

CMPT 413

Computational Linguistics

Anoop Sarkar

<http://www.cs.sfu.ca/~anoop>

3/19/12

1

Minimum Cost Edit Distance

- String edit distance: what is the minimum number of changes (char insertions or deletions) to transform the string *intention* into *execution* ?
- Assume cost of insertion is 1 and cost of deletion is 1
- Note that we assume that we can only change one character at a time

3/19/12

2

Levenshtein Distance

- Cost is fixed across characters
 - Insertion cost is 1
 - Deletion cost is 1
- Two different costs for substitutions
 - Substitution cost is 1 (transformation)
 - Substitution cost is 2 (one deletion + one insertion)

3/19/12

3

Minimum Cost Edit Distance

- Think of it as an alignment between target and source

t_1, t_2, \dots, t_n Find $D(n, m)$ recursively
 s_1, s_2, \dots, s_m

$$D(i, j) = \min \begin{cases} D(i-1, j) & + \text{cost}(t_i, \emptyset) \text{ insertion into target} \\ D(i-1, j-1) + \text{cost}(t_i, s_j) & \text{substitution/identity} \\ D(i, j-1) & + \text{cost}(\emptyset, s_j) \text{ deletion from source} \end{cases}$$

$$\begin{aligned}
 D(0, 0) &= 0 & D(i, 0) &= D(i-1, 0) + \text{cost}(t_i, \emptyset) \\
 & & D(0, j) &= D(0, j-1) + \text{cost}(\emptyset, s_j)
 \end{aligned}$$

3/19/12

4

```

Function MinEditDistance (target, source)

n = length(target)
m = length(source)
Create matrix D of size (n+1,m+1)
D[0,0] = 0

for i = 1 to n
    D[i,0] = D[i-1,0] + insert-cost

for j = 1 to m
    D[0,j] = D[0,j-1] + delete-cost

for i = 1 to n
    for j = 1 to m
        D[i,j] = MIN(D[i-1,j] + insert-cost,
                     D[i-1,j-1] + subst/eq-cost,
                     D[i,j-1] + delete-cost)

return D[n,m]
3/19/12

```

5

target

		g	a	m	b	l	e
	0	1	2	3	4	5	6
g	1	0 _e	1	2	3	4	5
u	2	1	2 _s	3	4	5	6
m	3	2	3	2 _e	3	4	5
b	4	3	4	3	2 _e	3 _i	4
o	5	4	5	4	3	4	5 _s

source

3/19/12

6

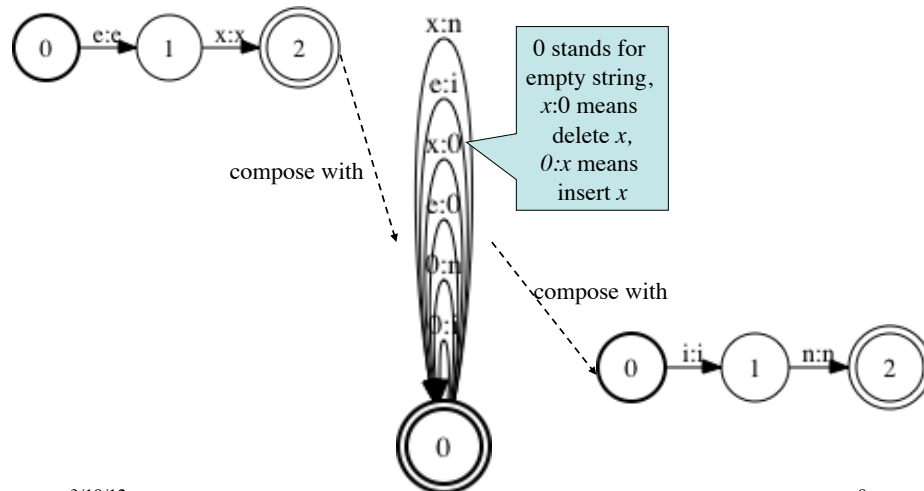
Edit Distance and FSTs

- Algorithm using a Finite-state transducer:
 - construct a finite-state transducer with all possible ways to transduce intention (source = input) into execution (target = output)
 - We do this transduction one char at a time
 - A transition $x:x$ gets zero cost and a transition on $\epsilon:x$ (insertion) or $x:\epsilon$ (deletion) for any char x gets cost 1
 - Finding minimum cost edit distance == Finding the shortest path from start state to final state

