Probabilistic logic programming with multiplicative modules

Roberto Maieli

maieli@uniroma3.it

TLLA 2021 - Rome, June 28, 2021
the quest of modularity

[...] all the problems concerning correctness and modularity of programs appeal in a deep way to the syntactic tradition, to proof theory.

[...] Heyting semantics is very original: it does not interpret the logical operations by themselves, but by abstract constructions. Now we can see that these constructions are nothing but typed i.e. modular programs.


Outline (this talk in 6 lines):

1. a **multiplicative module** is a "piece" of "multiplicative net" ⊆ MLL PNs;
2. the special case of **multiplicative bipoles** generalize Andreoli’s MLL bipoles (LP);
3. a **multiplicative module** is characterized by a **behavior** (a partitions set);
4. a **probability distribution function** is associated to each multiplicative module;
5. we deal with **non-determinism of processes** but no need for **additives** & , ⊕;
6. **correctness** of process transition is **LINEAR** (in the size of the behavior).
**multiplicative module**

**Def:** a **multiplicative module** $\mu$ is a triple $\langle I = \{i_1, ..., i_{n\geq0}\}, O = \{o_1, ..., o_{m\geq1}\}, B_\mu \rangle$

- $I$ is a possibly empty set of input indexes,
- $O$ is a non empty set of output indexes with $I \cap O = \emptyset$
- $B_\mu$ is a set of partitions (the **behavior** of $\mu$) over the border $B = I \cup O$ s.t.:
  1. all partitions $P_1, ..., P_h, ..., P_l$ in $B_\mu$ have **same size** (number of classes/blocks)

     $P_1 = \{\alpha^1_1, ..., \alpha^1_z\}$

     \[ \vdots \]

     $P_h = \{\alpha^j_1, ..., \alpha^j_z\}$

     \[ \vdots \]

     $P_l = \{\alpha^l_1, ..., \alpha^l_z\}$

  2. \(\forall i_j, \forall o_k, \exists P_h \in B_\mu\) s.t. $i_j$ and $o_k$ **occur together in a class** $\alpha^h_t$ of $P_h$;

  3. the **orthogonal** $(B_\mu)^\perp$ of $B_\mu$ must be not empty.
ORTHOGONALITY

**Def.** Two modules $\mu, \beta$ are **orthogonal** iff their behaviors (partitions sets) $B_\mu, B_\beta$ are orthogonal, $B_\mu \perp B_\beta$, iff they are pointwise orthogonal:

$$\forall P \in B_\mu \text{ and } \forall Q \in B_\beta, P \perp Q$$

"Orthogonality" $P \perp Q$ is defined by a topological condition: the bipartite graph obtained by linking together classes/blocks of each partition sharing an element is acyclic and connected.

**Example.**

\{(1,2), (3)\} is **not** orthogonal to \{(1,2,3)\} see $G_1$

\{(1,2), (3)\} is both orthogonal to \{(1,3), (2)\} and \{(1), (2,3)\} see $G_2, G_3$

\begin{align*}
G_1 : & \quad [1,2] \quad [3] \\
& \quad \bullet \quad \bullet \quad \bullet
\end{align*}

\begin{align*}
G_2 : & \quad [1,2] \quad [3] \\
& \quad \bullet \quad \bullet \quad \bullet
\end{align*}

\begin{align*}
G_3 : & \quad [1,2] \quad [3] \\
& \quad \bullet \quad \bullet \quad \bullet
\end{align*}

\begin{align*}
G_1 : & \quad [1,2,3] \\
& \quad \bullet
\end{align*}

\begin{align*}
G_2 : & \quad [1,3] \quad [2] \\
& \quad \bullet \quad \bullet
\end{align*}

\begin{align*}
G_3 : & \quad [1] \quad [2,3] \\
& \quad \bullet \quad \bullet
\end{align*}
Def. A **multiplicative bipole** is a special case of multiplicative module

\[ \beta : \langle I = \{i_1, ..., i_{n \geq 0}\}, O = \{o_1, ..., o_{m \geq 1}\}, \mathcal{B}_\beta \rangle \]

- with the condition that: for each partition \( P_h \) in \( \mathcal{B}_\beta \), all the elements of the output set \( O \) must belong to a single class (the **head class**) \( \alpha_h^t \) of \( P_h \).
- \( O \) is called the **head** of "method" \( \beta \): it plays the role of the "trigger" of \( \beta \);
- \( I \) is called the **body** of "method" \( \beta \).

