The

Five Greatest Applications

of

Markov Chains

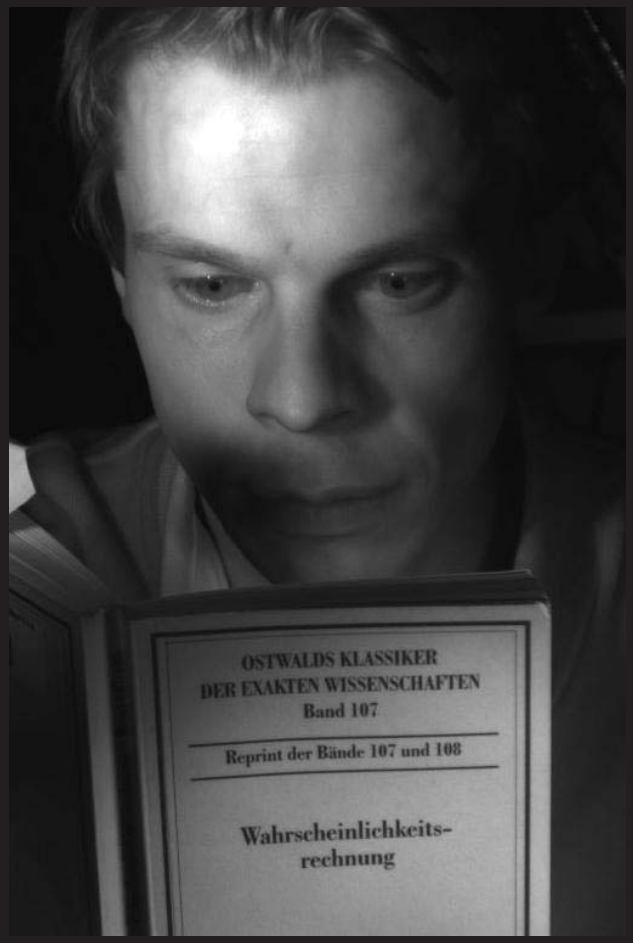
Amy Langville
College of Charleston
langvillea@cofc.edu

Philipp von Hilgers
Max Planck Institute
philgers@mpiwg-berlin.mpg.de

Markov Anniversary Meeting
6/13/2006

Philipp von Hilgers

working



pre-wedding modeling

Outline

Applications in Chronological Order

- Eugeny Onegin
- Information Theory
- Computer Performance Evaluation
- Hidden Markov Models
- PageRank

Proper Order

- Alphabetical?
- Markov number?
- Impact Factor?
- Our Method

Eugeny Onegin (1913)

- In 1913 Markov expands the 3rd edition of his book, "The Calculus of Probabilities," to celebrate the 200th anniversary of Jakob Bernoulli's "Ars Conjectandi."

- 3rd edition closes with Markov's famous application of chains.

  > "Let us finish the article and the whole book with a good example of the dependent trials, which approximately can be regarded as a simple chain."

- Markov studied sequence of 20,000 letters in A.S. Pushkin's poem "Eugeny Onegin"

- Markov also presented this work at the physico-mathematical meeting of the Academy of Sciences in 1913.

A.S.Pushkin. Eugeny Onegin (1-3 chapter)

* CHAPTER I

I My uncle was a man of virture, When he became quite old and sick, He sought respect and tried to teach me, His only heir, verte and weak. He had the fun, I had the sore, But grecious goodness! what a bore! To sit by bedplace day and night, Not doing even step aside, And what a cheep and cunning thing To entertain the sad, To serve around, make his bed, To fetch the pills, to mourn and grim, To sigh outloud, think along: `God damn old man, why ain't you gone?'	Мой дядя самых честных п равил, Когда не в шутку занемог, Он уважать себя заставил И лучше выдумать не мог. Его пример другим наука; Но, боже мой, какая скука С больным сидеть и днь и ночь Не отходя ни шагу прочь! Какое низкое коварство Полуживого забавлять, Ему подушки поправлять, Печально подносить лекарство, Вздыхать и думать про себя: Когда же черт возьмет тебя!
II So thought a playboy, young and funny, While riding through the dust of road, The only heir to the money, That got his folks with help of Lord.	Так думал молодой повеса, Летя в пыли на почтовых, Всевышней волею Зевеса Наследник всех своих родных. Друзья Людмилы и Руслана! С героем моего романа

Original: (english translation, all special characters except space removed, converted to small case):

my uncle a most worthy gentleman
when he fell seriously ill
constrained everyone to respect him
couldnt have done better if he tried
his behaviour was a lesson to us all
but god above what crushing boredom
to sit with the malingerer night and day
not moving even one footstep away
what demeaning hypocrisy
to amuse the halfdead codger
to fluff up his pillows and then
mournfully to bring him his medicine
to think to oneself and to sigh
when the devil will the old rascal die
so thought our young neerdowell
flying through the dust on a mail coach
by the supreme will of zeus
he was the inheritor of all his kin
good friends of ludmilla and ruslan
with the hero of my romance
allow me to make you acquainted ...

Top left corner of doublet count matrix:

	a	b	c	d	e	f	.	.
a	1	23	26	58	4	2		
b	16	1	0	0	51	0		
c	31	0	4	0	73	0		
d	36	0	2	7	76	0		
e	123	2	23	118	63	14		
f	31	0	0	2	38	17		
.								
.								

