

Weighted Visibly Pushdown Automata and Automated Music Transcription

Florent Jacquemard

April 29, 2021

1 SW Automata and Transducers

We follow the approach of [8] for the computation of distances between words with weighted transducers, and propose models of weighted automata and weighted transducers over infinite alphabets.

These models generalize weighted automata and transducers over finite alphabets, see e.g. [8], by labeling each transition with a weight functions that takes the input and output symbols as parameters, instead of a simple weight value. These functions are similar to the guards of symbolic automata [10], but they can return values in an arbitrary semiring, where the latter guards are restricted to the Boolean semiring.

1.1 Semirings

We shall consider semiring domains for weight values. A *semiring* $\langle \mathbb{S}, \oplus, \mathbb{0}, \otimes, \mathbb{1} \rangle$ is a structure with a domain \mathbb{S} , equipped with two associative binary operators \oplus and \otimes with respective neutral elements $\mathbb{0}$ and $\mathbb{1}$ and such that: \oplus is commutative, \otimes distributes over \oplus : $\forall x, y, z \in \mathbb{S}, x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$, and $\mathbb{0}$ is absorbing for \otimes : $\forall x \in \mathbb{S}, \mathbb{0} \otimes x = x \otimes \mathbb{0} = \mathbb{0}$.

A semiring \mathbb{S} is *commutative* if \otimes is commutative. It is *idempotent* if for each $x \in \text{dom}(\mathbb{S})$, $x \oplus x = x$. Following the terminology of [7], when $\forall x \in \text{dom}(\mathbb{S}), \mathbb{1} \oplus x = \mathbb{1}$, the semiring \mathbb{S} is called *bounded*. Note that every bounded semiring is idempotent: by boundedness, $\mathbb{1} \oplus \mathbb{1} = \mathbb{1}$, and idempotency follows by multiplying both sides by x and distributing.

A semiring \mathbb{S} is *monotonic wrt* a partial ordering \leq iff for all $x, y, z \in \mathbb{S}$, $x \leq y$ implies $x \oplus z \leq y \oplus z$, $x \otimes z \leq y \otimes z$ and $z \otimes x \leq z \otimes y$, and it is *superior wrt* \leq iff for all $x, y \in \mathbb{S}$, $x \leq x \otimes y$ and $y \leq x \otimes y$ [5]. The latter property corresponds to the *non-negative weights* condition in shortest-path algorithms [3]. Intuitively, it means that combining elements always increase their weight. Note that when \mathbb{S} is superior wrt \leq , then $\mathbb{1} \leq \mathbb{0}$ and moreover, for all $x \in \mathbb{S}$, $\mathbb{1} \leq x \leq \mathbb{0}$.

Every idempotent semiring \mathbb{S} induces a partial ordering $\leq_{\mathbb{S}}$ called the *natural ordering* of \mathbb{S} and defined by: for all x and y , $x \leq_{\mathbb{S}} y$ iff $x \oplus y = x$. This ordering is sometimes defined in the opposite direction [4]; The above definition follows [7], and coincides than the usual ordering on the Tropical semiring (*min-plus*). It holds that \mathbb{S} is monotonic wrt $\leq_{\mathbb{S}}$. An idempotent semiring \mathbb{S} is called *total* if it $\leq_{\mathbb{S}}$ is total *i.e.* when for all $x, y \in \mathbb{S}$, either $x \oplus y = x$ or $x \oplus y = y$.

We shall consider below infinite sums with \oplus . A semiring \mathbb{S} is called *complete* if for every family $(x_i)_{i \in I}$ of elements of $\text{dom}(\mathbb{S})$ over an index set $I \subseteq \mathbb{N}$, the infinite sum $\bigoplus_{i \in I} x_i$ is well-defined and in $\text{dom}(\mathbb{S})$, and the following properties hold:

- i. *infinite sums extend finite sums*: $\bigoplus_{i \in \emptyset} x_i = \mathbb{0}$, $\forall j \in \mathbb{N}$, $\bigoplus_{i \in \{j\}} x_i = x_j$,
 $\forall j, k \in \mathbb{N}, j \neq k$, $\bigoplus_{i \in \{j, k\}} x_i = x_j \oplus x_k$,
- ii. *associativity and commutativity*: for all $I \subseteq \mathbb{N}$ and all partition $(I_j)_{j \in J}$ of I , $\bigoplus_{j \in J} \bigoplus_{i \in I_j} x_i = \bigoplus_{i \in I} x_i$,
- iii. *distributivity of product over infinite sum*:
for all $I \subseteq \mathbb{N}$, $\bigoplus_{i \in I} (x \otimes y_i) = x \otimes \bigoplus_{i \in I} y_i$, and $\bigoplus_{i \in I} (x_i \otimes y) = (\bigoplus_{i \in I} x_i) \otimes y$.