![Diagram](image)

\[ P_1 = \{ \alpha_1^1 = (...o_1, ..., o_m,...), ..., \alpha_2^1 \} \]

... 

\[ P_h = \{ \alpha_1^h, ..., \alpha_t^h = (...o_1, ..., o_m,...), ..., \alpha_z^h \} \]

... 

\[ P_I = \{ \alpha_1^l, ..., \alpha_z^l = (...o_1, ..., o_m,...) \} \]
orthogonality guarantees bipoles Expansion ∼ Resolution

**Example** given:

- module $\pi$ with behavior $\mathcal{B}_\pi$ over the border $I = \{0, 1, 2, 3, 4\} \cup O = \{0\}$;
- bipole $\beta$ with behavior $\mathcal{B}_\beta$ over the border $I = \{5, 6, 7\} \cup O = \{1, 4\}$,

$$
\mathcal{B}_\pi = \begin{cases}
    p_1 : (1) \ (0, 2, 3) \ (4) \\
    p_2 : (2) \ (0, 1, 3) \ (4) \\
    p_3 : (1) \ (2, 3) \ (0, 4) \\
    p_4 : (2) \ (1, 3) \ (0, 4)
\end{cases} \quad \quad \quad
\mathcal{B}_\beta = \begin{cases}
    q_1 : (6) \ (5, 7, 1, 4) \\
    q_2 : (5) \ (6, 7, 1, 4)
\end{cases}.
$$

- the head $H = O : \{1, 4\}$ of $\beta$ is included in the body $I : \{1, 2, 3, 4\}$ of $\pi$;
- the restricted behaviors $(\mathcal{B}_\pi)\downarrow^H$ and $(\mathcal{B}_\beta)\downarrow^H$ are orthogonal, $\{(1, 4)\} \perp \{(1), (4)\}$;
- then, we can **expand** $\pi$ by $\beta$ and build the **multiplicative bipolar module/net** $\pi \circ \beta$:

$$
\mathcal{B}_{\pi \circ \beta} = \begin{cases}
    q_1.p_1 : (6) \ (5, 7) \ (0, 2, 3) \\
    q_2.p_1 : (5) \ (6, 7) \ (0, 2, 3) \\
    q_1.p_2 : (2) \ (6) \ (5, 7, 0, 3) := q_1.p_4 \\
    q_2.p_2 : (2) \ (5) \ (6, 7, 0, 3) := q_2.p_4 \\
    q_1.p_3 : (6) \ (5, 7, 0) \ (2, 3) \\
    q_2.p_3 : (5) \ (6, 7, 0) \ (2, 3).
\end{cases}
$$

Correctness of expansion is LINEAR in the size of the behavior of $\pi$. 
Example 1. $\beta$ is MLL definable/decomposable:

- border $I = \{a, b, c, d\}$, $O = \{h_1, h_2\}$
- behavior $B_\beta = \{((a, c, h_1, h_2), (b), (d)), ((a, d, h_1, h_2), (b), (c)),
  ((b, c, h_1, h_2), (a), (d)), ((b, d, h_1, h_2), (a), (c))\}.$

\[\exists\text{ a MLL proof structure } B \text{ (a bipole indeed) s.t. the behavior of } \beta \text{ corresponds to the set of partitions of the border of } B \text{ induced by all Danos-Regnier switchings: in a switching } S \text{ for } B, \text{ two points of the border stay in the same class iff they stay in a same connected component of } S.\]

\[\beta \text{ is a MLL bipole!}\]

Example 2. $\gamma$ is MLL definable: it is an MLL monopole:

