Université de Strasbourg

Apprentissage Collaboratif - M2 (SDSC et SDIA)

TP 1 and Project: Combining knowledge-based with data-driven approaches

Objective: The objective of this TP is to apply stream reasoning on a particular scenario. From this application, the idea is also to identify the advantages of this approach (knowledge-based), as well as to identify the limitations of this approach and how to combine it with other approaches (data-driven) to deal with them.

An illustrative case study is described below to highlight how stream reasoning can be applied to detect abnormal situations in an industrial scenario.

Illustrative case study description: The case study is based on a manufacturing production line (Figure 1), named `PL1`, composed of four machines: `M1`, `M2`, `M3` and `M4`. These machines are equipped with sensors on different components. The sensors collect data on the properties described in Figure 2.

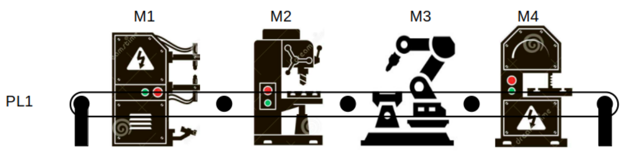

Figure 1: Production line.

Several abnormal situations that leads to failures in this scenario are defined by experts. Each of them is expressed as a set of constraints. We focus on situations covering the following types of failures: machines malfunctions and global malfunctions of the production line including hydraulic oil leakage and cooling system filter obstructions. The defined abnormal situations are shown in Figure 3, with a short description of what they represent. The constraints concerned by the abnormal situations are described along the associated properties in Figure 2. The abnormal situations are briefly describe below:

Set of constraints $\mathcal{C}$							
ID	Properties	Restriction	Device	ID	Properties	Restriction	Device
c_1	`Oil temp.`	$> 40°C$	M_1	c_{13}	`Power output`	< 500 kW	PL_1
c_2	`Oil temp.`	$> 60°C$	M_1	c_{14}	`Power output`	< 200 kW	PL_1
c_3	`Transformer temp.`	$> 45°C$	M_1T_1	c_{15}	`Conv. water temp.`	$> 60°C$	M_3Cv_1
c_4	`Controller temp.`	$> 40°C$	M_1Ct_1	c_{16}	`Conv. water temp.`	$> 80°C$	M_3Cv_1
c_5	`Generator curr.`	< 800 A	M_1G_1	c_{17}	`Trans. grid temp.`	$< 35°C$	M_3T_1
c_6	`Platform temp.`	$< 35°C$	PL_1	c_{18}	`Generator temp.`	$> 45°C$	M_3G_1
c_7	`Platform temp.`	$> 40°C$	PL_1	c_{19}	`Converter temp.`	$> 60°C$	M_3Cv_1
c_8	`Gearbox temp.`	$> 40°C$	M_2GB_1	c_{20}	`Converter temp.`	$> 80°C$	M_3Cv_1
c_9	`Gearbox temp.`	$> 60°C$	M_2GB_1	c_{21}	`Rotor speed`	< 200 rpm	M_4R_1
c_{10}	`Generator speed`	< 500 rpm	M_2G_1	c_{22}	`Rotor speed`	< 100 rpm	M_4R_1
c_{11}	`Environment temp.`	$< 25°C$	PL_1	c_{23}	`Rotor Pitch angle`	$< 5°$	M_4R_1
c_{12}	`Power output`	> 2000 kW	PL_1				

Figure 2: Constraints definition.