Top left corner of stochastic matrix:

	a	b	c	d	e	f	.	.
a	0.001	0.021	0.023	0.052	0.004	0.002		
b	0.081	0.005	0.000	0.000	0.259	0.000		
c	0.092	0.000	0.012	0.000	0.217	0.000		
d	0.056	0.000	0.003	0.011	0.119	0.000		
e	0.070	0.001	0.013	0.068	0.036	0.008		
f	0.083	0.000	0.000	0.005	0.102	0.046		
.								
.								

Random text generated by Markov chain:

abomy in tyoworeadshe ieabllllothinear tim s wifon cow ondowand abules ture in sethioriffour sshincore buspid re tome ifound ry ien antha winglysth f nd the a ncik othted f fle ttoulead thapagousud conghinshe r che i arer avelveg ornot tourig m dettonge ots bow ouco freed tyondevinglerolisps tul ad mo auspando mserin lisuriaset lurisy wigld wichare toutais ddethent tlouteunde anonlend be t s hthules ad iem ded adch he tly hean hesms ita oureliand te ppsushis atrues amby f hilell avecol t chasere alembow cotr itrever morericomanoffevetthin om ...

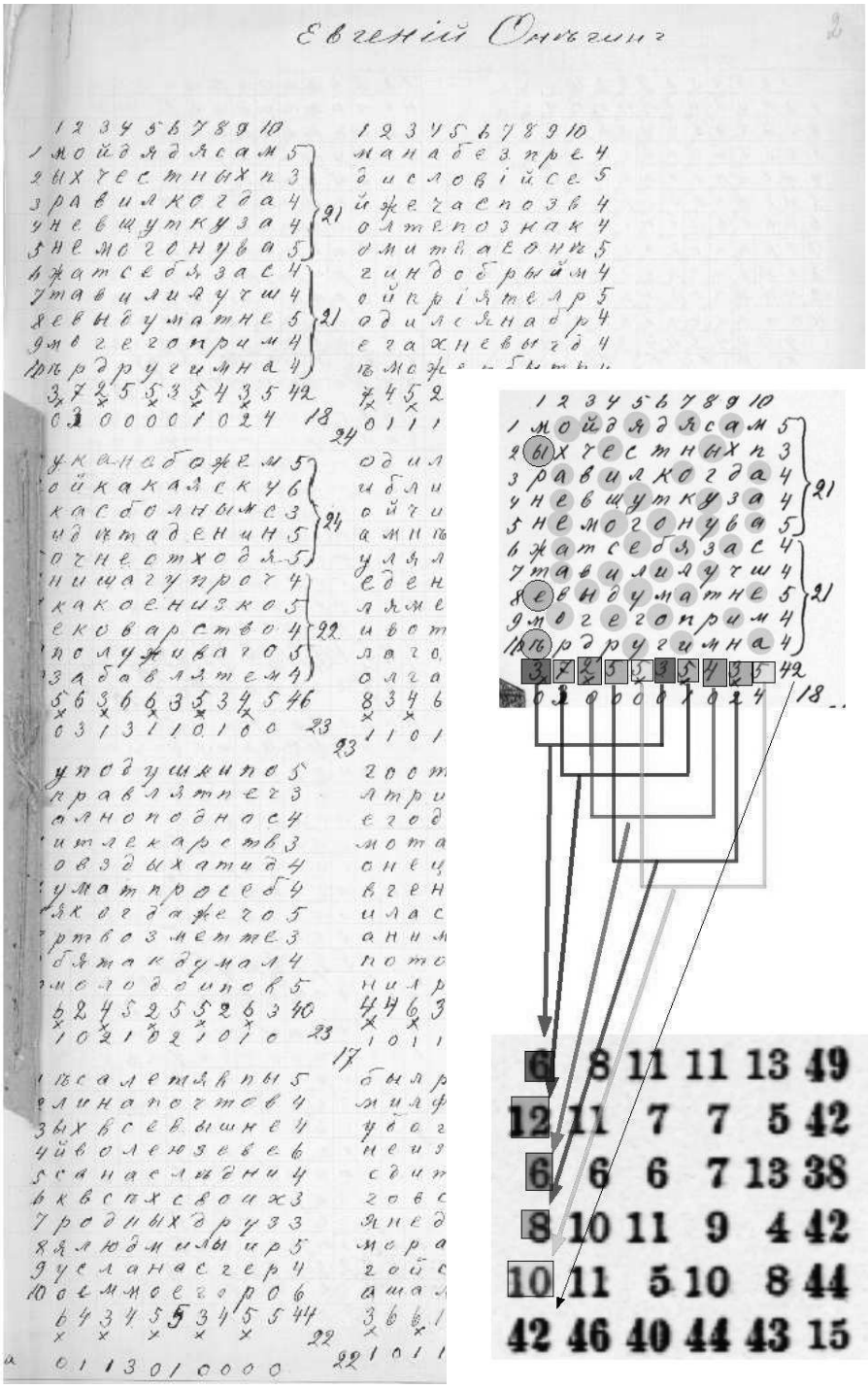


FIG. 2.1. *Left background: The first 800 letters of 20,000 total letters compiled by Markov and taken from the first one and a half chapters of Pushkin's poem "Eugeny Onegin." Markov omitted spaces and punctuation characters as he compiled the cyrillic letters from the poem. Right foreground: Markov's count of vowels in the first matrix of 40 total matrices of* 10×10 *letters. The last row of the* 6×6 *matrix of numbers can be used to show the fraction of vowels appearing in a sequence of 500 letters. Each column of the matrix gives more information. Specifically, it shows how the sums of counted vowels are composed by smaller units of counted vowels. Markov argued that if the vowels are counted in this way, then their number proved to be stochastically independent.*

Calculations for Eugeny Onegin

"The second question concerns an original statistical investigation which I have carried out and with which I propose to conclude my book. The character of the investigation, which embraces a sequence of 20,000 letters, is shown in the example below."

