

LING/C SC/PSYC 438/538

Lecture 18

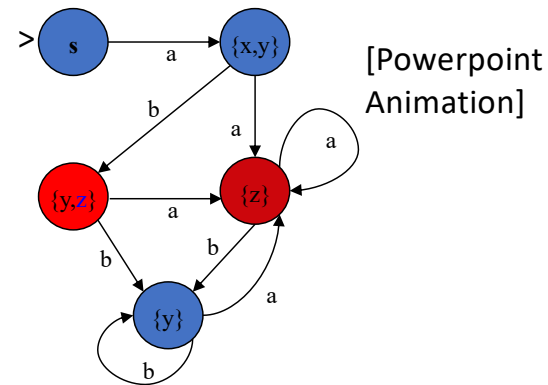
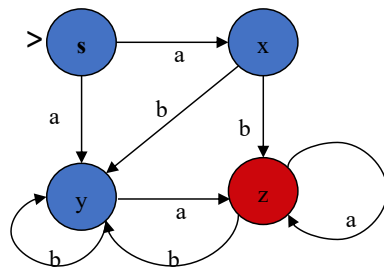
Prof. Sandiway Fong

Today's Topics

- FSA *contd.*
 - recall **non-determinism**?
 - NDFSA to DFSA conversion
- Markov Chains
- Homework 11: FSA Homework

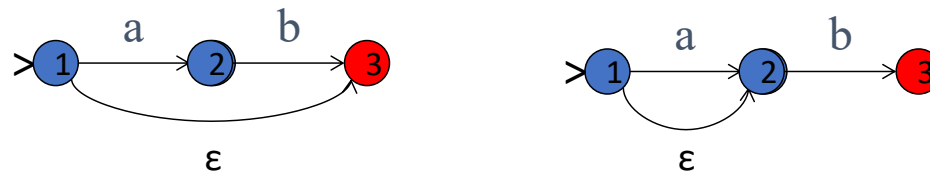
Last Time: NDFSA $\rightarrow$ DFSA

- **Idea:** construct a new machine
 - each state of the new machine represents the **set of possible states** of the original machine when stepping through the input alphabet (Σ)
- **Notes:**
 - new machine is equivalent to old one (*but has more states*)
 - new machine is deterministic!
- **example**