1.2 Label Theory

Let Σ and Δ be respectively an input and output *alphabets*, which are countable (finite or infinite) sets of symbols, and let \mathbb{S} be a commutative semiring. A *label theory* is made of 4 recursively enumerable sets: $\Phi_{\epsilon} \subseteq \mathbb{S}$, Φ_{Σ} and Φ_{Δ} , containing unary functions in $\Sigma \rightarrow \mathbb{S}$, resp. $\Delta \rightarrow \mathbb{S}$, and $\Phi_{\Sigma, \Delta}$ containing binary functions in $\Sigma \times \Delta \rightarrow \mathbb{S}$. Moreover, we assume that these sets are closed under the operators \oplus and \otimes of \mathbb{S} . More precisely, for all $\phi, \phi' \in \Phi_{\Sigma}$ all $\psi, \psi' \in \Phi_{\Delta}$, and $\eta, \eta' \in \Phi_{\Sigma, \Delta}$, the function

$$\phi \otimes \phi' : x \mapsto \phi(x) \otimes \phi'(x) \text{ belongs to } \Phi_{\Sigma},$$

$$\psi \otimes \psi' : y \mapsto \psi(y) \otimes \psi'(y) \text{ belongs to } \Phi_{\Delta},$$

$$\phi \otimes \eta : x, y \mapsto \phi(x) \otimes \eta(x, y) \text{ belongs to } \Phi_{\Sigma, \Delta},$$

$$\eta \otimes \psi : x, y \mapsto \eta(x, y) \otimes \psi(y) \text{ belongs to } \Phi_{\Sigma, \Delta},$$

$$\eta \otimes \eta' : x, y \mapsto \eta(x, y) \otimes \eta'(x, y) \text{ belongs to } \Phi_{\Sigma, \Delta}.$$

The same also holds for the binary sum operator \oplus .

Finally, it is assumed that the codomain of every function of Φ_{Σ} and Φ_{Δ} maybe not necessary

is a subset of Φ_ϵ . and all partial applications of functions $\Phi_{\Sigma,\Delta}$, resp. $f_a : y \mapsto f(a, y)$ for $a \in \Sigma$ and $y \in \Delta$ and $f_b : x \mapsto f(x, b)$ for $b \in \Delta$ and $x \in \Sigma$, belong resp. to Φ_Σ and Φ_Δ .

Definition 1 A symbolic-weighted transducer T over the input and output alphabet Σ and Δ and the semiring \mathbb{S} is a tuple $T = \langle Q, \text{in}, \mathbf{w}, \text{out} \rangle$, where Q is a finite set of states, $\text{in} : Q \rightarrow \mathbb{S}$, respectively $\text{out} : Q \rightarrow \mathbb{S}$, are functions defining the weight for entering, respectively leaving, a state, and \mathbf{w} is a transition function from $Q \times Q$ into $\langle \Phi_0, \Phi_\Sigma, \Phi_\Delta, \Phi_{\Sigma,\Delta} \rangle$.

We extend the above transition function into a function from $Q \times (\Sigma \cup \{\epsilon\}) \times (\Delta \cup \{\epsilon\}) \times Q$ into \mathbb{S} , also called \mathbf{w} for simplicity, such that for all $q, q' \in Q$, $a \in \Sigma$, $b \in \Delta$, and with $\langle \phi_\epsilon, \phi_\Sigma, \phi_\Delta, \phi_{\Sigma,\Delta} \rangle = \mathbf{w}(q, q')$,

$$\begin{aligned} \mathbf{w}(q, \epsilon, \epsilon, q') &= \phi_\epsilon \\ \mathbf{w}(q, a, \epsilon, q') &= \phi_\Sigma(a) \\ \mathbf{w}(q, \epsilon, b, q') &= \phi_\Delta(b) \\ \mathbf{w}(q, a, b, q') &= \phi_{\Sigma,\Delta}(a, b) \end{aligned}$$

These functions ϕ act as guards for the transducer's transitions, preventing a transition when they return the absorbing $\mathbb{0}$ of \mathbb{S} .