(Jan. 15, 1913 letter to Chuprov)

stationary probability of vowel = .432
stationary probability of consonant = .568
probability of a vowel following a vowel = .128
probability of a vowel following a consonant = .663

Besikovitch on Markov's book:

"great attention is paid to the simplest numerical examples which are discussed in unusual detail. And further, it is hardly possible to find a single mistake in these examples."

Markov on calculating:

"Many mathematicians apparently believe that going beyond the field of abstract reasoning into the sphere of effective calculations would be humiliating."

Table for Normal Distribution

produced a table with up to 11 digits. Table was used into the 1940s.

Tabelle der Werte von $\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt.$

x		0	1	2	3	4	5	6	7	8	9
1,90	0,99	2790	2821	2851	2881	2912	2942	2972	3001	3031	3061
1,91	0,99	3090	3119	3148	3178	3207	3235	3264	3293	3321	3350
1,92	0,99	3378	3406	3435	3463	3490	3518	3546	3574	3601	3628
1,93	0,99	3656	3683	3710	3737	3764	3790	3817	3844	3870	3896
1,94	0,99	3923	3949	3975	4001	4026	4052	4078	4103	4129	4154
1,95	0,99	4179	4204	4229	4254	4279	4304	4329	4353	4378	4402
1,96	0,99	4426	4450	4475	4498	4522	4546	4570	4593	4617	4640
1,97	0,99	4664	4687	4710	4733	4756	4779	4802	4824	4847	4870
1,98	0,99	4892	4914	4937	4959	4981	5003	5025	5047	5068	5090
1,99	0,99	5111	5133	5154	5176	5197	5218	5239	5260	5281	5302
2,00	0,99	5322	5343	5363	5384	5404	5425	5445	5465	5485	5505
2,01	0,99	5525	5545	5564	5584	5604	5623	5643	5662	5681	5700
2,02	0,99	5719	5738	5757	5776	5795	5814	5832	5851	5870	5888
2,03	0,99	5906	5925	5943	5961	5979	5997	6015	6033	6050	6068
2,04	0,99	6086	6103	6121	6138	6156	6173	6190	6207	6224	6241
2,05	0,99	6258	6275	6292	6308	6325	6342	6358	6375	6391	6407
2,06	0,99	6423	6440	6456	6472	6488	6504	6519	6535	6551	6567
2,07	0,99	6582	6598	6613	6628	6644	6659	6674	6689	6704	6719
2,08	0,99	6734	6749	6764	6779	6794	6808	6823	6837	6852	6866
2,09	0,99	6880	6895	6909	6923	6937	6951	6965	6979	6993	7007
2,10	0,99	7021	7034	7048	7061	7075	7088	7102	7115	7128	7142
2,11	0,99	7155	7168	7181	7194	7207	7220	7233	7246	7258	7271
2,12	0,99	7284	7296	7309	7321	7334	7346	7358	7371	7383	7395
2,13	0,99	7407	7419	7431	7443	7455	7467	7479	7490	7502	7514
2,14	0,99	7525	7537	7548	7560	7571	7583	7594	7605	7616	7627
2,15	0,99	7639	7650	7661	7672	7683	7693	7704	7715	7726	7737
2,16	0,99	7747	7758	7768	7779	7789	7800	7810	7820	7831	7841
2,17	0,99	7851	7861	7871	7881	7891	7901	7911	7921	7931	7941
2,18	0,99	7951	7960	7970	7980	7989	7999	8008	8018	8027	8037
2,19	0,99	8046	8055	8065	8074	8083	8092	8101	8110	8119	8128
2,20	0,99	8137	8146	8155	8164	8173	8181	8190	8199	8207	8216
2,21	0,99	8224	8233	8241	8250	8258	8267	8275	8283	8292	8300
2,22	0,99	8308	8316	8324	8332	8340	8348	8356	8364	8372	8380
2,23	0,99	8388	8396	8403	8411	8419	8426	8434	8442	8449	8457
2,24	0,99	8464	8472	8479	8486	8494	8501	8508	8516	8523	8530
2,25	0,99	8537	8544	8552	8559	8566	8573	8580	8586	8593	8600
2,26	0,99	8607	8614	8621	8627	8634	8641	8648	8654	8661	8667
2,27	0,99	8674	8680	8687	8693	8700	8706	8712	8719	8725	8731
2,28	0,99	8738	8744	8750	8756	8762	8768	8775	8781	8787	8793
2,29	0,99	8799	8805	8810	8816	8822	8828	8834	8840	8845	8851
2,30	0,99	8857	8862	8868	8874	8879	8885	8890	8896	8902	8907
2,31	0,99	8912	8918	8923	8929	8934	8939	8945	8950	8955	8960
2,32	0,99	8966	8971	8976	8981	8986	8991	8996	9001	9006	9011
2,33	0,99	9016	9021	9026	9031	9036	9041	9045	9050	9055	9060
2,34	0,99	9065	9069	9074	9079	9083	9088	9093	9097	9102	9106
2,35	0,99	9111	9115	9120	9124	9129	9133	9137	9142	9146	9150
2,36	0,99	9155	9159	9163	9168	9172	9176	9180	9184	9189	9193
2,37	0,99	9197	9201	9205	9209	9213	9217	9221	9225	9229	9233
2,38	0,99	9237	9241	9245	9249	9252	9256	9260	9264	9268	9271
2,39	0,99	9275	9279	9282	9286	9290	9293	9297	9301	9304	9308