- border $I = \emptyset$, $O = \{h_1, \ldots, h_n\}$
- behavior $B_\beta = \{((h_1, \ldots, h_n))\}$ (a singleton)

\[\gamma \text{ is a MLL monopole!}\]
multiplicative bipolar net that are MLL definable

\[ B_\pi = \begin{cases} p_1 : (1, 0, 2, 3) \\ p_2 : (2, 0, 1, 3) \\ p_3 : (1, 2, 3) \\ p_4 : (2, 1, 3) \end{cases}, \quad B_\beta = \begin{cases} q_1 : (6, 5, 7, 1, 4) \\ q_2 : (5, 6, 7, 1, 4) \end{cases}, \quad B_{\pi \circ \beta} = \begin{cases} q_1.p_1 : (6, 5, 7) \\ q_1.p_2 : (5, 6, 7) \\ q_2.p_1 : (2, 6) \\ q_2.p_2 : (5, 6, 7, 0) \end{cases}. \]

There are three equivalent ways to perform the bipolar proof construction in the MLL case:
- by sets (orthogonal behaviors i.e., partitions sets)
- by graphs (proof net expansion)
- by trees (sequent calculus expansion)

**Theorem** Given a set of MLL methods/bipoles \( U = \{ \beta_1, ..., \beta_n \} \) (LP) and a goal \( G \) (a multi-set of atoms \( \{ a_1, ..., a_m \} \)) then \( U \vdash_{\text{MLL foc}} G \) iff \( \exists \mu : \langle I : \{ i_1, ..., i_{n \geq 0} \}, O : \{ o_1, ..., o_{m \geq 1} \}, B_\mu \rangle \) s.t.:

1. \( O = \{ a_1, ..., a_m \} \) and
2. \( B_\mu \) is built by expanding \( \beta_1, ..., \beta_n \).
"primitive" multiplicative bipoles that are NOT MLL definable

\[ \{ \text{MLL bipoles} \} \subsetneq \{ \text{multiplicative bipoles} \} \]

\( \gamma \) is NOT MLL definable.

\[ B_\gamma = \{ \{ (i_1, o_1, o_2), (i_2, o_3, o_4) \}, \{ (i_1, o_2, o_3), (i_2, o_4, o_1) \} \} \]

\( \beta \) is NOT MLL definable.

\[ B_\beta = \{ \{ (i_1, i_3, o_5, o_6), (i_2), (i_4) \}, \{ (i_2, i_4, o_5, o_6), (i_1), (i_3) \} \} \]

\( B_\gamma \perp B_\beta \):

\( B_\gamma \) restricted to \( O_\gamma = \{ o_1, o_2, o_3, o_4 \} \) and \( B_\beta \), restricted to \( I_\beta = \{ i_1, i_2, i_3, i_4 \} \) are orthogonal modulo the unification \( I_\beta \leftrightarrow O_\gamma: i_1 = o_1, i_2 = o_2, i_3 = o_3, i_4 = o_4 \).
the unfolding of "primitive" bipoles

\( \gamma \) can be interpreted as the **union** of the behaviors of two pairs of "concurrent" bipoles:

\[
B_\gamma = B_{\gamma_1} \cup B_{\gamma_2} \text{ with } \gamma_1 = \alpha_1 \otimes \alpha_2 \text{ and } \gamma_2 = \alpha_1' \otimes \alpha_2' \]

\[
B_\gamma = \{ \{(i_1, o_1, o_2), (i_2, o_3, o_4)\}, \\
\{(i_1, o_2, o_3), (i_2, o_4, o_1)\} \}
\]

\[
B_\gamma = \{ B_{\gamma_1} = \{\{(i_1, o_1, o_2), (i_2, o_3, o_4)\}\} \cup B_{\gamma_2} = \{\{(i_1, o_2, o_3), (i_2, o_4, o_1)\}\}\}
\]

We say that \( \gamma \) can be **unfolded** in to \( \{\gamma_1, \gamma_2\} \) called the **unfolding trace/family of** \( \gamma \).
the unfolding of "primitive" bipoles

Dually, $\beta$ can be interpreted as the intersection of a pair of MLL bipoles, $\beta_1$ and $\beta_2$, with the same "skeleton" and whose input borders only differ by the cyclic permutation of the input sequence $(i_1, i_2, i_3, i_4)$, that is:

$$B_\beta = B_{\beta_1} \cap B_{\beta_2}$$

We say that $\beta$ can be unfolded in to $\{\beta_1, \beta_2\}$ called the unfolding trace/family of $\beta$.

