Architecture of a Production System

Expert Systems Tools: History

- 1943: Post production rules
- 1957: GPS started (Newell / Simon)
- 1958: LISP
- 1973: MYCIN (expert system, medical diagn.)
- 1977: OPS
- 1979: Rete algorithm
- 1985: CLIPS (NASA)
- 1995: JESS (Sandia Corp.)

Early History of Production Systems

Production System for „P minus Q“

Working Memory: (START) (P 5) (Q 3)
Variables: X1 X2
Production Rules:

R1: (START)
    → DELETE((START)), ADD((COUNT 0))
R2: (COUNT X1) & (P X2) & (Q X2))
    → PRINT(X1), STOP
R3: (COUNT X1) & (Q X2)
    → REPLACE (X1, SUCC(X1), (COUNT X1)),
        REPLACE (X2, SUCC(X2), (Q X2))

DB1: (COUNT 0) (P 5) (Q 3) <R1>
DB2: (COUNT 1) (P 5) (Q 4) <R3>
DB3: (COUNT 2) (P 5) (Q 5) <R3>
    “2” <R2>
JESS: Example „counting“

```prolog
(deffacts init (START) (P 5) (Q 3))

(defrule R1
  ?x <- (START)
  =>
  (retract ?x)
  (assert (COUNT 0)))

(defrule R2
  (COUNT ?x1)
  (P ?x2)
  (Q ?x2)
  =>
  (printout t "RESULT: " ?x1 crlf)
  (halt))
```

```
(printfn t "RESULT: 2")
(explicit termination!)
```

Fundamental Terms of Production Systems

Each production (synonym: inference rule) consists of two parts:

<table>
<thead>
<tr>
<th>A → B</th>
<th>A ← B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premise</td>
<td>Condition</td>
</tr>
<tr>
<td>Antecedence</td>
<td>Left Side (LHS)</td>
</tr>
<tr>
<td>Evidence</td>
<td>If - part</td>
</tr>
<tr>
<td>Right Side (RHS)</td>
<td>Action</td>
</tr>
</tbody>
</table>

Inference rules with evidence values

(A → B) (EV₁) → Z, EV₃
(B ∨ C) → Z, EV₄
(D → E) → Z, EV₅
(F → G, EV₆)
(G → H, EV₇)
(H → I, EV₈)

Evaluation Modes for Production Systems

Forward Chaining

1. Determine set K of all rules whose premises are fulfilled by WM
2. Choose a rule R out of K using the conflict resolution strategy
3. WM := evaluation result obtained by applying R on WM

Backward Chaining

1. Determine set K of all rules whose conclusion can be unified with P
2. Choose a rule R out of K using the conflict resolution strategy
3. If premise of R not in WM, perform backward chaining (premise of R)

Multiple Derivation as Multiple And/Or-Tree

Productions are evaluated over a „database“. „Database“ is also referred to as „Working Memory“ (WM) or in Cognitive Psychology as „Short Term Memory“ (STM).

There are two different evaluation modes for productions:
Pros and Cons of Production Systems

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expressive power of representation language</td>
<td>Extensibility</td>
<td>Very limited syntax and semantics</td>
</tr>
<tr>
<td>Extensibility</td>
<td>Static analysis of control flow</td>
<td>Indirect rule interaction with working memory</td>
</tr>
<tr>
<td>Static analysis of control flow</td>
<td>Realization of complex control structures</td>
<td>Programming with side effects</td>
</tr>
<tr>
<td>Machine Understanding</td>
<td>Modular Knowledge Representation</td>
<td>Consistency check</td>
</tr>
<tr>
<td>Knowledge acquisition</td>
<td>Explanation component</td>
<td>Control structures</td>
</tr>
<tr>
<td>Error correction</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Use of Meta-Rules in Rule-Based Systems

Meta-knowledge is knowledge that refers to other knowledge units in the knowledge base. Meta-rules refer to form and content of the productions that are stored in the knowledge base.

