Natural Language Processing (FS 2022)

Transliteration with WFSTs

09.11.2022

1. WFSA to WFST

Any WFSA can be encoded as a WFST

learning / 0.2 - WFSA: 0/1.0 formal / 0.2 natural / 0.3 language / 0.6 is / 0.5 are / 0.5 fun / 1.0 data / 0.3 languages / 0.4 are / 1.0 (learning, learning) / 0.2 - WFST: (is, is) 1.0 0/1.0 (formal, formal) / 0.2 (gatural, natural) / 0.3 (language, language) / 0.6 (data, data) / 0.3 (is, is) / 0.5 (are, are) / 0.5 (fun, fun) / 1.0 (languages, languages) / 0.4 (are, are) / 1.0

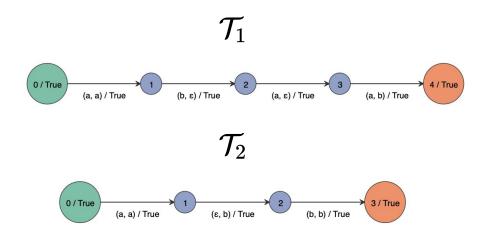
2. Transducer Composition

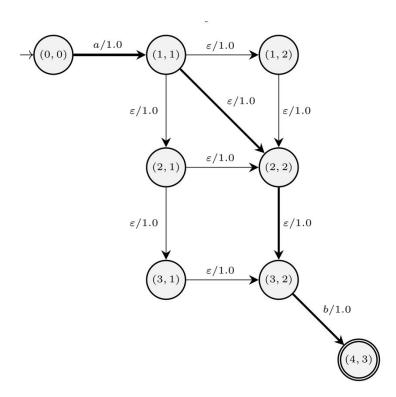
- Given two transducers $\,\mathcal{T}_1$ and $\,\mathcal{T}_2$ we can construct the composition $\,\mathcal{T}_2$

$$\mathcal{T}\left(\mathbf{x},\mathbf{y}
ight) = igoplus_{\mathbf{z} \in \mathbf{\Omega}^*} \mathcal{T}_1\left(\mathbf{x},\mathbf{z}
ight) \otimes \mathcal{T}_2\left(\mathbf{z},\mathbf{y}
ight)$$

- For each ${f Z}$, we must match one path in ${\cal T}_1$ with output ${f Z}$ and one path in ${\cal T}_2$ with input ${f Z}$
- Standard algorithms fail for $oldsymbol{\mathcal{E}}$ transitions

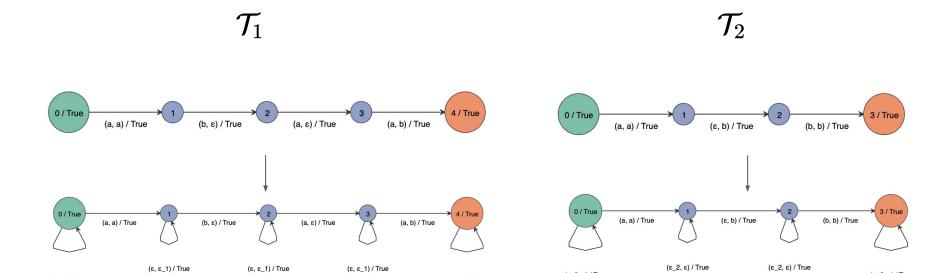
2. Transducer Composition





2. Augumentation

(ε, ε_1) / True

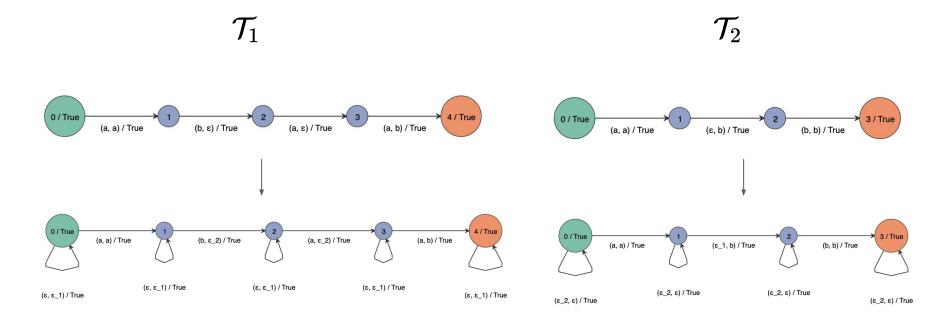


(ε, ε_1) / True

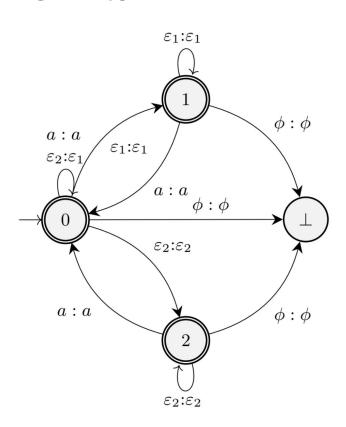
(ε_2, ε) / True

(ε_2, ε) / True

2. Relabeling



2. ε -Filter



- Each state keeps track of what actions have been performed previously.
- The blocking state ensures that only one path is allowed.
- a is any symbol in the alphabet.
- $\phi : \phi$ is anything that is not defined.