

Application and Research Highlights

# Hidden Markov Models & Finite State Transducers

Lecturer : Giorgio Satta

We introduce new finite state models that are **generalizations** of FSA

- weighted and probabilistic FSA
- hidden Markov models
- finite state transducers

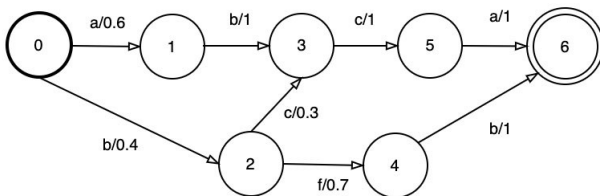
**Computational biology:** gene prediction, protein folding, DNA motif discovery, alignment of bio-sequences

**Speech recognition:** Siri & Alexa earlier versions (now replaced by neural networks)

**Machine translation:** Google translate earlier versions (now replaced by neural networks)

# Weighted FSA

In a weighted FSA, each arc is associated with a real number, called **weigh**t



The weight of a computation  $\gamma$  is the product of the weights of all its arcs  $\alpha$

$$w(\gamma) = \prod_{\alpha:\gamma} w(\alpha)$$

The weight of a string  $x \in \Sigma^*$  is the sum of the weights of all of the computations  $\gamma$  for  $x$

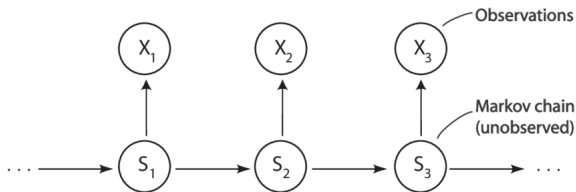
$$w(x) = \sum_{\gamma:x} w(\gamma) = \sum_{\gamma:x} \prod_{\alpha:\gamma} w(\alpha)$$

A probabilistic FSA is a weighted FSA such that the weights are **probabilities**

- real numbers in  $[0, 1]$
- at each state  $q$ , sum of probabilities at arcs leaving  $q$  is always 1

# Hidden Markov models

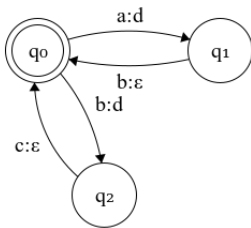
Hidden Markov models (HMM) are special types of probabilistic FSA, where input symbols are associated with states, not with arcs (technical details omitted)



Two type of probabilities: **transition** Pr and **emission** Pr

# Finite state transducers

Finite state transducers, or FST for short, have arcs where input symbols are paired with output symbols/strings

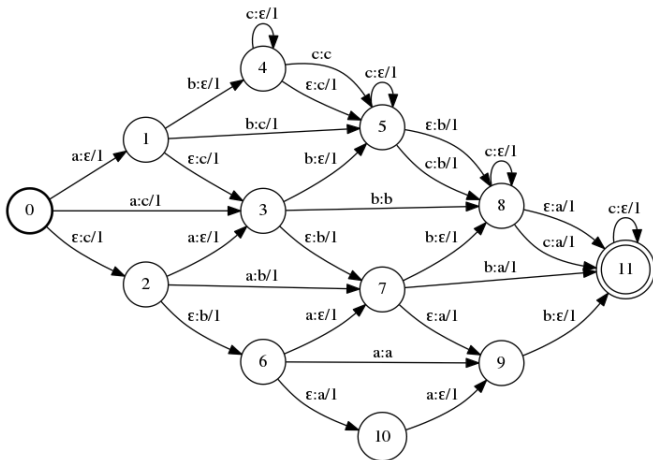


FST are **translation models**, that is, they map strings to strings



# Weighted finite state transducers

A weighted FST is a FST combined with weights (or probabilities)



Model parameters for HMM and probabilistic FST can be **estimated** from data using machine learning techniques

Several libraries implement HMM and WFST

In many applications, these finite state technologies are being replaced by **recurrent neural networks**, such as LSTM