Note this unfoldable module expresses a kind of non-deterministic super-position ($\cap$): only one of them or both simultaneously may partecipate to the net expansion.

Roberto Maieli (maieli@uniroma3.it)
Prob. Log. Programm. with multiplicative modules
TLLA 2021 - Rome, June 28, 2021 11 / 18
in **standard logic programming**, conditional probability values are assigned to method (MLL bipoles) and a-priori probability values are assigned to fact (MLL monopole):

\[ H : \neg B_1, \ldots, B_n \quad p(H \mid \bigcap_i B_i) \quad \text{conditional probability} \]

\[ H : \neg . \quad p(H) \quad \text{a-priori probability} \]

with **multiplicative unfoldable modules**, we assign a **probability distribution function** to a unfoldable bipolar module: this function describes all possible values and likelihoods that a random variable can take within a given range.
Let $\beta$ be a **multiplicative unfoldable bipole**

with behavior $B_\beta$ over the border $I = \{i_1, \ldots, i_n\} \uplus O = \{o_1, \ldots, o_m\}$;

Let $\beta_1, \ldots, \beta_k$ be the **unfolding trace** (the unfolding family of MLL bipoles) of $\beta$.

We call a **probability distribution** for $\beta$ a (finite) set of real number values,

$$P(O|I)_\beta = \{p(\beta_i) \mid 0 < p(\beta_i) \in \mathbb{R} \leq 1 \text{ and } \beta_i \text{ is in the trace of } \beta\}$$

with the condition that in case that $B_\beta = \bigcup_i B_{\beta_i}$ then, $\sum_{i=1}^{k} p(\beta_i) = 1$.

In particular, if $\beta$ is a **MLL bipole** then, $P(O|I) = \{p(\beta)\}$ (a singleton):

- if $\beta$ is a method with $I \neq \emptyset$ then $p(\beta)$ is the conditional probability $p(O|I)$,
- if $\beta$ is a fact (i.e., $I = \emptyset$) then, $p(\beta)$ is an a-priori probability $p(O)$. 
probability distribution of unfoldable bipoles

In case \( B_\beta = \bigcup_{i=1}^{k} B_{\beta_i} \) then \( P_\beta(O|I) = \{p(\gamma_1), p(\gamma_2)\} \) s.t. \( p(\gamma_1) + p(\gamma_2) = 1 \).

\( p(O|I) \) expresses the variation of probability over an aleatory variable \( O = \{o_1, o_2, o_3, o_4\} \):

\[
P(O|I) = \{p(\gamma_1), p(\gamma_2)\} \quad \text{s.t.} \quad p(\gamma_1) + p(\gamma_2) = 1.
\]

Example.

Assume for simplification reasons that \( I = \emptyset \) then, \( p(O) \) expresses the variation of probability over the aleatory "variable" \( O = \{o_1, o_2, o_3, o_4\} \):

- \( p(\gamma_1) \) denotes the a-priori probability \( p(E_1) \) of the event \( E_1 \):
  "resource \( o_1 \) occurs together with \( o_2 \) while resource \( o_3 \) occurs together with resource \( o_4 \);"

- \( p(\gamma_2) \) denotes the a-priori probability \( p(E_2) \) of the event \( E_2 \):
  "resource \( o_2 \) occurs together with resource \( o_3 \) while resources \( o_4 \) occurs together with \( o_1 \)."
probability distribution of unfoldable bipoles

otherwise, in case \( B_\beta = \bigcap_{i=1}^{k} B_{\beta_i} \) then \( P_\beta(O|I) = \{p(\gamma_1), p(\gamma_2)\} \) where every \( p(\beta_i) \) expresses a condition probability \( p(O|I) \)