Set of situations $\mathcal{S}$					
Situation	Constraint ($\mathcal{T}$)	Description	Situation	Constraint ($\mathcal{T}$)	Description
s_1	c_1,c_3,c_4,c_5,c_6	M_1 oil leakage	s_8	c_6,c_{17},c_{19}	M_3Cv_1 malfunction
s_2	$\mathbf{c_1},c_2,c_3,c_4,c_5,c_6$	M_1 oil leakage	s_9	$c_6,c_{17},\mathbf{c_{19}},c_{20}$	M_3Cv_1 malfunction
s_3	c_6,c_8,c_{10},c_{11}	Increase M_2 oil temp.	s_{10}	$\mathbf{c_{15}},c_{16},c_{17},c_{18},c_{19}$	M_3Cv_1 malfunction
s_4	$c_6,\mathbf{c_8},c_9,c_{10},c_{11}$	M_2 oil temp.	s_{11}	c_{12},c_{21},c_{23}	M_4R_1 malfunction
s_5	$c_7,\mathbf{c_8},c_9,c_{10},c_{11}$	Increase M_2 oil temp.	s_{12}	$c_{12},\mathbf{c_{21}},c_{22},c_{23}$	M_4R_1 malfunction
s_6	c_{15},c_{17},c_{18}	M_3 filter obstruction	s_{13}	c_{13},c_{21},c_{23}	PL global malfunction
s_7	$\mathbf{c_{15}},c_{16},c_{17},c_{18}$	M_3 filter obstruction	s_{14}	$\mathbf{c_{13}},c_{14},c_{21},c_{23}$	PL global malfunction

Figure 3: Situations and their constraints (the constraints that are implied by other constraints are in bold face)

