Probabilistic logic programming with multiplicative modules

Roberto Maieli

maieli@uniroma3.it

TLLA 2021 - Rome, June 28, 2021

the quest of modularity

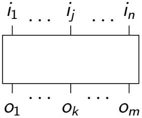
> [...] all the problems concerning correctness and modularity of programs appeal in a deep way to the syntactic tradition, to proof theory.
>
> [...] Heyting semantics is very original: it does not interpret the logical operations by themselves, but by abstract constructions. Now we can see that these constructions are nothing but typed i.e. modular programs.
>
> *J.-Y. Girard, Proofs and Types, 1989.*

Outline (this talk in 6 lines):

1. a **multiplicative module** is a "piece" of "multiplicative net" $\supsetneq$ 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** $\&, \oplus$;
6. **correctness** of process transition is **LINEAR** (in the size of the behavior).

multiplicative module

DEF: a **multiplicative module** μ is a triple $\langle I = \{i_1, ..., i_{n\geq 0}\}, O = \{o_1, ..., o_{m\geq 1}\}, \mathcal{B}_\mu \rangle$

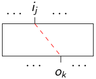


– I is a possibly empty set of input indexes,
– O is a non empty set of output indexes with $I \cap O = \emptyset$
– $\mathcal{B}_\mu$ is a set of partitions (the **behavior** of μ) over the border $B = I \cup O$ s.t.:

1. all partitions $P_1, ..., P_h, ..., P_l$ in $\mathcal{B}_\mu$ 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. $\forall i_j, \forall o_k, \exists P_h \in \mathcal{B}_\mu$ s.t. i_j and o_k **occur together in a class** α^h_t of P_h;



3. the **orthogonal** $(\mathcal{B}_\mu)^\perp$ of $\mathcal{B}_\mu$ must be not empty.

orthogonality

DEF: two modules μ, β are **orthogonal** iff their behaviors (partitions sets) $\mathcal{B}_\mu, \mathcal{B}_\beta$ are orthogonal, $\mathcal{B}_\mu \perp \mathcal{B}_\beta$, iff they are pointwise orthogonal:

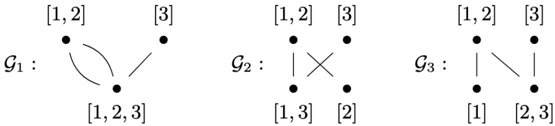
$$\forall P \in \mathcal{B}_\mu \text{ and } \forall \mathcal{Q} \in \mathcal{B}_\beta, \mathcal{P} \perp \mathcal{Q}$$

"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 $\mathcal{G}_1$

$\{(1,2),(3)\}$ is both orthogonal to $\{(1,3),(2)\}$ and $\{(1),(2,3)\}$ see $\mathcal{G}_2, \mathcal{G}_3$

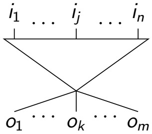


multiplicative bipole

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**) α_t^h of P_h.
- O is called the **head** of "method" β: it plays the role of the "trigger" of β; I is called the **body** of "method" β.



$$P_1 = \{\alpha_1^1 = (...o_1, ..., o_m, ...), ..., \alpha_z^1\}$$

$$\vdots$$

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

$$\vdots$$

$$P_l = \{\alpha_1^l, ..., \alpha_z^l = (...o_1, ..., o_m, ...)\}$$

orthogonality guarantees bipoles Expansion ∼ Resolution

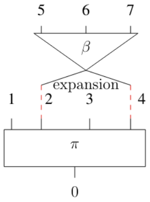
EXAMPLE given:

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

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

- the head $H = O : \{1,4\}$ of β is included in the body $I : \{1,2,3,4\}$ of π
- 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** π by β and build the **multiplicative bipolar module/net** $\pi \circ \beta$:

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



Correctness of expansion is LINEAR in the size of the behavior of π.

multiplicative bipoles that are MLL definable

EXAMPLE 1. β is **MLL definable/decomposable**:

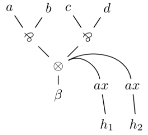
– border $I = \{a, b, c, d\}$, $O = \{h_1, h_2\}$
– behavior $\mathcal{B}_\beta = \{\{(a, c, h_1, h_2), (b), (d)\}, \{(a, d, h_1, h_2), (b), (c)\},$
$\{(b, c, h_1, h_2), (a), (d)\}, \{(b, d, h_1, h_2), (a), (c)\}\}.$