The symbolic-weighted transducer T defines a mapping from the pairs of strings of $\Sigma^* \times \Delta^*$ into the weights of \mathbb{S} , based on the following intermediate function weight_T defined recursively for every $q, q' \in Q$, for every strings of $s \in \Sigma^*$, $t \in \Delta^*$:

$$\begin{aligned} \text{weight}_T(q, s, t, q') &= \\ &\oplus \bigoplus_{\substack{q'' \in Q \\ s=au, a \in \Sigma}} \mathbf{w}(q, \epsilon, \epsilon, q'') \otimes \text{weight}_T(q'', u, t, q') \\ &\oplus \bigoplus_{\substack{q'' \in Q \\ t=bv, b \in \Delta}} \mathbf{w}(q, \epsilon, b, q'') \otimes \text{weight}_T(q'', s, v, q') \\ &\oplus \bigoplus_{\substack{q'' \in Q \\ s=au, a \in \Sigma \\ t=bv, b \in \Delta}} \mathbf{w}(q, a, b, q'') \otimes \text{weight}_T(q'', u, v, q') \end{aligned}$$

Recall that by convention, an empty sum with \oplus is $\mathbb{0}$. The weight associated by T to $\langle s, t \rangle \in \Sigma^* \times \Delta^*$ is then defined as follows:

$$T(s, t) = \bigoplus_{q, q' \in Q} \text{in}(q) \otimes \text{weight}_T(q, s, t, q') \otimes \text{out}(q').$$

A symbolic weighted automata (SWA) $A = \langle Q, \text{in}, \text{weight}, \text{out} \rangle$ over Σ and \mathbb{S} is defined in a similar way by simply omitting the output symbols, i.e. \mathbf{w}

is a function of $Q \times Q$ into $\langle \Phi_0, \Phi_\Sigma \rangle$, or equivalently from $Q \times (\Sigma \cup \{\epsilon\}) \times Q$ into \mathbb{S} .

Proposition 2 *Given a SWT T over Σ , Δ and \mathbb{S} , and a word $s \in \Sigma^*$, one can construct a SWA $A_{s,T}$ such that for all $t \in \Delta^*$, $A_{s,T}(t) = T(s, t)$.*

The construction time and size of $A_{s,T}$ are $O(|s| \cdot \|T\|)$.

2 SW Visibly Pushdown Automata

The following model generalizes Symbolic VPA [2] from Boolean semirings to arbitrary semiring weight domains.

Let Σ be a countable alphabet that we assume partitioned into :

- a set Σ_i of *internal symbols* denoted a ,
- a set Σ_c of *call symbols* denoted $\langle a$,
- a set Σ_r of *return symbols* denoted \rangle_a .

Let \mathbb{S} be a commutative semiring and let $(\Phi_\epsilon, \Phi_c, \Phi_r, \Phi_{cr})$ be a label theory over \mathbb{S} where Φ_c , Φ_r and Φ_{cr} stand respectively for Φ_{Σ_c} , Φ_{Σ_r} and $\Phi_{\Sigma_c, \Sigma_r}$. Moreover, we extend this theory with a set Φ_i of unary functions in $\Sigma_i \rightarrow \mathbb{S}$, closed under \oplus and \otimes .

Definition 3 *A Symbolic Weighted Visibly Pushdown Automata (SWVPA) A over $\Sigma = \Sigma_i \uplus \Sigma_c \uplus \Sigma_r$ and \mathbb{S} is a tuple $T = \langle Q, P, \text{in}, w_i, w_c, w_r, w_e, \text{out} \rangle$, where Q is a finite set of states, P is a finite set of stack symbols, $\text{in} : Q \rightarrow \mathbb{S}$, respectively $\text{out} : Q \rightarrow \mathbb{S}$, are functions defining the weight for entering, respectively leaving, a state, and $w_i : Q \times Q \rightarrow \Phi_i$, $w_c : Q \times Q \times P \rightarrow \Phi_c$, $w_r : Q \times P \times Q \rightarrow \Phi_r$, $w_e : Q \times Q \rightarrow \Phi_{cr}$, are transition functions.*