Reaction to Markov's Application

- Nikolai A. Morozov at a 1915 meeting of the Academy of Sciences in St. Petersburg called Markov's method

  "a new weapon for the analysis of ancient scripts."

- To demonstrate Morozov provided some statistics that could help identify the style of some authors.

- Markov found Morozov's experiments unconvincing, but did mention that a more advanced model and extended data set might enable author identification solely by mathematical analysis of this writings.

- By 1916 Markov's vision was nearly completely lost to glaucoma, and his analysis of texts ceased.

Information Theory (1948)

- Claude E. Shannon's work "A Mathematical Theory of Communication" created the field of information theory.

- Idea: any source transmitting data is a Markov process.

- In a communication system with symbols such as letters, one must determine transition probabilities for a letter following any other letter. (e.g., prob. of "u" following "q" is high)

- Shannon used these chains to
  - predict communication channel load and optimize utilization of bandwidth.
  - generate text using n-grams. First attempt with tri-grams resulted in sentence:

Information Theory (1948)

- Claude E. Shannon's work "A Mathematical Theory of Communication" created the field of information theory.

- Idea: any source transmitting data is a Markov process.

- In a communication system with symbols such as letters, one must determine transition probabilities for a letter following any other letter. (e.g., prob. of "u" following "q" is high)

- Shannon used these chains to
  - predict communication channel load and optimize utilization of bandwidth.
  - generate text using n-grams. First attempt with tri-grams resulted in sentence:

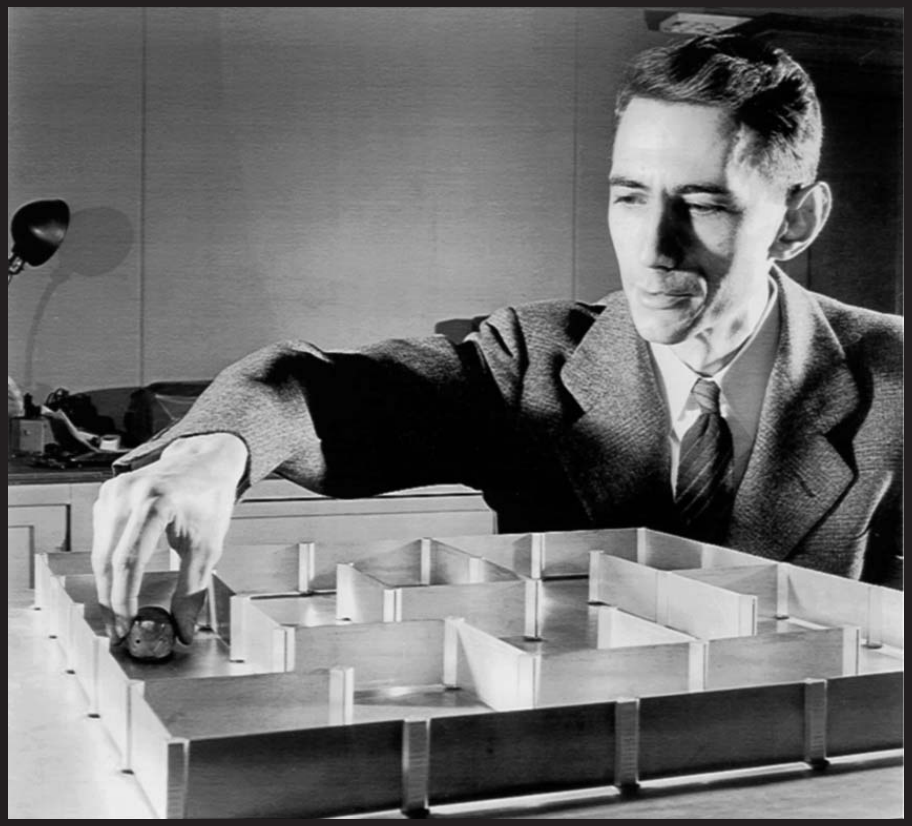
    IN NO IST LAT WHEY CRATICT FOURE BIRS GROCID.

- Markov chains have since been used to generate text, images, and music, as well as in speech recognition.

Shannon and his Machines

A tinkerer, Shannon created several electromechanical machines.