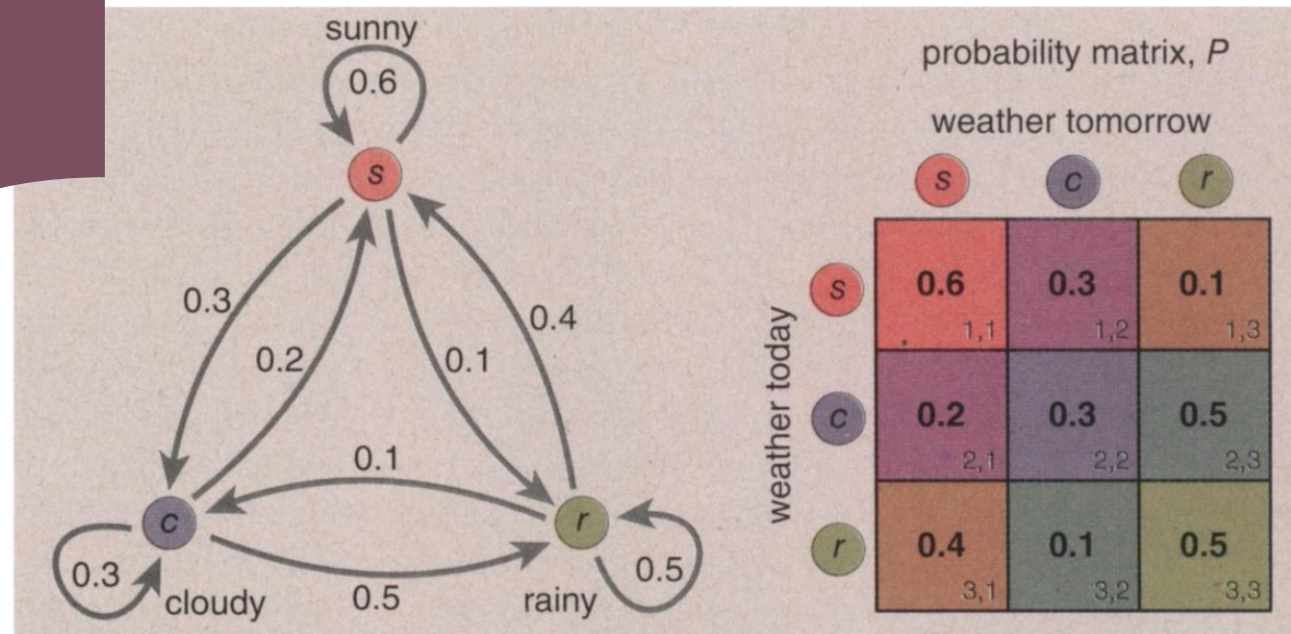
Worked EXERCISE

- Let's check our understanding:
 - apply the set-of-states construction technique to the two machines on the ϵ -transition slide from the previous lecture (*repeated below*)
- How to check your answer?
 - should confirm the machine produced is actually **deterministic** and accurately simulates its ϵ -transition counterpart



Markov Chain:
when events
are *not*
independent

Note: exit transition **probabilities** must sum to 1 (otherwise just *weights*)
Any state here can be a start state!



(Hayes 2013)

Markovian Assumption

Markov's scheme for extending the laws of probability beyond the realm of independent variables has one crucial restriction: The probabilities must depend only on the present state of the system, not on its earlier history. The Markovian analysis of Monopoly, for example, considers a player's current position on the board but not how he or she got there. This limitation is serious. After all, life presents itself as a long sequence of contingent events—kingdoms are lost for want of a nail, hurricanes are spawned by butterflies in Amazonia—but these causal chains extending into the distant past are not Markov chains.

Hayes (2013)

20K letter sequence (Russian): _____

- 8638 vowels / 11362 consonants
- 1104 vowel | vowel (by counting: *circling* (?)
- p_1 vowel | vowel = 0.128
- p_0 vowel | consonant = 0.663
- computed by:
 - $8638 - 1104 = 7534$ (vowel | consonants);
 - $7534 / 11362 = 0.663$
- q_0 consonant | consonant = 0.337
 - $11362 - 7534 = 3827$ (consonant | consonant);
 - $3827 / 11362 = 0.337$
- q_1 consonant | vowel = 0.872
 - $7534 / 8638 = 0.872$

Markov, A. A. 1913. *An example of statistical investigation of the text Eugene Onegin concerning the connection of samples in chains (in Russian)*. Bulletin of the Imperial Academy of Sciences of St. Petersburg 7(3):153-162.

Investigation of a Remarkable Case of Dependent Samples" [Markov 1907a]

Eugene Onegin

Russian (33 letters):

- twenty consonants: ⟨б⟩, ⟨в⟩, ⟨г⟩, ⟨д⟩, ⟨ж⟩, ⟨з⟩, ⟨к⟩, ⟨л⟩, ⟨м⟩, ⟨н⟩, ⟨п⟩, ⟨р⟩, ⟨с⟩, ⟨т⟩, ⟨ф⟩, ⟨х⟩, ⟨ц⟩, ⟨ч⟩, ⟨ш⟩, ⟨щ⟩
- ten vowels: ⟨а⟩, ⟨е⟩, ⟨ё⟩, ⟨и⟩, ⟨о⟩, ⟨у⟩, ⟨ы⟩, ⟨э⟩, ⟨ю⟩, ⟨я⟩
- a semivowel / consonant (⟨й⟩)
- two modifier letters or "signs" ⟨ъ⟩, ⟨ь⟩

20K letters (*Eugene Onegin*, in Russian, 1st chapter, 16 stanzas of the 2nd chapter).

- vowels: 8638; consonants: 11362
- $p(\text{vowel}) = 0.43$
- $p(\text{vowel}|\text{vowel}) = 0.43^2 = 0.19$ (assuming independence)
- $p(\text{consonant}|\text{consonant}) = 0.32$ (assuming independence)

ЕВГЕНІЙ ОНЕГІНЪ,

РОМАНЪ ВЪ СТИХАХЪ.

