Weighted Visibly Pushdown Automata and Automated Music Transcription

April 27, 2021

Abstract

...

Symbolic Weighted (SW) extension of symbolic automata where...

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 $\otimes$: $\forall x \in \mathbb{S}$, $\mathbb{0} \otimes x = x \otimes \mathbb{0} = \mathbb{0}$. In the application presented in this paper, intuitively, $\oplus$ selects an optimal value amongst two values and $\otimes$ combines two values into a single value.

A semiring $\mathbb{S}$ is *commutative* if $\otimes$ is commutative. It *bounded* [4] if $\forall x \in dom(\mathbb{S}), \mathbb{1} \oplus x = \mathbb{1}$, and *idempotent* if for all $x \in \mathbb{S}$, $x \oplus x = x$. 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 $y \leq x \otimes y$ [3]. The latter property corresponds to the *non-negative weights* condition in shortest-path algorithms [1]. 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 [2]; The above definition follows [4], 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 $dom(\mathbb{S})$ over an index set $I \subset \mathbb{N}$, the infinite sum $\bigoplus_{i \in I} x_i$ is well-defined and in $dom(\mathbb{S})$, 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 SW Visibly Pushdown Automata

We follow the approach of [5] for the computation of distances...

1.1 SW Automata and Transducers

The following definition of weighted transducers over infinite alphabets generalizes weighted transducers over finite alphabets, see e.g. [5], by considering weight functions generalizing the guards of symbolic automata

Let Σ and Δ be respectively an input and output *alphabets*, which are finite or infinite sets of symbols, and let $\mathbb{S}$ be a semiring. A *label theory* is a 4-uplet of recursively enumerable sets: Φ_0 containing constant functions valued in $\mathbb{S}$, Φ_Σ and Φ_Δ, containing unary functions in $\Sigma \to \mathbb{S}$, resp. $\Delta \to \mathbb{S}$, and $\Phi_{\Sigma,\Delta}$ containing binary functions in $\Sigma \times \Delta \to \mathbb{S}$. Moreover, we assume that each of these sets is closed under $\oplus$ and $\otimes$, 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* weighted transducer T *over the input and output alphabet* Σ *and* Δ *and the semiring* $\mathbb{S}$ *is a tuple* $T = \langle Q, \mathsf{in}, \mathsf{weight}, \mathsf{out}\rangle$, *where* Q *is a finite set of states,* $\mathsf{in} : Q \to \mathbb{S}$, *respectively* $\mathsf{out} : Q \to \mathbb{S}$, *is a function defining the weight for entering, respectively leaving, a state, and* weight *is a transition function of* $Q \times Q$ *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 weight for simplicity, such that for all $q, q' \in Q$, $a \in \Sigma$, $b \in \Delta$, and with $\langle \phi_\epsilon, \phi_\Sigma, \phi_\Delta, \phi_{\Sigma,\Delta}\rangle = \mathsf{weight}(q, q')$,

$$\begin{array}{rcl}
\mathsf{weight}(q, \epsilon, \epsilon, q') & = & \phi_\epsilon \\
\mathsf{weight}(q, a, \epsilon, q') & = & \phi_\Sigma(a) \\
\mathsf{weight}(q, \epsilon, b, q') & = & \phi_\Delta(b) \\
\mathsf{weight}(q, a, b, q') & = & \phi_{\Sigma,\Delta}(a, b)
\end{array}$$

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 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 $\mathsf{weight}_{\mathcal{A}}$ defined recursively for every $q, q' \in Q$, for every strings of $s \in \Sigma^*$, $t \in \Delta^*$:

$$\begin{aligned}
\mathsf{weight}_{\mathcal{A}}(q, s, t, q') \quad = & \quad \mathsf{weight}(q, \epsilon, \epsilon, q') \\
\oplus & \bigoplus_{\substack{q'' \in Q \\ s = au, a \in \Sigma}} \mathsf{weight}(q, a, \epsilon, q'') \otimes \mathsf{weight}_{\mathcal{A}}(q'', u, t, q') \\
\oplus & \bigoplus_{\substack{q'' \in Q \\ t = bv, b \in \Delta}} \mathsf{weight}(q, \epsilon, b, q'') \otimes \mathsf{weight}_{\mathcal{A}}(q'', s, v, q') \\
\oplus & \bigoplus_{\substack{q'' \in Q \\ s = au, a \in \Sigma \\ t = bv, b \in \Delta}} \mathsf{weight}(q, a, b, q'') \otimes \mathsf{weight}_{\mathcal{A}}(q'', u, v, q')
\end{aligned}$$

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:

$$T(s, t) = \bigoplus_{q, q' \in Q} \mathsf{in}(q) \otimes \mathsf{weight}_{\mathcal{A}}(q, s, t, q') \otimes \mathsf{out}(q').$$

A *weighted automata* $T = \langle Q, \mathsf{in}, \mathsf{weight}, \mathsf{out} \rangle$ over Σ and $\mathbb{S}$ is defined in a similar way by simply omitting the output symbols, *i.e.* weight 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 $\mathbb{S}$.

1.2 Distance between words or languages

distance d: defined over $\Sigma^* \times \Sigma^*$ into a semiring $\mathbb{S} = (\mathbb{S}, \oplus, \mathbb{0}, \otimes, \mathbb{1})$.

Edit-Distance. ...algebraic definition of edit-distance of Mohri, in [5] 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 Ω : $\delta : \Omega \to S$, extended to Ω^* as follows (for $\omega \in \Omega^*$): $\delta(\omega) = \bigotimes_{0 \leq i < |\omega|} \delta(\omega_i)$.

Definition 2 *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 $(a, b) \in \Omega$.

1.3 SW Visibly Pushdown Automata

2 Application

Symbolic Automated Music Transcription

2.1 Representations

Performance.

Score.

2.2 Transducer for Distance Computation

References

[1] E. W. Dijkstra. A note on two problems in connexion with graphs. *NUMERISCHE MATHEMATIK*, 1(1):269–271, 1959.

[2] M. Droste and W. Kuich. Semirings and formal power series. In *Handbook of Weighted Automata*, pages 3–28. Springer, 2009.

[3] L. Huang. Advanced dynamic programming in semiring and hypergraph frameworks. In *In COLING*, 2008.

[4] M. Mohri. Semiring frameworks and algorithms for shortest-distance problems. *Journal of Automata, Languages and Combinatorics*, 7(3):321–350, 2002.

[5] M. Mohri. Edit-distance of weighted automata: General definitions and algorithms. *International Journal of Foundations of Computer Science*, 14(06):957–982, 2003.