(1) Meta-rules for conflict resolution
Consider an expert system with the following rules that gives advise in the case of a chemical contamination:
R1: If contamination caused by sulfuric acid, use procedure A.
R2: If contamination caused by sulfuric acid, use procedure B.

Forward chaining can lead to conflict set (R1, R2).
Assume the knowledge base includes the following information:
- Procedure A is expensive, procedure B is cheap
- Procedure A is not dangerous, procedure B is dangerous
- R1 entered by expert Mr. Meyer, R2 entered by novice Mr. Schulz

Meta-rules:
R3: Prefer rules that recommend inexpensive procedures instead of expensive procedures.
R4: Prefer rules that recommend less dangerous procedures instead of dangerous procedures.
R5: Prefer rules that were entered by experts instead of rules that were entered by novices.

Conflict Resolution with Meta-Rules
(1) R4, R5: R1 before R2
(2) R3: R2 before R1
The decision depends on:
- for common interpreter for rules and meta-rules: evaluation
- for separated interpreters for meta-rules:
  - for random selection: (1) more likely
  - for ‘majority decision’: (1)

Reflexive use of meta rules as meta-meta-rules
- Assume that the knowledge base includes the following information:
  - R4 was entered by expert Mr. Meyer,
  - R3 was entered by novice Mr. Schulz
  - R5: R4 before R3: R1 before R2
- Assume that the knowledge base includes the following information:
  - Using R3 is dangerous,
  - using R5 is less dangerous,
  - using R4 is not dangerous,
  - R4: R4 before R3, R5: R1 before R2

In the case of a conflict between meta-rules, a further conflict resolution strategy could be needed.
An infinite regress will be prevented on a certain meta-level by selecting a production out of the conflict set randomly.
In practice, more than 2 meta-levels haven’t been necessary so far.

The Rule Pyramid

The rule pyramid shows typical relative sizes of object-rule levels and meta-rule levels in production systems.

Because of the lower number of rules on the higher meta-levels, the probability of a conflict set with more than one element becomes smaller.

The application of intelligent conflict resolution (more resources needed) in the sense of a best-first search strategy makes only sense on the lower levels of the rule pyramid.
Conflict Resolution Strategies

In the working memory, the data elements are ordered from left to right with respect to their increasing dwell time.

(P S) and (Q T) stored in the latest cycle, (P T) and (R V) stored two cycles ago, (Q S) and (P V) stored 3 cycles ago, (W V) and (W T) were stored 101 cycles ago in the working memory.

Productions are stored with respect to the order of their input. Variables are prefixed with ‘=’.

(1) Conflict resolution based on rule priorities

- PO1: Total order of rules, e.g. order of entry.
  with that: P1 is selected.
- PO2: Partial order of rules, P1 dominates P2, P3 dominates P4.
  with that: P1 and P3 are selected.
PO2 is relevant for structured rule-sets.

Conflict Resolution Strategies

(2) Conflict resolution based on specialization relations between instantiations and productions

- SC1: S is a special case of production A, if
  * S has at least as many premises as A and
  * for each premise in A including constants, exists a corresponding premise in S which includes these constants as a subset and
  * S and A are not identical
Thus, P4 is a special case of P1. If P4 is instantiated, the instantiations of P1 are excluded. There are no specialization relations in the sense of SC1 between P2, P3 and the other productions.

- SC2: S is a special case of instantiation A, if S includes the data elements of A as a proper subset.
  With that, I3 is a special case of I1

- SC3: Besides SC2, the production of S must include more premises than the production of A.
  With that, I4

Conflict Resolution Strategies

... based on dwell time in working memory

- R1: Order on the basis of the element of an instantiation with the shortest dwell time in the working memory. Time is defined as the number of executed actions since the storage of the element.

- R2: Same as R1, whereby time is defined as the number of interpreter cycles since the last storage.

- R3: Check the oldest elements of all instantiations. Choose the instantiation with the shortest dwell time in the sense of R1 and its element.

- R4: Time is defined as in R2. Check the oldest elements of all instantiations as in R3. The instantiations with an element \( \leq 100 \) dwell time cycles are preferred compared to instantiations with a dwell time \( \geq 101 \) cycles.