3/19/12

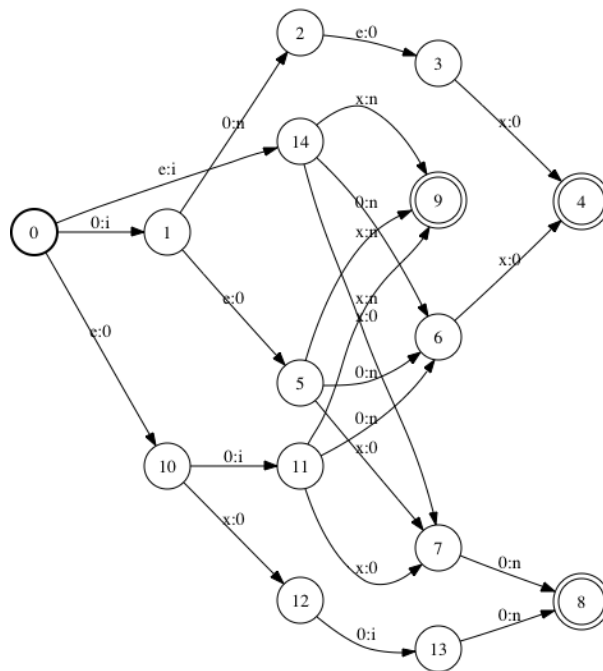
7

Edit distance and FSTs



3/19/12

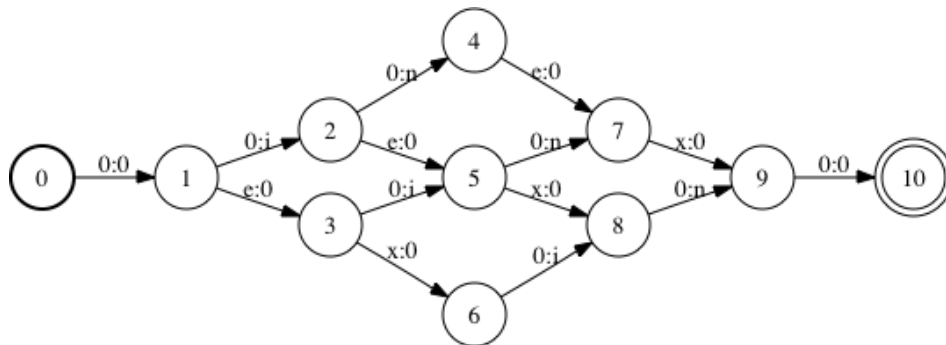
8



3/19/12

9

Edit distance and FSTs



3/19/12

10

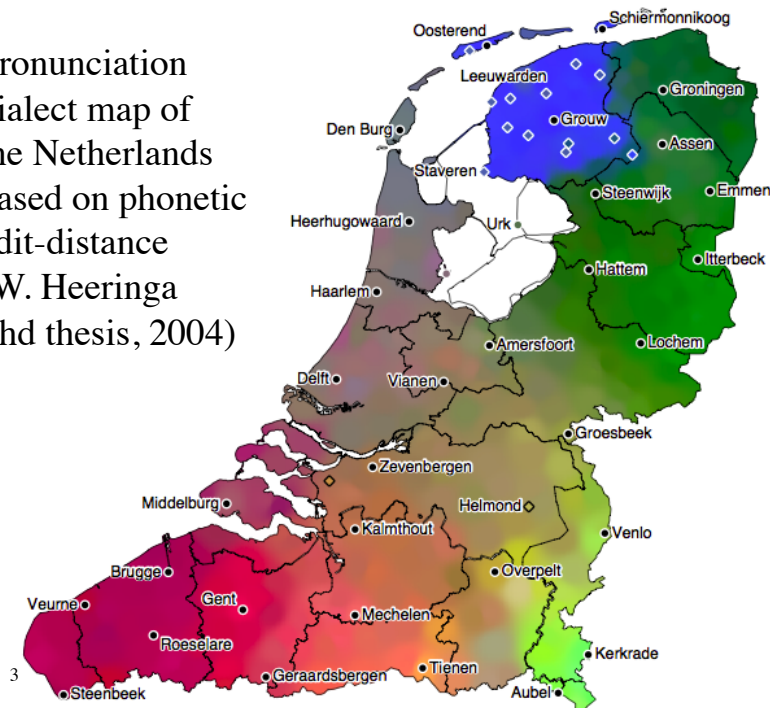
Edit distance

- Useful in many NLP applications
- In some cases, we need edits with multiple characters, e.g. 2 chars deleted for one cost
- Comparing system output with human output, e.g. input: ibm output: IBM vs. Ibm (TrueCasing of speech recognition output)
- Error correction
- Defined over character edits or word edits, e.g. MT evaluation:
 - Foreign investment in Jiangsu 's agriculture on the increase
 - Foreign investment in Jiangsu agricultural investment increased

3/19/12

11

Pronunciation
dialect map of
the Netherlands
based on phonetic
edit-distance
(W. Heeringa
Phd thesis, 2004)



3

12

Variable Cost Edit Distance

- So far, we have seen edit distance with uniform insert/delete cost
- In different applications, we might want different insert/delete costs for different items
- For example, consider the simple application of spelling correction
- Users typing on a qwerty keyboard will make certain errors more frequently than others
- So we can consider insert/delete costs in terms of a probability that a certain alignment occurs between the *correct* word and the *typo* word