Theseus

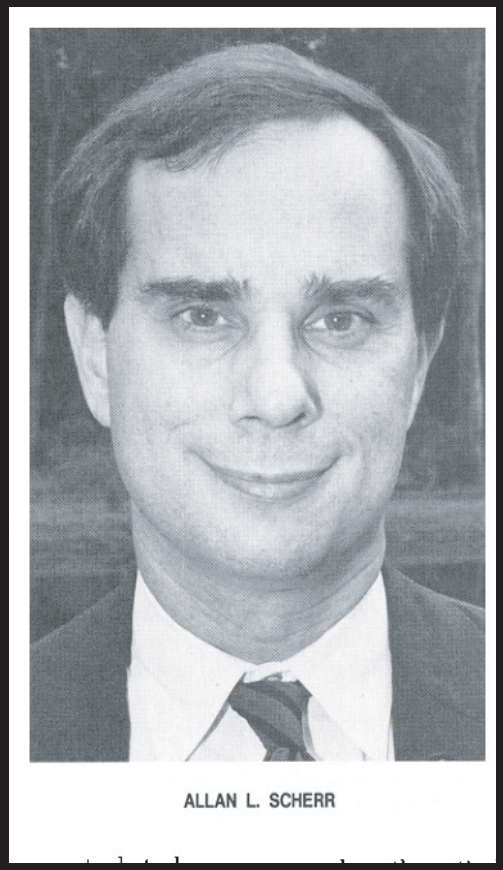


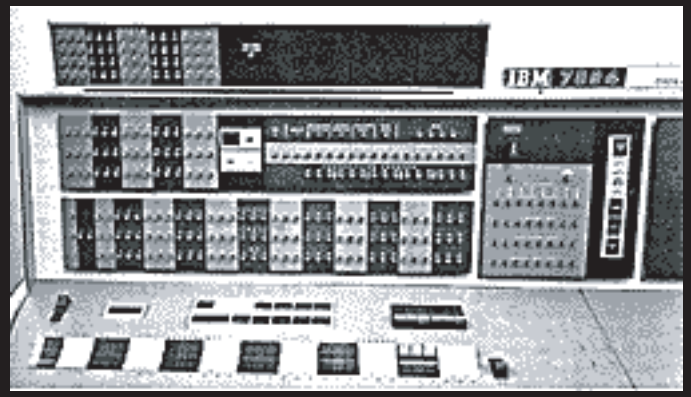
Mind-Reading Machine

- first implementation of Markov chain in electrical device.
- Philipp's applet: `http://www.ekphorie.de/engramm/mrm/`

Computer Performance Evaluation (1965)

- Allan L. Scherr's MIT thesis "An Analysis of Time-Shared Computer Systems" created the field of performance evaluation.





- CTSS allowed 300 users to interactively access computer. Scherr's goal: to characterize system usage.

Scherr's Thesis Work (Ph.D. in Elec.Eng.)

- created his own measurements of system performance.
- collected user data from MIT's CTSS system.
- compared this with simulation studies to predict these usage measures.
- sent draft of thesis off to his committee.

Scherr's Thesis Work (Ph.D. in Elec.Eng.)

- created his own measurements of system performance.
- collected user data from MIT's CTSS system.
- compared this with simulation studies to predict these usage measures.
- sent draft of thesis off to his committee.

Committee Member Comment

- one member complained "that there weren't enough mathematical formulas."

Scherr's Thesis Work (Ph.D. in Elec.Eng.)

- created his own measurements of system performance.
- collected user data from MIT's CTSS system.
- compared this with simulation studies to predict these usage measures.
- sent draft of thesis off to his committee.

Committee Member Comment

- one member complained "that there weren't enough mathematical formulas."

Scherr's Response

- He added an idea, CTMCs, from his recent O.R. course to produce a "very quick and dirty" mathematical analysis.

$$A = \begin{bmatrix} -\frac{n}{T} & \frac{n}{T} & 0 & 0 & 0 & \cdots & & & \\ \frac{1}{P} & -(\frac{1}{P}+\frac{n-1}{T}) & \frac{n-1}{T} & 0 & 0 & \cdots & & & \\ 0 & \frac{1}{P} & -(\frac{1}{P}+\frac{n-2}{T}) & \frac{n-2}{T} & 0 & \cdots & & & \\ 0 & 0 & & & & \cdots & & & \\ 0 & & & & & \cdots & & & \\ & \cdots & & & & & & & \\ & & \cdots & & \frac{1}{P} & -(\frac{1}{P}+\frac{3}{T}) & \frac{3}{T} & 0 & 0 \\ & & \cdots & & 0 & \frac{1}{P} & -(\frac{1}{P}+\frac{2}{T}) & \frac{2}{T} & 0 \\ & & \cdots & & 0 & 0 & \frac{1}{P} & -(\frac{1}{P}+\frac{1}{T}) & \frac{1}{T} \\ & & \cdots & & 0 & 0 & 0 & \frac{1}{P} & -\frac{1}{P} \end{bmatrix}$$

Let $\prod = \{\pi_0, \pi_1, \pi_2, \ldots, \pi_n\}$ be the vector of the steady-state probabilities of occupancy. Then,

$$\prod A = 0 \quad .$$

The equations for the π's are:

$$-\frac{n}{T}\pi_0 + \frac{1}{P}\pi_1 = 0$$

$$\cancel{\frac{n}{T}\pi_0} - \cancel{\frac{1}{P}\pi_1} - \frac{n-1}{T}\pi_1 + \frac{1}{P}\pi_2 = 0$$

etc.

yielding

$$\pi_1 = n\frac{P}{T}\pi_0$$

$$\pi_2 = n(n-1)\left(\frac{P}{T}\right)^2\pi_0$$

etc.