Similarly as in Section 1, we extend the above transition functions as follows for all $q, q' \in Q$, $p \in P$, $a \in \Sigma_i$, $\langle c \in \Sigma_c$, $r \rangle \in \Sigma_r$, overloading the names for simplicity:

$$\begin{array}{lll}
w_i : Q \times \Sigma_i \times Q \rightarrow \mathbb{S} & w_i(q, a, q') = \phi_i(a) & \text{where } \phi_i = w_i(q, q'), \\
w_c : Q \times \Sigma_c \times Q \times P \rightarrow \mathbb{S} & w_c(q, \langle c, q', p) = \phi_c(\langle c) & \text{where } \phi_c = w_c(q, q', p), \\
w_r : Q \times \Sigma_r \times P \times Q \rightarrow \mathbb{S} & w_r(q, \langle c, p, r \rangle, q') = \phi_r(\langle c, r \rangle) & \text{where } \phi_r = w_r(q, p, q'), \\
w_e : Q \times \Sigma_r \times Q \rightarrow \mathbb{S} & w_e(q, r, q') = \phi_e(r) & \text{where } \phi_e = w_e(q, q').
\end{array}$$

The intuition is the following for the above transitions.

w_i : read the input internal symbol a , change state to q' .

w_c : read the input call symbol $\langle c$, push it to the stack along with p , change state to q' .

w_r : when the stack is not empty, read and pop from stack a pair made of \langle_c and p , read the input return symbol r , change state to q' . In this case, the weight function ϕ_r computes the value of matching between the call and return symbols. It might be 0 to express that the symbols do not match.

w_e : when the stack is empty, read the input symbol \langle_r , change state to q' .

We give now a formal definition of these transitions of the automaton A in term of a weight value computed by an intermediate function weight_A . In the case of a pushdown automaton, a configuration is composed of a state $q \in Q$ and a stack content $\theta \in \Theta^*$, where $\Theta = \Sigma_c \times P$. Therefore, weight_A is a function from $Q \times \Theta^* \times \Sigma^* \times Q \times \Theta^*$ into \mathbb{S} .

$$\begin{aligned} \text{weight}_A([q_\theta], au, [q']_{\theta'}) &= \bigoplus_{q'' \in Q} w_i(q, a, q'') \otimes \text{weight}_A([q''_\theta], u, [q']_{\theta'}) \\ \text{weight}_A([q_\theta], \langle_c u, [q']_{\theta'}) &= \bigoplus_{\substack{q'' \in Q \\ p \in P}} w_c(q, \langle_c, q'', p) \otimes \text{weight}_A([q''_{\langle_c p \cdot \theta}], u, [q']_{\theta'}) \\ \text{weight}_A([q_{\langle_c p \cdot \theta}], r)u, [q']_{\theta'}) &= \bigoplus_{q'' \in Q} w_r(q, \langle_c, p, r, q'') \otimes \text{weight}_A([q''_\theta], u, [q']_{\theta'}) \\ \text{weight}_A([q_\perp], r)u, [q']_{\theta'}) &= \bigoplus_{q'' \in Q} w_e(q, r, q'') \otimes \text{weight}_A([q''_\perp], u, [q']_{\theta'}) \end{aligned}$$

where \perp denotes the empty stack and $\langle_c p \cdot \theta$ denotes a stack with the pair made of \langle_c and p on its top and θ as the rest of stack.

The weight associated by A to $s \in \Sigma^*$ is then defined as follows, following empty stack semantics:

$$A(s) = \bigoplus_{q, q' \in Q} \text{in}(q) \otimes \text{weight}_A([q_\perp], s, [q']_\perp) \otimes \text{out}(q').$$

2.1 1-best Computation

Regarding the infinite sum operator, it must hold that $\bigoplus_{x \in \Phi_\Sigma} \phi(x)$, $\bigoplus_{y \in \Phi_\Delta} \psi(y)$, and

property of bounded convexity of weight functions:

3 Application

Symbolic Automated Music Transcription and analysis of music performances

3.1 Time Scales

Real-Time Unit (RTU) = seconds

Musical-Time Unit (MTU) = number of measures

conversion via tempo value

3.2 Representation of Music Performances

A musical performance is represented symbolically as a finite sequence of timestamped events. similar to piano roll representation [9] chap.1
We consider an infinite alphabet Σ of MIDI-like events made of:

- a timestamp in RTU
or IOI in RTU
- a pitch value in 0..127
- ON | OFF flag
- a velocity value in 0..127

3.3 Representation of Music Scores

we consider here the case of monodies for the sake of simplicity.