\[
B_\beta = \{ (i_1, i_2, o_5, o_6), (i_2, i_4) \} = B_{\beta_1} : \{(i_1, i_3, o_5, o_6), (i_2, i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1, i_3)\}, \{(i_1, i_4, o_5, o_6), (i_2, i_3)\}, \{(i_2, i_3, i_4, o_5, o_6), (i_1)\}, \{(i_3, i_2, i_4, o_5, o_6), (i_1)\}, \{(i_4, i_3, i_2, i_4, o_5, o_6), (i_1)\}, \{(i_4, i_3, i_2, o_5, o_6), (i_1)\}, \{(i_4, i_3, o_5, o_6), (i_1)\}, \{(i_4, o_5, o_6), (i_1)\}, \{(o_5, o_6), (i_1)\} \\
B_{\beta_2} : \{(i_1, i_3, o_5, o_6), (i_2, i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1, i_3)\}, \{(i_1, i_4, o_5, o_6), (i_2, i_3)\}, \{(i_2, i_3, i_4, o_5, o_6), (i_1)\}, \{(i_3, i_2, i_4, o_5, o_6), (i_1)\}, \{(i_3, i_4, o_5, o_6), (i_1)\}, \{(i_3, o_5, o_6), (i_1)\}, \{(o_5, o_6), (i_1)\}
\]

**Example.**

- \( p(\beta_1) \) expresses the **conditional probability** \( p(E|E_1) \) that:

  "we observe the event \( E \), in which resource \( o_5 \) stays together with resource \( o_6 \), if occurs the event \( E_1 \) that resources \( i_1 \) stays together with \( i_2 \) while \( i_3 \) stays together \( i_4 \)";

- \( p(\beta_2) \) expresses the **conditional probability** \( p(E|E_2) \) that:

  "we observe the event \( E \), in which resource \( o_5 \) stays together with resource \( o_6 \), if occurs the event \( E_2 \) that resources \( i_2 \) stays together with \( i_3 \) while \( i_4 \) stays together with \( i_1 \)."
There are two directions of the information flow in our net construction model:

1. **net expansion** $\uparrow$: the first direction consists in the bottom-up construction of the net, by module expansions;

2. **info propagation** $\downarrow$: the second direction intervenes when the net construction is successfully completed; in that case, we can invert the direction of the information and propagate the probability information from the top (that is, the a-priori probabilities associated to the axiom-bipoles/facts) to the bottom.
Net unfolding and Naive Bayesian Classification

An example inspired to Naive Bayesian Classifier (used e.g. in Machine Learning):

Let us classify a new instance of the event $E = (o_5, o_6)$ according either to event $E_1$ or to $E_2$;

Assume the sub-net $T_2$ is the trained Naive Bayesian model.

Unfolding the trained model $T_2$ allows us to calculate the \textbf{a-posteriori probabilities} that:

"if event $E$ occurs then, we could expect event $E_1$ (net $T'_2$) rather than event $E_2$ (net $T''_2$)"

\textbf{Bayes’ Theorem:} \hspace{1cm} 

\[ p(E_1|E) = \frac{p(E|E_1)p(E_1)}{p(E)} : T'_2, \hspace{0.5cm} p(E_2|E) = \frac{p(E|E_2)p(E_2)}{p(E)} : T''_2 \]

where:

$- p(E) = \sum_{i=1}^{2} p(E|E_i).p(E_i)$ is the \textbf{absolute probability} that event $E$ will occur;

$- p(E|E_1).p(E_1) = p(\beta_1).p(\gamma_1)$ and $p(E|E_2).p(E_2) = p(\beta_2).p(\gamma_2)$. 

Roberto Maieli (maieli@uniroma3.it)
**conclusion & further woks**

**CONCLUSIONS:**
- Probabilistic choice, where each branch of a choice is weighted according to a probability distribution, is an established approach for modelling processes;
- this task is often carried out by using additives $\&$, $\oplus$;
- why should I use unfolding modules instead of "standard" additives?

1. correctness of additive (MALL) proof structure is NON-LINEAR while correctness of generalized multiplicatives is LINEAR (in the behavior size);
2. additives have global effects while here we propose a (non-deterministic) "local choice behavior" inherent in multiplicatives.

**FURTHER WORKS:**
- connection with Girard's Transcendental Syntax (see yesterday Boris Eng's talk)
- a Naive Bayesian Classifier for Machine Learning based on modules/rules.