САНКТПЕТЕРБУРГЪ.

Въ типографіи Александра Смирдина.

1833.

Eugene Onegin (English translation)

$P(x|y)$ means
probability of event x
given event y (has
happened)

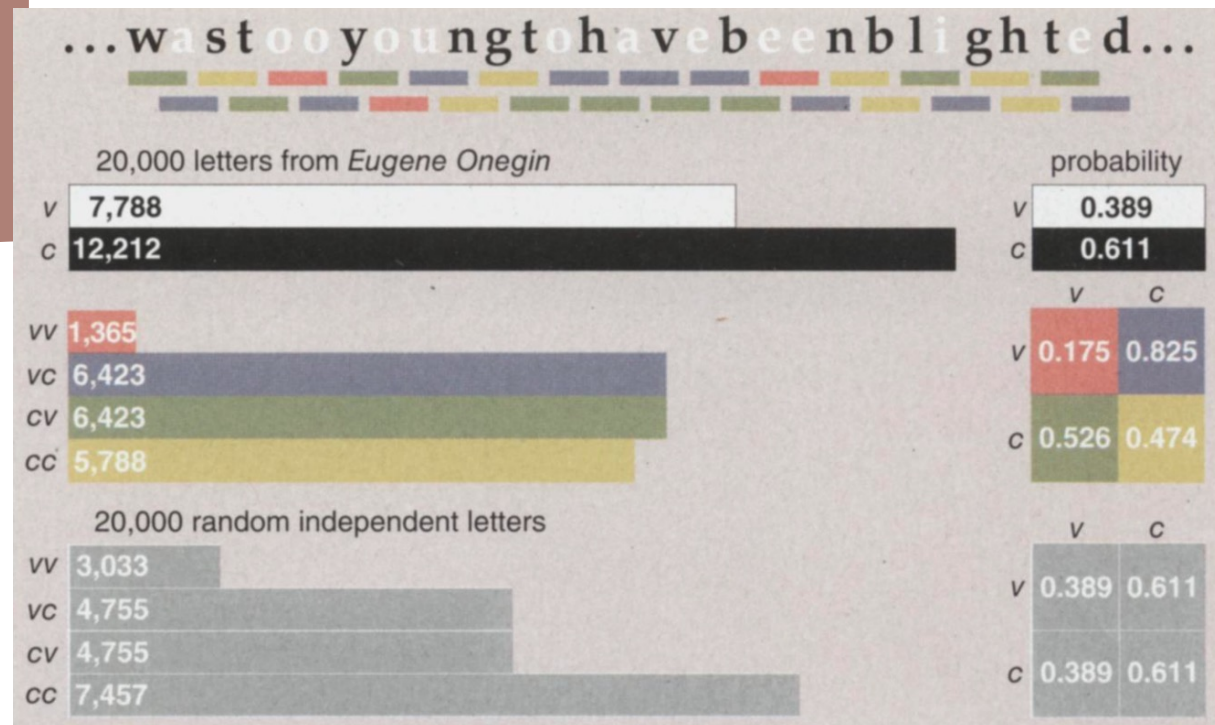
Table cells represent

$P(v|c)$

$P(v|v)$

$P(c|v)$

$P(c|c)$



Markovian Assumption in real life?

- 13th C poet:

- *The wise tell us that a nail keeps a shoe, a shoe keeps a horse, a horse keeps a knight, a knight, who can fight, keeps a castle.*