In general,

$$\pi_i = \frac{n!}{(n-i)!}\left(\frac{P}{T}\right)^i\pi_0 \quad .$$

Making use of the fact that $\sum_{i=0}^{n}\pi_i = 1$, the following equation results:

$$1 = \pi_0[1 + n\frac{P}{T} + n(n-1)\left(\frac{P}{T}\right)^2 + \ldots + \frac{n!}{(n-j)!}\left(\frac{P}{T}\right)^j + \ldots + n!\left(\frac{P}{T}\right)^n]$$

Letting $\frac{P}{T} = r$, and solving for π_0 :

$$\pi_0 = \frac{1}{\sum_{j=0}^{n}\frac{n!}{(n-j)!}r^j} \quad .$$

Thus

$$\pi_i = \frac{\frac{n!}{(n-i)!} r^i}{\sum_{j=0}^{n} \frac{n!}{(n-j)!} r^j} .$$

π_0 is the steady-state probability thatthe processor is idle.

Now, let W be equal to the mean length of time a user spends in the working part of the interaction (i.e., the mean response time). Let $\overline{Q}$ be the average number of users waiting for service. Then

$$\overline{Q} = n \frac{W}{W+T} = \sum_{i=0}^{n} i\pi_i$$

and

$$\overline{Q} = \frac{\sum_{i=0}^{n} \frac{in!}{(n-i)!} r^i}{\sum_{i=0}^{n} \frac{n!}{(n-i)!} r^i} = n \frac{W}{W+T} .$$

Solving for W ;

$$W = \frac{\sum_{i=0}^{n} \frac{ir^i}{(n-i)!}}{\sum_{i=0}^{n} \frac{r^i}{(n-i-1)!}} T$$

Dividing both sides by P and using the definition of r :

$$\frac{W}{P} = \frac{\sum_{i=0}^{n} \frac{ir^i}{(n-i)!}}{r\sum_{i=0}^{n} \frac{r^i}{(n-i-1)!}} \quad .$$

Expressed in terms of π_0 ,

$$\frac{W}{P} = \frac{n}{(1-\pi_0)} - \frac{1}{r} \qquad \text{or}$$

$$W = \frac{nP}{(1-\pi_0)} - T \quad .$$

It is interesting to note that the rate matrix and the resulting calculations would be the same if it were assumed that each program were run a finite quantum of time or if all programs were run to completion, i.e., batch processed! This is due to the fact that there is no swapping loss and that the time distributions are exponential. Use of other types of distribution functions would not, in general, yield the same results for any quantum size.

Hidden Markov Models (1989)

History

- In the late 1960s L. E. Baum discovered a method that, given a long sequence of observations, could identify the underlying model that generated the sequence.
- In the 1970s, based on Baum's work, James Baker created a speech recognition system that outperformed the current systems. Unlike most systems, his relied on no linguistic knowledge, instead only advanced statistics, to do pattern recognition of speech signals.
- Baker's approach intrigued Lee Neuwirth, the director of the Institute of Defense Analysis, who named the models "Hidden Markov Models."
- Lawrence R. Rabiner's tutorial in IEEE popularized HMMs.

Hidden Markov Models

HMMs vs MCs

- MCs require a priori computation or knowledge of transition probabilities. HMMs can be used to compute some hidden transition probabilities.

HMMs used in

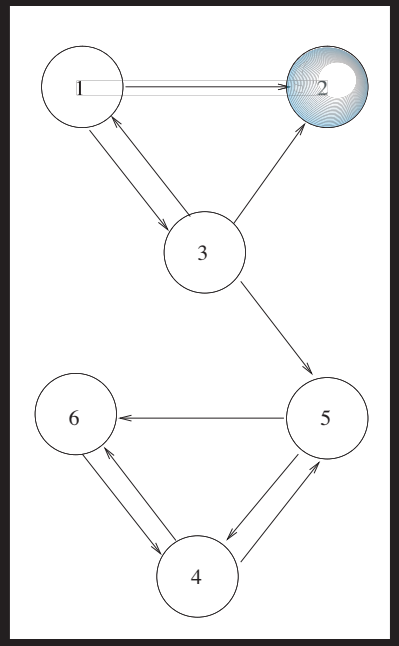
- pattern recognition
- sequence analysis of genes
- speech recognition
- handwriting recognition

What could be the next technological step for HMM?

• Perhaps a 3-dimensional radio pattern recognition of books – Markov would have like it!

