Rightarrow$  output diagram



- Task: Incorporate several decision diagrams into one
- Approach:
  - implement a plugin that allows decision diagram handling with dlvhex
  - provide predefined operators
- input diagrams ⇒ graphconverter ⇒
  ⇒ dlvhex ⇒ answer set ⇒
  - ⇒ graphconverter ⇒ output diagram

### Advantages

- Easy-to-use rapid prototyping tool
- No manual incorporation of diagrams
- Allows changes of the scenario without redoing routine tasks



set of facts

merging plan

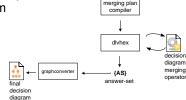
graphconverter

decision diagrams

- Task: Incorporate several decision diagrams into one
- Approach:
  - implement a plugin that allows decision diagram handling with dlvhex
  - provide predefined operators
- input diagrams ⇒ graphconverter ⇒
  - $\Rightarrow$  dlvhex  $\Rightarrow$  answer set  $\Rightarrow$
  - $\Rightarrow$  graphconverter  $\Rightarrow$  output diagram

### Advantages

- Easy-to-use rapid prototyping tool
- No manual incorporation of diagrams
- Allows changes of the scenario without redoing routine tasks
- ⇒ dd-plugin = mergingplugin + dd processing with dlvhex + dd operators



set of facts

graphconverter

decision diagrams