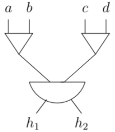
$\exists$ a MLL proof structure B (a bipole indeed) s.t. the behavior of β corresponds to the set of partitions of the border of B induced by all Danos-Regnier switchings: in a switching S for B, two points of the border stay in the same class iff they stay in a same connected component of S.

method β
$[h_1, h_2] : -[a, b], [c, d]$

$(h_1^\perp \otimes h_2^\perp) \otimes (a \parr b) \otimes (c \parr d)$

$((h_1 \parr h_2) \multimap (a \parr b) \otimes (c \parr d))^\perp$





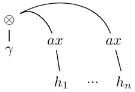
β is a MLL bipole!

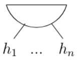
EXAMPLE 2. γ is MLL definable: it is an **MLL monopole**:

– border $I = \emptyset$, $O = \{h_1, ..., h_n\}$
– behavior $\mathcal{B}_\beta = \{\{(h_1, ..., h_n)\}\}$ (a singleton)

fact γ
$[h_1, ..., h_n] : -.$

$(h_1 \otimes ... \otimes h_n)$



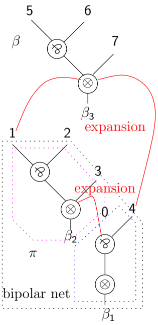


γ is a MLL monopole!

multiplicative bipolar net that are MLL definable

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

$$\cfrac{\cfrac{\cfrac{\vdash 5,6 \qquad \vdash 7,2 \qquad \vdash 4,4^\perp \qquad \vdash 1,1^\perp}{\vdash \beta_3, 1, 2, 4}\,\beta_3 \qquad \vdash 3 \qquad \vdash 0, 0^\perp}{\vdash \beta_2, \beta_3, 0, 4}\,\beta_2}{\vdash \beta_1, \beta_2, \beta_3}\,\beta_1$$



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)

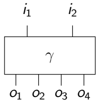
THEOREM Given a set of MLL methods/bipoles $\mathcal{U} = \{\beta_1, .., \beta_n\}$ (LP) and a goal G (a multi-set of atoms $\{a_1, ..., a_m\}$) then $\mathcal{U} \vdash_{MLLfoc} G$ iff $\exists \mu : \langle I : \{i_1, ..., i_{n\geq 0}\}, O : \{o_1, ..., o_{m\geq 1}\}, \mathcal{B}_\mu \rangle$ s.t.:

1. $O = \{a_1, ..., a_m\}$ and
2. $\mathcal{B}_\mu$ is built by expanding $\beta_1, ..., \beta_n$.

"primitive" multiplicative bipoles that are NOT MLL definable

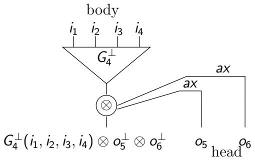
{MLL bipoles}$\subsetneq${multiplicative bipoles}

γ is **NOT** MLL definable.



$$\mathcal{B}_\gamma = \{ \quad \{(i_1, o_1, o_2), (i_2, o_3, o_4)\}, \{(i_1, o_2, o_3), (i_2, o_4, o_1)\}, \quad \}$$

β is **NOT** MLL definable.



$$\mathcal{B}_\beta = \{ \quad \{(i_1, i_3, o_5, o_6), (i_2), (i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1), (i_3)\} \quad \}$$

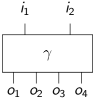
$\mathcal{B}_\gamma \perp \mathcal{B}_\beta$:

$\mathcal{B}_\gamma$ restricted to $O_\gamma = \{o_1, o_2, o_3, o_4\}$ and $\mathcal{B}_\beta$, restricted to $I_\beta = \{i_1, i_2, i_3, i_4\}$ are orthogonal modulo the unification $I_\beta \leftrightarrow O_\gamma$: $i_1 = o_1, i_2 = o_2, i_3 = o_3, i_4 = o_4$.