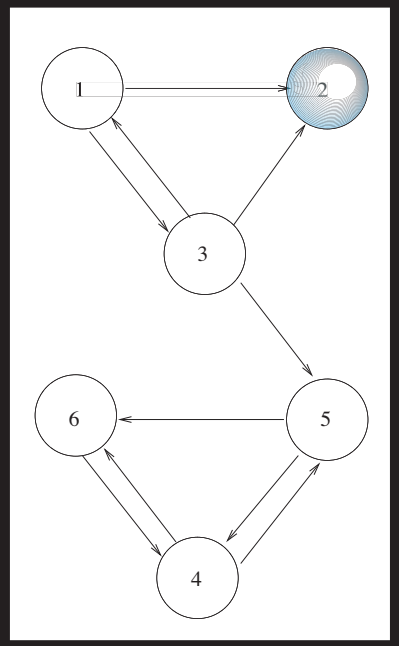
PageRank (1998)

Sergey Brin and Larry Page , then at Stanford, think of a web hyperlink as a recommendation.



$$
\mathbf{H} = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{array}
\begin{array}{c}
\begin{array}{cccccc} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 \end{array} \\
\begin{pmatrix}
0 & 1/2 & 1/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\end{array}
$$

PageRank Example



The Magic Eigenvector $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{G}$,

$$\mathbf{G} = \alpha \mathbf{H} + \alpha\, \mathbf{a}\, \mathbf{v}^T + (1-\alpha)\mathbf{E}$$

$$\boldsymbol{\pi}^T = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} .03721 & .05396 & .04151 & .3751 & .206 & .2862 \end{pmatrix} \end{array}$$

Global ranking of pages = $[\,4 \quad 6 \quad 5 \quad 2 \quad 3 \quad 1\,]$

Query-independent way of ranking relevant set

An Appropriate Order

- Chronological

Chronological Order

1. Eugeny Onegin
2. Information Theory
3. Computer Perf. Eval.
4. Hidden Markov Models
5. PageRank

Chronological Order

1. Eugeny Onegin
2. Information Theory
3. Computer Perf. Eval.
4. Hidden Markov Models
5. PageRank

vs.

1. PageRank
2. Hidden Markov Models
3. Computer Perf. Eval.
4. Information Theory
5. Eugeny Onegin

An Appropriate Order

- Chronological
- Alphabetical

Alphabetical Order

- by author or application?
- what about multiple authors? Brin and Page or Page and Brin?

An Appropriate Order

- Chronological
- Alphabetical
- Markov Number

Order by Markov Number

- Clearly, for Eugeny Onegin, Markov # = 0.
- What about others? Did Shannon co-author a paper with one of Markov's co-authors?

Order by Markov Number

- Clearly, for Eugeny Onegin, Markov # = 0.
- What about others? Did Shannon co-author a paper with one of Markov's co-authors?
- Did Markov actually have many co-authors?

  On the 5th of March 1921 A. A. Markov communicated that on account of the absence of footwear he is not able to attend meetings of the Academy. A few weeks later the KUBU (Committee for Improvement of the Existence of Scientists), meeting under the chairmenship of A. M. Gorky, fulfilled the prosaic request of the famous mathematician. Time, however, provided a colorful sequel, of sorts, to this. At the meeting of the physico-mathematical section of the Academy of Science on the 25th May, Andrei Andreevich announced: "Finally, I received footwear; not only, however, is it stupidly stitched together, it does not in essence accord with my measurements. Thus, as before, I cannot attend meetings of the Academy. I propose placing the footwear received by me in the Ethnographic Museum as an example of the material culture of the current time, to which end I am ready to sacrifice it."

An Appropriate Order

- Chronological
- Alphabetical
- Markov Number
- Impact Factor

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

— we used Google Scholar to find all the references for this paper?*

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

— we used Google Scholar to find all the references for this paper?*

* We did.

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

— we used Google Scholar to find all the references for this talk?*

— we used Google Image to find all the pictures in this talk?**

* We did.

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

— we used Google Scholar to find all the references for this talk?*

— we used Google Image to find all the pictures in this talk?**

* We did.

** We did.

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

— we used Google Scholar to find all the references for this talk?*
— we used Google Image to find all the pictures in this talk?**
— Google has become a noun, verb, participle, adj., proper noun (for pets)?***

* We did.
** We did.

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

— we used Google Scholar to find all the references for this talk?*
— we used Google Image to find all the pictures in this talk?**
— Google has become a noun, verb, participle, adj., proper noun (for pets)?***

* We did.
** We did.
*** It has.

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

— we used Google Scholar to find all the references for this talk?*
— we used Google Image to find all the pictures in this talk?**
— Google has become a noun, verb, participle, adj., proper noun (for pets)?***
— Amy bought her shoes through Froogle?****

* We did.
** We did.
*** It has.

Order by Impact Factor

How should we measure impact . . .

- As applications that create new fields, like information theory or computer performance evaluation?
- Or spread of influence of an application?

Should we consider PageRank the greatest application of Markov chains just because

— we used Google Scholar to find all the references for this talk?*
— we used Google Image to find all the pictures in this talk?**
— Google has become a noun, verb, participle, adj., proper noun (for pets)?***
— Amy bought her shoes through Froogle?****

* We did.
** We did.
*** It has.
**** She did.

There must be some way to rank these

(our thought process)

Hmmm... What are popular ranking systems?

Ahh-Haa! PageRank!

