Weighted Finite-State Transducers in Speech and Natural Language Processing
Problem Set 4:
Weighted Composition

June 22, 2009

Directions: Explain your answers and show your work. Submit all computer code (e.g., Python scripts) used in completing the problems.

Let $L(A)$ and $L(B)$ denote the languages accepted by finite-state automata $A$ and $B$, respectively. Let $L(C) = L(A) \cap L(B)$ be the language accepted by automata $C$, which is constructed by the function $\text{intersect}(A, B)$ in fsa.py.

1. Prove that the FSA intersection algorithm given in class is correct.

2. Provide the pseudocode for weighted composition.

3. Prove that the pseudocode you have written is correct.

4. Using the scripts provided as a basis, determine if the word lattice in 22_23_T18c0204_T17c0210-1.fsa includes the hypotheses:

   • <s> IT GOT TO THE POINT WHERE YOU HAD A BETTER CHANCE BEING STRUCK BY LIGHTNING THAN BEING ABLE TO SELL NEW PAPER MS. THOMPSON SAYS </s>

   • <s> IT GOT TO THE POINT WHERE YOU HAD A BETTER CHANCE OF BEING STRUCK BY LIGHTNING THAN BEING ABLE TO SELL NEW PAPER MR. WILSON SAID </s>

5. Modify the automata classes and intersection algorithm to allow for $\epsilon$-transitions on the $A$ side.

6. Modify the automata classes and intersection algorithm to perform weighted composition on the tropical semiring.