3/19/12

13

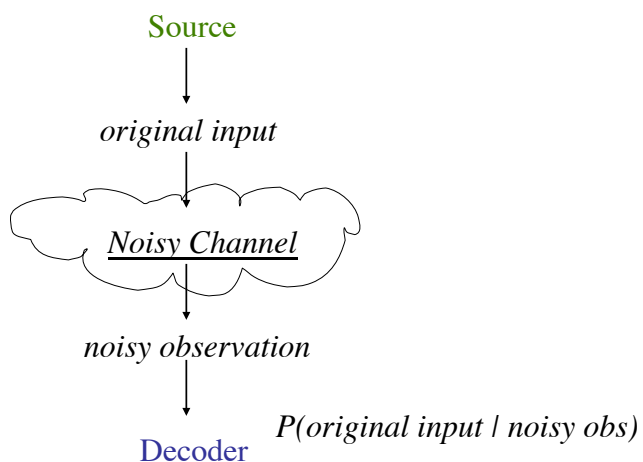
Spelling Correction

- Types of spelling correction
 - non-word error detection
e.g. *hte* for *the*
 - isolated word error detection
e.g. *acres* vs. *access* (cannot decide if it is the right word for the context)
 - context-dependent error detection (real world errors)
e.g. *she is a talented acres* vs. *she is a talented actress*
- For simplicity, we will consider the case with exactly 1 error

3/19/12

14

Noisy Channel Model



3/19/12

15

Bayes Rule: *computing $P(\text{orig} \mid \text{noisy})$*

- let $x = \text{original input}$, $y = \text{noisy observation}$

$$p(x \mid y) = \frac{p(x, y)}{p(y)} \quad p(y \mid x) = \frac{p(y, x)}{p(x)}$$

$$p(x, y) = p(y, x)$$

$$p(x \mid y) \times p(y) = p(y \mid x) \times p(x)$$

$$p(x \mid y) = \frac{p(y \mid x) \times p(x)}{\cancel{p(y)}} \quad \underline{\text{Bayes Rule}}$$

3/19/12

16

Chain Rule

$$p(a,b,c \mid d) = p(a \mid b,c,d) \times \\ p(b \mid c,d) \times \\ p(c \mid d)$$

Approximations: Bias vs. Variance

$$p(a \mid b,c,d) \approx \frac{p(a \mid b,c)}{p(a \mid b)} \text{ less } \textit{bias} \\ p(a) \text{ less } \textit{variance}$$

3/19/12

17

Single Error Spelling Correction

- Insertion (addition)
 - acress vs. cress
- Deletion
 - acress vs. actress
- Substitution
 - acress vs. access
- Transposition (reversal)
 - acress vs. caress

3/19/12

18

Noisy Channel Model for Spelling Correction (Kernighan, Church and Gale, 1990)

- t is the word with a single typo and c is the correct word

$$P(c | t) = p(t | c) \times p(c) \quad \text{Bayes Rule}$$

- Find the best candidate for the correct word

$$\hat{c} = \arg \max_{c \in C} P(t | c) \times P(c)$$


$$P(t | c) = ?? \quad P(c) = \frac{f(c)}{N}$$

3/19/12

C is all the words in the vocabulary; |C| = N

19

Noisy Channel Model for Spelling Correction (Kernighan, Church and Gale, 1990) single error, condition on previous letter




$P(\text{potion} | \text{potion})$

$P(t | c) =$

$$\left\{ \begin{array}{l} \frac{\text{del}[c_{p-1}, c_p]}{\text{chars}[c_{p-1}, c_p]} (xy)_c \text{ typed as } (x)_t \\ \frac{\text{ins}[c_{p-1}, t_p]}{\text{chars}[c_{p-1}]} (x)_c \text{ typed as } (xy)_t \\ \frac{\text{sub}[t_p, c_p]}{\text{chars}[c_p]} (y)_c \text{ typed as } (x)_t \\ \frac{\text{rev}[c_p, c_{p+1}]}{\text{chars}[c_p, c_{p+1}]} (xy)_c \text{ typed as } (yx)_t \end{array} \right.$$

$t = \text{potion}$
 $c = \text{potion}$
 $\text{del}[t, i] = 427$
 $\text{chars}[t, i] = 575$
 $P = .7426$

$P(\text{potion} | \text{piton})$



$t = \text{potion}$
 $c = \text{piton}$
 $\text{sub}[o, i] = 568$
 $\text{chars}[i] = 1406$
 $P = .4039$

3/19/12

20

Noisy Channel model for Spelling Correction

- The *del*, *ins*, *sub*, *rev* matrix values need data in which contain known errors
(training data)
e.g. Birbeck spelling error corpus (from 1984!)
- Accuracy on single errors on unseen data
(test data)

3/19/12

21

Noisy Channel model for Spelling Correction

- Easily extended to multiple spelling errors in a word using edit distance algorithm (however, using learned costs for ins, del, replace)
- Experiments: 87% accuracy for machine vs. 98% average human accuracy
- What are the limitations of this model?
... *was called a “stellar and versatile **acress** whose combination of sass and glamour has defined her*
...

KCG model best guess is **acres**

3/19/12

22