- R5: Time is defined as the number of executed actions. Compare the elements of all instantiations. First, compare the elements with the shortest dwell time. If the elements have the same dwell time, compare elements with second shortest dwell time. Once an instantiation is found that includes an element with a shorter dwell time than the comparison element, this instantiation is selected. Instantiations that have not enough elements for the successively comparison can not be selected. In R5, only instantiations with identical elements are equal.
### The Basic Syntax of OPS5

1) The structure of working memory elements

\[
\text{<element>} ::= \text{<vector-element>} | \text{<av-element>}
\]

\[
\text{<vector-element>} ::= (\{\text{<value>}\}^+) \\
\text{<av-element>} ::= (\text{<object>}\{^<\text{attribute}><\text{value}>\}^+)
\]

2) The structure of production rules

\[
\text{<rule>} ::= (\text{P <rule-name>} \text{<antecedent>} \rightarrow \text{<consequent>}) \\
\text{<antecedent>} ::= (\text{<condition>})^+ \\
\text{<condition>} ::= (\text{<pattern>} | -\text{<pattern>}) \\
\text{<pattern>} ::= (\text{<object>}\{^<\text{attribute}><\text{value}>\}^+) \\
\text{<consequent>} ::= (\text{<action>})^+ \\
\text{<action>} ::= (\text{MAKE <object>}\{^<\text{attribute}><\text{value}>\}^+) | \\
(\text{MODIFY <pattern-number>}\{^<\text{attribute}><\text{value}>\}^+) | \\
(\text{REMOVE <pattern-number>}) | \\
(\text{WRITE {<value>}})
\]

### Conflict Resolution Strategies in the OPS5 Rule Interpreter

OPS5 includes two conflict resolution strategies: LEX and MEA. In each strategy, multiple filters are arranged in a cascade.

- **LEX** (default strategy)

  The term ‘LEX’ is based on the fact, that the lexicographical order order plays an important role in the second filter.

  LEX is primarily suitable for small rule-sets and for problems where the order in which the subtasks are solved is not important.

- **MEA**

  The term ‘MEA’ should hint at the problem solving method ‘Means Ends Analysis’.

  The prioritization of the first single constraint is used to control the system. Thereby, this constraint describes the current subtask.
Conflict Resolution Strategies in the OPS5 Rule Interpreter

✓ Uniqueness

All instantiations are filtered out which are already fired but still remained in the conflict set because their conditions were still fulfilled by the elements of the working memory.

The underlying heuristic for this filter: firing the same instantiation once again will bring nothing new.

✓ Novelty

This second filter underlies a dwell time strategy in which instantiations based on the newest data elements are preferred.

The time stamps of an instantiation are arranged in descending order and can be understood as a vector according to their size.

In this order, the biggest vector belongs to the instantiation that is able to pass this filter.

Conflict Resolution Strategies in the OPS5 Rule Interpreter

✓ Novelty

It is possible that a few instantiations pass the filter because several instantiations can have the same vector of time stamps.

If a vector is included in another vector, the longer vector is seen as the bigger one in this order. In doing so, besides a dwell time strategy, a special case strategy is also realized.

Example:

<table>
<thead>
<tr>
<th>Input</th>
<th>Output of the filter:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule_3 27 19 11 6</td>
<td>Rule_3 27 19 11 6</td>
</tr>
<tr>
<td>Rule_1 27 19 11 6</td>
<td>Rule_1 27 19 11 6</td>
</tr>
<tr>
<td>Rule_2 27 19 11</td>
<td>Rule_2 27 19 11</td>
</tr>
<tr>
<td>Rule_7 27 19 11</td>
<td>Rule_7 27 19 11</td>
</tr>
<tr>
<td>Rule_2 21 19 11</td>
<td>Rule_2 21 19 11</td>
</tr>
<tr>
<td>Rule_4 17 12 13 6 2 1</td>
<td>Rule_4 17 12 13 6 2 1</td>
</tr>
</tbody>
</table>