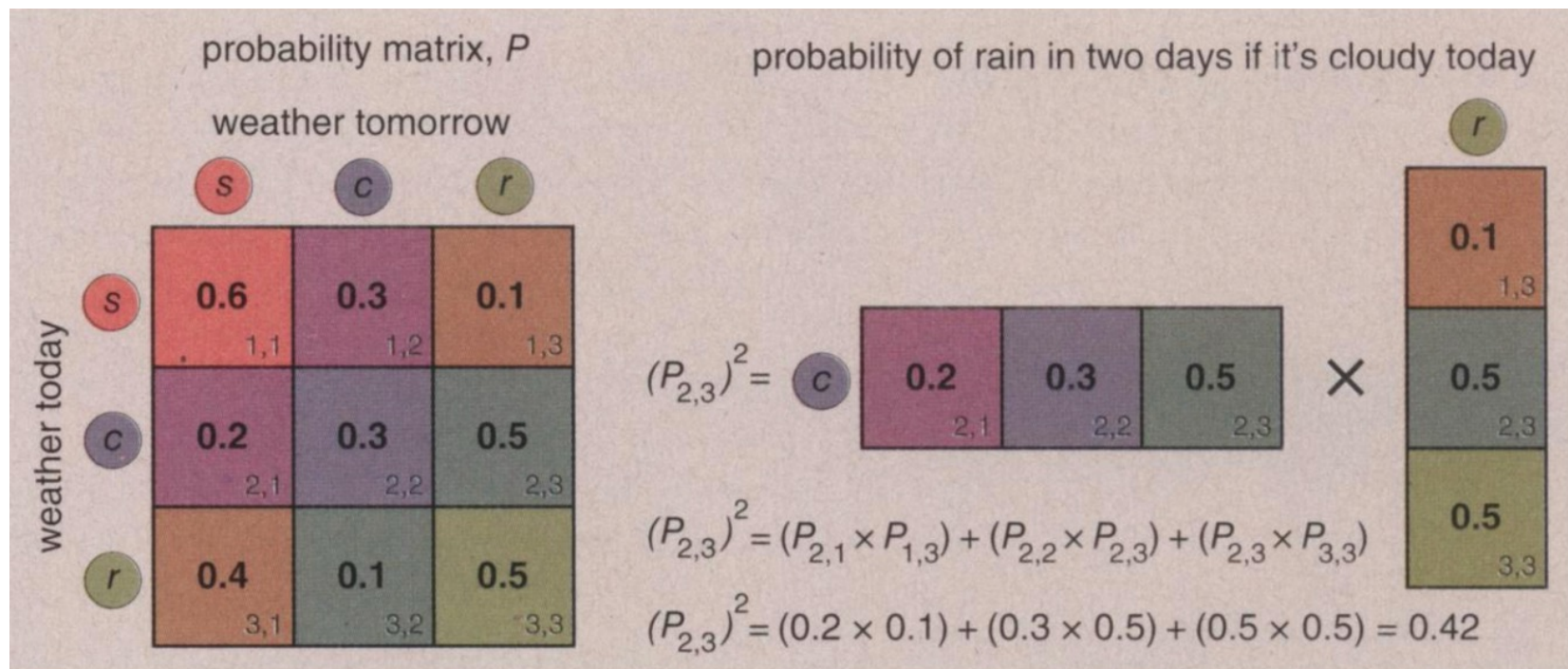
- Shakespeare:

- *A horse! a horse! my kingdom for a horse!*
- (Richard III, Act V, Scene IV)

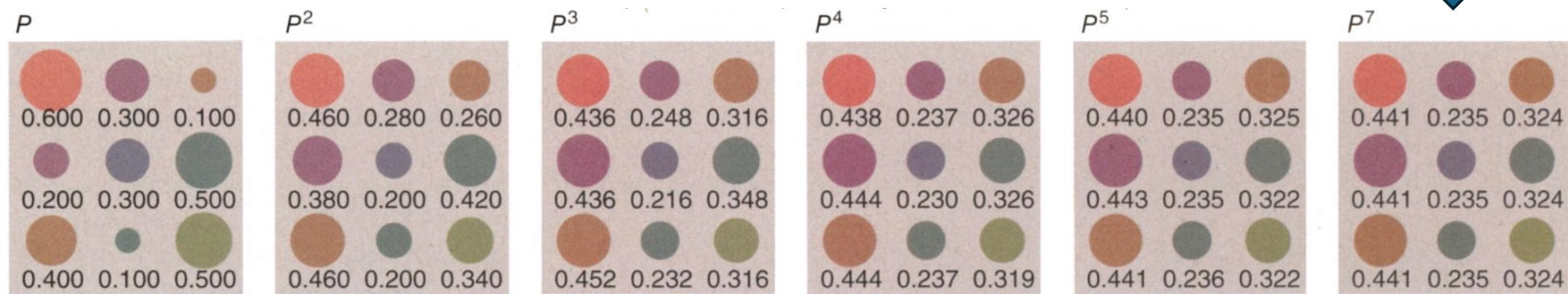
- Butterfly Effect:

- *Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?*
- E.N. Lorenz (Chaos theory mathematician)
- *weather*: 0.506 vs. 0.506¹²⁷ resulted in a completely different simulation outcome (1961)

Markov Chain



Markov Chain



Text Generation by Markov Chain

First order

*Theg sheso pa lyiklg ut. cout Scrpauscricre cobaives wingervet Ners, whe ilened te o
wn taulie wom uld atimore rteansouroocono weveiknt hef ia ngry'sif farll t mmat and,
tr iscond frnid riliofr th Gureckpeag*

Third order

*At oness, and no fall makestic to us, infessed Russion-bently our then a man thous al-
ways, and toops in he roquestill shoed to dispric! Is Olga's up. Italked fore declaimsel
the Juan's conven night toget nothem,*

Fifth order

*Meanwhile with jealousy bench, and so it was his time. But she trick. Let message
we visits at dared here bored my sweet, who sets no inclination, and Homer, so prose,
weight, my goods and envy and kin.*

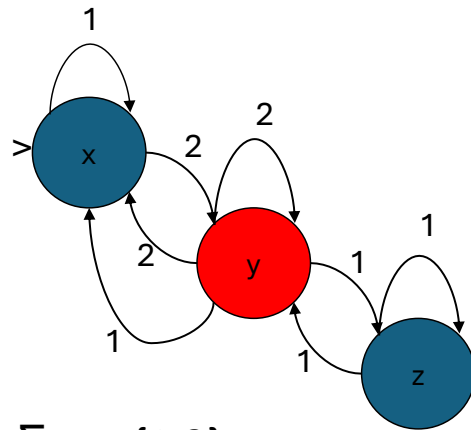
Seventh order

*My sorrow her breast, over the dumb torment of her veil, with our poor head is stoop-
ing. But now Aurora's crimson finger, your christening glow. Farewell. Evgeny loved
one, honoured fate by calmly, not yet seeking?*

In a k th-order random-text Markov model the states are sequences of k characters, and transitions are defined for all characters that can follow each of these sequences.

Homework 11

- Consider the following NDFSA: Q1: Why is it a NDFSA?



$\Sigma = \{1,2\}$

- Q2: What strings of length 3 from Σ^* does it accept? How many are there?
- Q3: Which strings of length 4 from Σ^* does it **not** accept? How many are there?
- Q4: Convert our NDFSA into a DFSA. How many states does the DFSA have? How many final states?
 - use the construction shown in class

Homework 11

- 538 Q5 (438 Extra Credit):
 - Implement your DFSA in Perl or Python (only)
 - Then devise a way to calculate how many strings there are in the language (of the DFSA) with string length **not exceeding** 10?
 - e.g. 221 (length 3); 121122 (length 6)
 - Generally, given $\Sigma=\{1,2\}$ the number of strings of length n is 2^n .
 - $2^1 = \{^*1, 2\}$
 - $2^2 = \{^*11, 12, ^*21, 22\}$
 - **Bash** (terminal):
 - `$ eval echo ${printf '{1,2}%.0s' $(seq 3)}`
 - 111 112 121 122 211 212 221 222

Homework 11

- Usual rules ...
 - One PDF file (*no links to file*)
 - Subject: 438/538 Homework 11 *YOUR NAME*
 - Due date: Sunday midnight
- You can draw your machine for Q4 by hand (*make it legible*)
 - put in the set of states {...}
- For the 538 Q5 question:
 - copy/paste your Perl/Python code into your PDF file
 - attachments don't work well with UA Mail nuking them