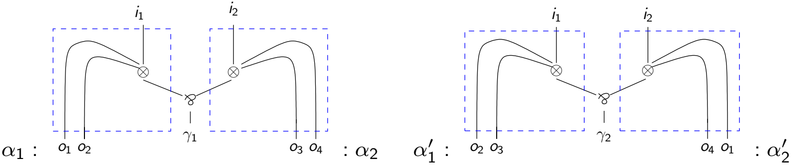
the unfolding of "primitive" bipoles

γ can be interpreted as the **union** of the behaviors of two pairs of "concurrent" bipoles:

$$\mathcal{B}_\gamma = \mathcal{B}_{\gamma_1} \cup \mathcal{B}_{\gamma_2} \text{ with } \gamma_1 = \alpha_1 \parr \alpha_2 \text{ and } \gamma_2 = \alpha'_1 \parr \alpha'_2$$



$$\mathcal{B}_\gamma = \{ \quad \{(i_1, o_1, o_2), (i_2, o_3, o_4)\}, \\ \{(i_1, o_2, o_3), (i_2, o_4, o_1)\}, \quad \}$$



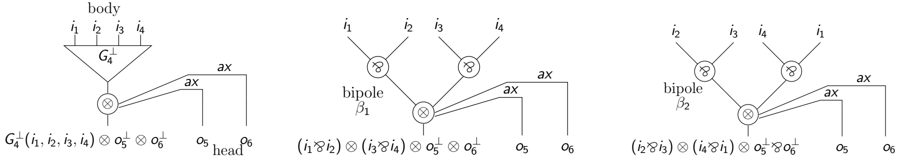
$$\mathcal{B}_\gamma = \{\mathcal{B}_{\gamma_1} = \{\{(i_1, o_1, o_2), (i_2, o_3, o_4)\}\} \quad \cup \quad \mathcal{B}_{\gamma_2} = \{\{(i_1, o_2, o_3), (i_2, o_4, o_1)\}\}\}$$

We say that γ can be **unfolded** in to $\{\gamma_1, \gamma_2\}$ called the **unfolding trace/family of** γ.

the unfolding of "primitive" bipoles

Dually, β can be interpreted as the **intersection** of a pair of MLL bipoles, β_1 and β_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:

$$\mathcal{B}_\beta = \mathcal{B}_{\beta_1} \cap \mathcal{B}_{\beta_2}$$



$$\mathcal{B}_\beta = \{ \{(i_1, i_3, o_5, o_6), (i_2), (i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1), (i_3)\} \} =$$

$$\mathcal{B}_{\beta_1} : \{\{(i_1, i_3, o_5, o_6), (i_2), (i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1), (i_3)\}, \{(i_1, i_4, o_5, o_6), (i_2), (i_3)\}, \{(i_2, i_3, i_4, o_5), (i_1), (i_4)\}\}$$
$$\cap$$
$$\mathcal{B}_{\beta_2} : \{\{(i_1, i_3, o_5, o_6), (i_2), (i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1), (i_3)\}, \{(i_1, i_2, o_5, o_6), (i_3), (i_4)\}, \{(i_3, i_4, o_5, o_6), (i_1), (i_2)\}\}$$

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

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.

logic programming with probabilities

- in **standard logic programming**, conditional probability values are assigned to method (MLL bipoles) and a-priori probability values are assigned to fact (MLL monopole):

$$H :- B_1, ..., B_n \qquad p(H \mid \bigcap_i B_i) \quad \textbf{conditional probability}$$

$$H :- . \qquad p(H) \quad \textbf{a-priori probability}$$

- 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.

probability distribution function of unfoldable bipoles

- Let β be a **multiplicative unfoldable bipole**
  with behavior $\mathcal{B}_\beta$ over the border $I = \{i_1, ..., i_n\} \uplus O = \{o_1, ..., o_m\}$;
  Let $\beta_1, ..., \beta_k$ be the **unfolding trace** (the unfolding family of MLL bipoles) of β.

- We call a **probability distribution** for β a (finite) set of real number values,

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

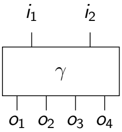
  with the condition that in case that $\mathcal{B}_\beta = \bigcup_i \mathcal{B}_{\beta_i}$ then, $\sum_{i=1}^k p(\beta_i) = 1$.

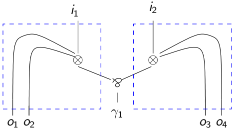
- In particular, if β **is a MLL bipole** then, $P(O|I) = \{p(\beta)\}$ (a singleton):
  - if β is a method with $I \neq \emptyset$ then $p(\beta)$ is the conditional probability $p(O|I)$,
  - if β is a fact (i.e., $I = \emptyset$) then, $p(\beta)$ is an a-priori probability $p(O)$.