- The S1 situation is associated with an oil leak on the M1 machine. This situation happens when the constraints OilTemp $>$ 40°C (c_1), T_temp $>$ 45°C (c_3), C_temp $>$ 40°C (c_4), G_current $>$ 800 A (c_5) and P_temp $<$ 35°C (c_6) are satisfied at least once in a 20 second period of time.
- The S2 situation is a more severe form of S1, also associated with an oil leak on the M1 machine. The S2 situation happens when the constraints OilTemp $>$ 60°C (c_2), T_temp $>$ 45°C (c_3), C_temp $>$ 40°C (c_4), G_current $>$ 800 A (c_5) and P_temp $<$ 35°C (c_6) are satisfied at least once in a 20 second period of time. S2 is more severe than S1 because the oil temperature exceeds 60°C. More urgent actions are required to avoid the failure of the M1 machine.
- The S3 situation is associated with an increase in the oil temperature of the M2 machine. This situation happens when the constraints P_temp $<$ 35°C (c_6), GB_temp $>$ 40°C (c_8), G_speed $<$ 500 rpm (c_{10}) and E_temp $<$ 25°C (c_{11}) are satisfied at least once in a 25 second period of time.
- As with S1 and S2, the S4 situation is a more severe version of S3. The S4 situation happens when the constraints P_temp $<$ 35°C (c_6), GB_temp $>$ 60°C (c_9), G_speed $<$ 500 rpm (c_{10}) and E_temp $<$ 25°C (c_{11}) are satisfied at least once in a 25 second period of time. Since the GB temperature exceeds 60°C, immediate actions are required to avoid the failure of the M2 machine.
- The S5 situation is an even more severe situation. It happens when the constraints P_temp $>$ 40°C (c_7), GB_temp $>$ 60°C (c_9), G_speed $<$ 500 rpm (c_{10}) and E_temp $<$ 25°C (c_{11}) are satisfied at least once in a 25 second period of time. S5 is more severe than S3 and S4 because the GB temperature exceeds 60°C and the platform temperature exceeds 40°C. Due to its degree of severity, drastic actions must be taken to avoid the failure of the M2 machine.
- The S6 situation is associated with a cooling system failure of the M3 machine. This situation happens when the constraints C_Wtemp $>$ 60°C (c_{15}), TG_temp $<$ 35°C (c_{17}) and G_temp $>$ 45°C (c_{18}) are satisfied at least once in a 15 second period of time.
- As the S6 situation, the S7 situation is associated with a cooling system failure of the M3 machine. The S7 situation happens when the constraints C_Wtemp $>$ 80°C (c_{16}), TG_temp $<$ 35°C (c_{17}) and G_temp $>$ 45°C (c_{18}) are satisfied at least once in a 15 second period of time. S7 is more severe than S6 because the converter water temperature exceeds 80°C. In the same way, that the most severe situations above the actions to take must be prompt to avoid in this case the failure of the cooling system of the M3 machine.
- The S8 situation is associated with a malfunction of the M3 machine. This situation happens when the constraints P_temp $<$ 35°C (c_6), TG_temp $<$ 35°C (c_{17}) and Conv_temp $>$ 60°C (c_{19}) are satisfied at least once in a 20 second period of time.
- The S9 situation is more severe than S8. The S9 situation happens when the constraints P_temp $<$ 35°C (c_6), TG_temp $<$ 35°C (c_{17}) and Conv_temp $>$ 80°C (c_{20}) are satisfied in at least once in a 20 second period of time. In this case, as the converter temperature exceeds 80°C, more urgent actions are required to avoid the failure of the M3 machine.
- The S10 situation is associated with a malfunction of the M3 machine as the two previous situations. The S10 situation happens when the constraints C_Wtemp $>$ 80°C (c_{16}), TG_temp $<$ 35°C (c_{17}), G_temp $<$ 45°C (c_{18}) and Conv_temp $>$ 60°C (c_{19}) are satisfied at least once in a 20 second period of time. S10 is more severe than S8 and S9 because the GB temperature exceeds 60°C and the platform temperature exceeds 40°C. In order to avoid a failure of M3, immediate actions must be taken.
- The S11 situation is associated with a malfunction of the M4 machine. This situation happens when the constraints PowerOutput $>$ 2000 KW (c_{12}), R_speed $<$ 200 rpm (c_{21}) and R_Pangle $<$ 5° (c_{23}) are satisfied at least once in a 30 second period of time.
- The S12 situation is associated with the same malfunction as the S11 situation. The S12 situation happens when the constraints PowerOutput $>$ 2000 KW (c_{12}), R_speed $<$ 100 rpm (c_{22}) and R_Pangle $<$ 5° (c_{23}) are satisfied in least once in a 30 second period of time. S12 is more severe than S11 because the rotor speed of the M4 machine is lower than 100 rpm. It requires immediate action to prevent the failure of the M4 machine.
- The S13 situation is associated with a global malfunction of the PL1 production line. This situation happens when the constraints PowerOutput $<$ 500 KW (c_{13}), R_speed $<$ 200 rpm (c_{21}) and R_Pangle $<$ 5° (c_{23}) are satisfied at least once in a 35 second period of time.
- As the S13 situation, the S14 situation is also associated with a malfunction of the PL1 production line. The S14 situation happens when the constraints PowerOutput $<$ 200 KW (c_{14}), R_speed $<$ 200 rpm (c_{21}) and R_Pangle $<$ 5° (c_{23}) are satisfied at least once in a 35 second period of time. S14 is more severe version of S13 because the power output of the PL1 production line is lower than 200 KW.

Exercises:
Work in groups of 3 students. The idea is not to parallelize the tasks but to have exchanges and discussions during the carrying out of each of the tasks.

1. From the case study description:
   (a) Identify which concepts, relations and instances are relevant to build the ontology and the data streams.
   (b) Build the ontology, you can use Protégé[1], OWLAPI[2] (it is a Java API and reference implmentation for creating and manipulating OWL Ontologies) or Owlready2[3] (it is a package for ontology-oriented programming in Python).
   (c) Take a look at the SSN[4] ontology.
2. Take some time to familiarize yourself with the source code. More particularly:
   (a) Download the project available on the course Moodle site.
   (b) Read the **readme.md** file to execute the project.
   (c) Identify what is the purpose of each of the classes as well as how they are used for the execution of the stream reasoner.
   (d) Execute the example provided in the main class which may help you to understand how the project works (a more detailed description is given below).
   (e) Write the C-SPARQL[5][6] queries for the detection of the situations described above. To do this, you must also generate the data streams for the properties involved in those situations.
   (f) Modify the code so that when a situation is detected it is added as an instance to the ontology.
3. Perform an analysis of the approach applied to the selected case study (you can also think in general and not only in the selected case study), highlighting the advantages and limitations of the approach as well as how to deal with the limitations. Some points to be considered (this list is not exhaustive):
   - Implementation/Set up of the approach (need of expert knowledge, query formulation, . . .)
   - Scalability (size of the ontology, size of the data streams, speed of data generation/collection, number of queries, . . .)