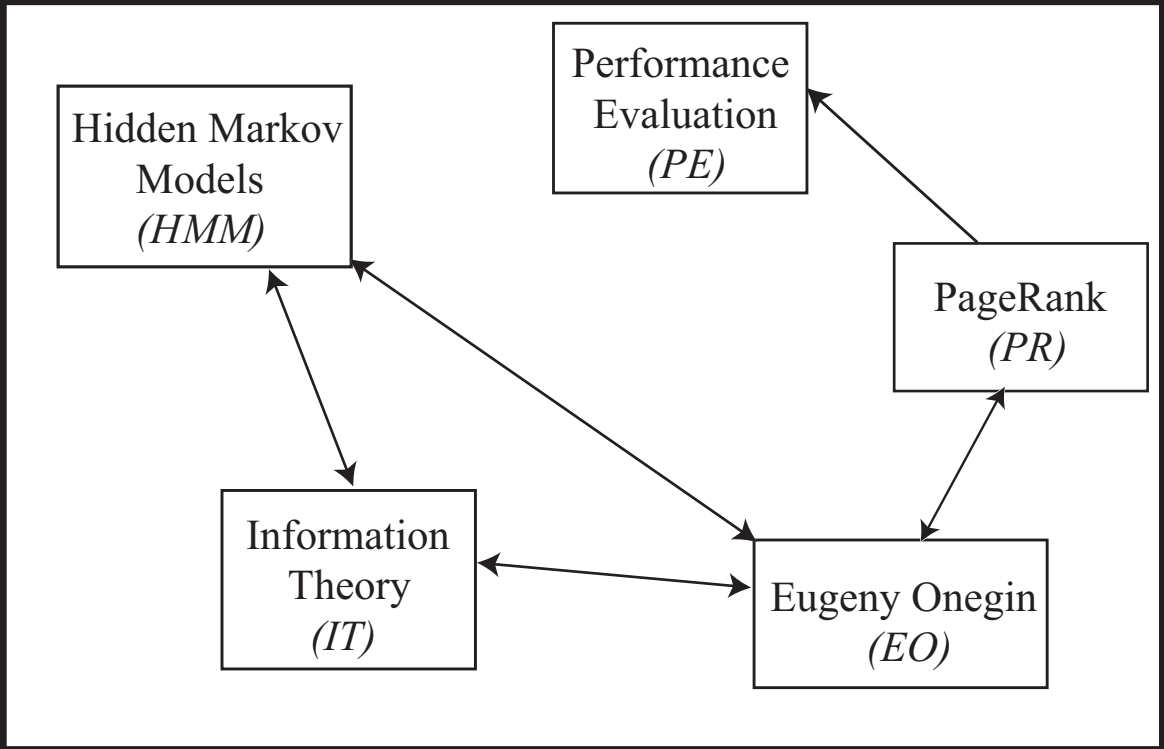
Eeeww. But dangerous. Is it safe to use PageRank to rank items, one of which is PageRank?

Ooohh. Use HITS instead!

An Appropriate Order

- Chronological
- Alphabetical
- Markov Number
- Impact Factor
- HITS ranking vector

Order by HITS



The Proper Order

5. Scherr's Computer Performance Evaluation

The Proper Order

4. Brin and Page's PageRank

5. Scherr's Computer Performance Evaluation

The Proper Order

3. Baum's Hidden Markov Models
4. Brin and Page's PageRank
5. Scherr's Computer Performance Evaluation

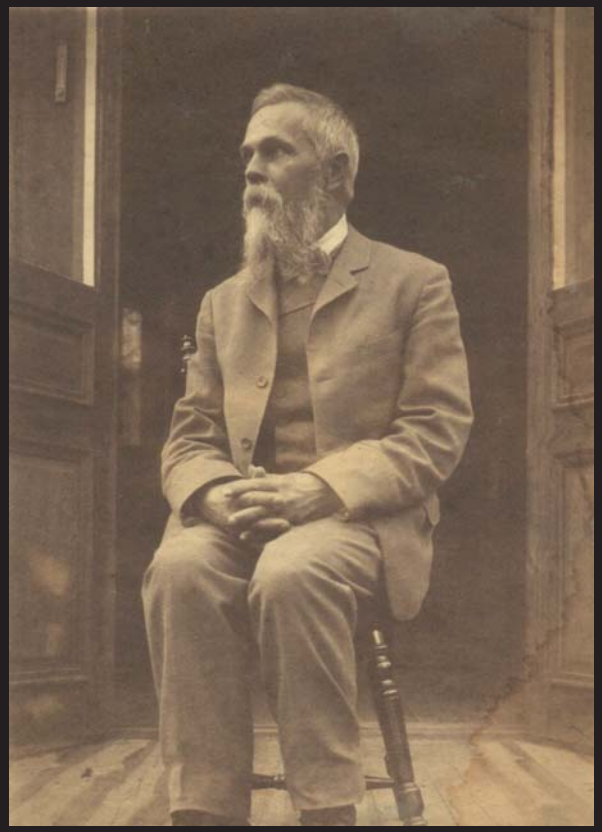
The Proper Order

2. Shannon's Information Theory
3. Baum's Hidden Markov Models
4. Brin and Page's PageRank
5. Scherr's Computer Performance Evaluation

The Proper Order

1. Markov's Eugeny Onegin
2. Shannon's Information Theory
3. Baum's Hidden Markov Models
4. Brin and Page's PageRank
5. Scherr's Computer Performance Evaluation

Happy Birthday, Markov!




A. LORENS
St PETERSBOURG