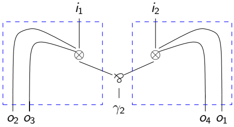
probability distribution of unfoldable bipoles

In case $\mathcal{B}_\beta = \bigcup_{i=1}^{k} \mathcal{B}_{\beta_i}$ = 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 $O = \{o_1, o_2, o_3, o_4\}$:



α_1 :  : α_2

α'_1 :  : α'_2

$$\mathcal{B}_\gamma = \{ \begin{array}{l} \{(i_1, o_1, o_2), (i_2, o_3, o_4)\}, \\ \{(i_1, o_2, o_3), (i_2, o_4, o_1)\}, \end{array} \} = \{\mathcal{B}_{\gamma_1} = \{\{(i_1, o_1, o_2), (i_2, o_3, o_4)\}\} \cup \mathcal{B}_{\gamma_2} = \{\{(i_1, o_2, o_3), (i_2, o_4, o_1)\}\}\}$$

EXAMPLE.

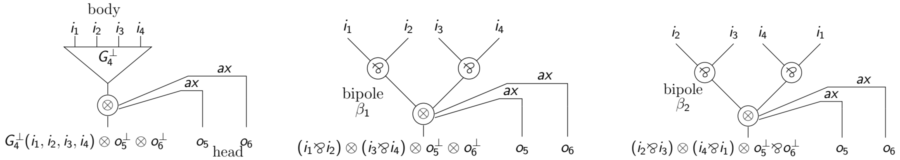
Assume for simplification reasons that $I = \emptyset$ then, $p(O)$ expresses the variation of probability over the aleatory "variable" $O = \{o_1, o_2, o_3, o_4\}$:

- $p(\gamma_1)$ 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 $\mathcal{B}_\beta = \bigcap_{i=1}^{k} \mathcal{B}_{\beta_i} =$ then $P_\beta(O|I) = \{p(\gamma_1), p(\gamma_2)\}$ where every $p(\beta_i)$ expresses a condition probability $p(O|I)$



$$\mathcal{B}_\beta = \{ \quad \{(i_1, i_3, o_5, o_6), (i_2), (i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1), (i_3)\} \quad \} \quad = \quad \begin{array}{l} \mathcal{B}_{\beta_1} : \{\{(i_1, i_3, o_5, o_6), (i_2), (i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1), (i_3)\}, \{(i_1, i_4, o_5, o_6), (i_2), (i_3)\}, \{(i_2, i_3, i_4, o_5), (i_1), (i_4)\}\} \\ \qquad \cap \\ \mathcal{B}_{\beta_2} : \{\{(i_1, i_3, o_5, o_6), (i_2), (i_4)\}, \{(i_2, i_4, o_5, o_6), (i_1), (i_3)\}, \{(i_1, i_2, o_5, o_6), (i_3), (i_4)\}, \{(i_3, i_4, o_5, o_6), (i_1), (i_2)\}\} \end{array}$$

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".

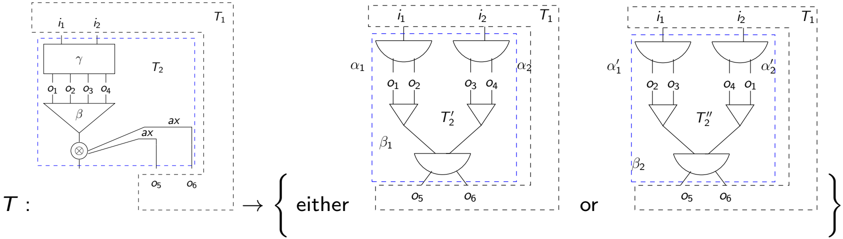
net expansion vs info propagation

There are two directions of the information flow in our net construction model:

1. **net expansion ↑:** the first direction consists in the bottom-up construction of the net, by module expansions;
2. **info propagation ↓:** 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 **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'')"

Bayes' Theorem: $p(E_1|E) = \dfrac{p(E|E_1)p(E_1)}{p(E)} : T_2'$, $\quad p(E_2|E) = \dfrac{p(E|E_2)p(E_2)}{p(E)} : T_2''$

where:
– $p(E) = \sum_{i=1}^{2} p(E|E_i).p(E_i)$ is the **absolute probability** that event E will occur;
– $p(E|E_1).p(E_1) = p(\beta_1).p(\gamma_1)$ and $p(E|E_2).p(E_2) = p(\beta_2).p(\gamma_2)$.

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.