To be submitted in the Moodle site of the course (all in one compressed file `NOM1-NOM2-NOM3.tar.gz` or `NOM1-NOM2-NOM3.zip`) **Deadline 18/12 at 23:59 (CET)**:

- source code (just the modified `.java` files)
- a report (4 pages maximum): *(1)* Introduction briefly explaining the application scenario. *(2)* Description of the developed ontology. *(3)* Description of the RDF streams generated as well as the required C-SPARQL queries. *(4)* Analysis of the limitations and how to potentially overcome them. *(5)* Possible applications of this approach in a collaborative learning framework. Points raised in the lectures can be discussed here in more detail.

[1] https://protege.stanford.edu/
[2] https://github.com/owlcs/owlapi
[3] https://owlready2.readthedocs.io/en/v0.47/
[4] https://www.w3.org/TR/vocab-ssn/
[5] https://github.com/streamreasoning/CSPARQL-engine
[6] RSP4J is a more recent version and it is available here https://github.com/streamreasoning/rsp4j

Description of a C-SPARQL query: Here a particular C-SPARQL query is described. We must emphasize that for this case study all the data streams of the properties (`ObservableProperties`) defined in table 2 are generated using the *RDFStream* class provided by C-SPARQL to generate RDF streams. Queries can also be executed on data streams that are generated and published by other systems or users, however it is necessary to know the structure of the RDF stream to be able to execute queries on them.

The C-SPARQL query presented in Listing 1 has the purpose of detecting a situation `S`. The query name is registered on line 1 and prefixes used in the query are declared on lines 2 and 3. The query is executed on a RDF stream that correspond to the property `C_Wtemp` in a time window of 15 seconds, sliding the window by 5 seconds (line 5). The chosen time window is arbitrary and can be changed as desired. This query produces pairs of values (line 4): the machine name (`?m`) and the production line of which it is a part of (`?pl`). In order to obtain the production line to which the machine belongs, we indicate in the query that the C-SPARQL engine must use our ontological model as background knowledge (line 6). Line 8 enables to obtain the production line to which the machine belongs. To get the observation's values, `?o1` individuals are respectively bound with the data values `?v1` through the appropriate properties (`sosa:madeObservation` and `sosa:hasSimpleResult`) (line 10-11). Finally, the list of output pairs are filtered out to include only the ones where the observation's values satisfy the restrictions in the `FILTER` clause (line 12-14).

```
1   REGISTER QUERY S6-detection AS
2   PREFIX : <http:<@//@>semanticweb.org/Ontology#>
3   PREFIX sosa: <http:<@//@>www.w3.org/ns/sosa/>
4   SELECT ?m ?pl
5   FROM STREAM <Stream_C_Wtemp> [RANGE 15s STEPS 5s]
6   FROM <http:<@//@>semanticweb.org/Ontology#>
7   WHERE {
8      ?m           :isPartOf              ?pl .
9      ?m           sosa:hosts             sosa:S_C_Wtemp .
10     :S_C_Wtemp   sosa:madeObservation   ?o1 .
11     ?o1          sosa:hasSimpleResult   ?v1 .
12  FILTER (
13     ?v1 > 60.0 ) .
14  }
```

Listing 1: Example of a C-SPARQL query.