A music score is represented as a structured word = sequence of quantified events + markups, see nested words [1].

We consider an alphabet Δ , every symbol of which is composed of a tag, in a finite set Ξ , and an IOI duration value. Moreover, Δ is partitioned into $\Delta = \Delta_i \uplus \Delta_c \uplus \Delta_r$, like in Section 2.

The elements of Δ_i represent events: whose tags are one of:

- continuation (0) : tie or dot
- note, grace note (pitch) or chord (pitch+)
- rest
- ...

The elements of Δ_c and Δ_r are markups for the representation of groups of events (linearization of rhythm trees [6]...) - parentheses for time divisions : triplets, bars... tag contain info such as tuple number, beaming policy...

The date or duration of events, in MTU (rational), can be computed with the markups and tags (e.g. grace note has duration 0).

There are simultaneous events, since grace notes has duration 0. They are ordered.

Finite bound on the number of duration ratio. ?

3.4 Performance/Score Distance Computation

with a transducer

References

- [1] R. Alur and P. Madhusudan. Adding nesting structure to words. *Journal of the ACM (JACM)*, 56(3):1–43, 2009.
- [2] L. D’Antoni and R. Alur. Symbolic visibly pushdown automata. In *International Conference on Computer Aided Verification*, pages 209–225. Springer, 2014.
- [3] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.
- [4] M. Droste and W. Kuich. Semirings and formal power series. In *Handbook of Weighted Automata*, pages 3–28. Springer, 2009.
- [5] L. Huang. Advanced dynamic programming in semiring and hypergraph frameworks. In *In COLING*, 2008.
- [6] F. Jacquemard, P. Donat-Bouillud, and J. Bresson. A Structural Theory of Rhythm Notation based on Tree Representations and Term Rewriting. In D. M. Tom Collins and A. Volk, editors, *Mathematics and Computation in Music: 5th International Conference, MCM 2015*, volume 9110 of *Lecture Notes in Artificial Intelligence*, page 12, London, United Kingdom, June 2015. Oscar Bandtlow and Elaine Chew, Springer.
- [7] M. Mohri. Semiring frameworks and algorithms for shortest-distance problems. *Journal of Automata, Languages and Combinatorics*, 7(3):321–350, 2002.
- [8] M. Mohri. Edit-distance of weighted automata: General definitions and algorithms. *International Journal of Foundations of Computer Science*, 14(06):957–982, 2003.
- [9] M. Müller. *Fundamentals of music processing: Audio, analysis, algorithms, applications*. Springer, 2015.
- [10] M. Veanes. Applications of symbolic finite automata. In *International Conference on Implementation and Application of Automata*, pages 16–23. Springer, 2013.

A Edit-Distance

...algebraic definition of edit-distance of Mohri, in [8] distance d over $\Sigma^* \times \Sigma^*$ into a semiring $\mathbb{S} = (\mathbb{S}, \oplus, 0, \otimes, \mathbb{1})$.

Let $\Omega = \Sigma \cup \{\epsilon\} \times \Sigma \cup \{\epsilon\} \setminus \{(\epsilon, \epsilon)\}$, and let h be the morphism from Ω^* into $\Sigma^* \times \Sigma^*$ defined over the concatenation of strings of Σ^* (that removes the ϵ 's). An *alignment* between 2 strings $s, t \in \Sigma^*$ is an element $\omega \in \Omega^*$ such that $h(\omega) = (s, t)$. We assume a base cost function $\delta : \Omega \rightarrow S$, extended to Ω^* as follows (for $\omega \in \Omega^*$): $\delta(\omega) = \bigotimes_{0 \leq i < |\omega|} \delta(\omega_i)$.

Definition 4 For $s, t \in \Sigma^*$, the edit-distance between s and t is $d(s, t) = \bigoplus_{\omega \in \Omega^* \ h(\omega) = (s, t)} \delta(\omega)$.

e.g. Levenstein edit-distance: S is min-plus and $\delta(a, b) = 1$ for all $(a, b) \in \Omega$.