研究論文

An introductory research of application of Markov process to literary study

Chae-Heung Park
Dept. of Business Administration, Ewha Womans University

Abstract

This paper attempts to apply Markove chain theory to literary works. The objective of this paper is to show the trend of authors and his works by way of Markovian Models. In this introductory research, it is useful to make simple criteria as virtue and evil, good and bad, justice and injustice, positive and negative, man and woman, etc. Markovian transition matrixes which are derived by M.L.E. and A.U.E are almost same. In case of Hamlet, we are able to know the author of Hamlet or his works are inconsistent and fickle.

1. INTRODUCTION

In literary works, there are various alternative states which are exchanged in the course of theevent deployment. The Markovian assum ption greatly simplifies both the possible behavior of the process and the problem of specifying the process.

The assumption is : only the last state occupied by the process is relevant in determining its future behavior.

With this assumption,

$$P\{s(n+1)=j \mid s(n)=i, s(n-1)=k, \cdots, s(0)=m\}$$

Becomes

$$P\{s(n+1)=j \mid s(n)=i\}$$

Thus, the probability of making a transition to each state of the process depends only on the state presently occupied. Therefore, even if the history of the system before the last state occupied does influence future behavior, the Markovian assumption may still be satisfied by a change in the last structure.

Markov models represent the first outpost in the domain of dependent models that is powerful both in capturing the essence of

many dependent system observed in practice and producing the analytical and computational results necessary to gain insight into their behavior.

There are numerous contributors using consistent application of the fundamental principles. In 1920, Markov's interesting application of Markov chain is carried out for alternative vowels and consonants in written alphabets.「Cinlar, 1979」 Polson and Huizinga has studied the development of general statistical methods for absorbing markov-chain models of learning and of other psychological processes. 「Polson and Huizinga, 1974」 The output of a Markov process producing good and bad items in sequence is inspected for the bad items. A measure proposed of the work involved and the estimation required for this measurement is discussed by S. R. Broadbent. 「Broadbent, 1958」 Sign changes and behavior sequences are another interesting application of Markov process.「Lynch, 1979」, 「Cane, 1956」 Cane got and described the some experiments on birds, fish and insects which had long records of steady-state behavior and considered the relative merits of three simple models of behavior. The author has not read the analytical and computational results applying the Markov process in literature. But the exchange of states are numerous in literary works. Black and white in Conrad's heart of darkness, justice and injustice in Mark Twain's Huckleberry Finn are most prominent example of alternative states exchange. Sex, color, symbol (positive and negative) ideology, are categories dividing the states in the deployment. In this paper, the author is going to present the canonical model which is able to apply the Markov chains to the literature. The criteria dividing the states are ambiguous and may be different from the subjective opinion. Therefore, most simple and regular case which has two states is available to show the model.

In section 2, we analysis the states of the literary works and formulate the criteria dividing the states. And maximum likelihood estimation(M.L.E.) and almost unbiased estimation(A.U.E.) will be compared in two states Markov process in literature. In section 3, the trend of author and his works in literature is interpreted by the transition probability. Section 3, also contains discussion of possible areas of further study.

2. Formulation of criteria dividing categories and estimation

1. Formulation of criteria dividing categories.

As mentioned in introduction, the objective

of this paper is to show the trend of author and his works by way of Markovian Models. But it is very difficult to make a same criteria which is applied to all authors and their works in all ages and countries. Therefore, in this introductory research, it is useful to make such simple criteria as virtue and evil, good and bad, justice and injustice, positive and negative, man and woman, etc. Because of symbolism of language, such simple criteria will have adverse criticism in application of real problem. Francis Fergusson have provided as with clause to many of Hamlet's archetypal mysteries. So we first find out the case of Hamlet. He discloses point by point how the scenes in Shakespeare's play follow the same ritual pattern as those in Greek tragedy.

To appreciate how closely the moral norms in Shakespeare's play are related to those of ancient vegetation myths, we need only to note how after images of disease and corruption are used to symbolize the evil that has brighten Hamlet's Denmark. In this paper, the righteous characters stand proxy for virtue and bad characters stand proxy for evil to obtain simplification. Let the righteous characters (Hamlet, Horatio, Laertes, Fortinbras, Ophelia, Ghost of Hamlet's father) be 0, and the bad characters(otherwise) be 1, then we may consider that the process has such a sequence as .....0 1 0 1 0 1 0 0 .....

We now suppose that this sequence is a simple Markov chain i.e., the probability any character be 0 to 1, depends only on the previous character. That is to say, the probability which 1 is preceeded by 1 is $1-\beta$, and the probability which 1 is preceeded by 0 is α.

The transition matrix is therefore

$$\begin{bmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{bmatrix}$$

where $0<\alpha<1$ and $0<\beta<1$

This sequence is an ineducible Markov chain with two ergodic states.

2. Maximum Likelihood estimation (M. L. E.)

Let n_{00} : the number of transition from state 0 to state 0

n_{01} : the number of transition from state 0 to state 1

n_{10} : the number of transition from state 1 to state 0

n_{11} : the number of transition from state 1 to state 1

Then, the 〈table 1〉 show the cumulative number of transitions every transitions.

〈Table 1〉 Cumulative Number of Transitions

No.of Transitons	n_{00}	n_{01}	n_{10}	n_{11}
100	21	25	25	29
200	81	45	45	29
300	115	58	57	70
400	115	92	92	101
500	140	119	1250	121
600	188	140	141	131
700	190	185	185	140
800	191	224	224	161
900	201	256	256	187
1000	220	285	286	199
1100	259	315	315	201
1141	274	326	327	204

Let P_{00} : the probability of transition from state 0 to state 0

P_{01} : the probability of transition from state 0 to state 1

P_{10} : the probability of transition from state 1 to state 0

P_{11} : the probability of transition from state 1 to state 1

Then, the maximum likelihood estimators satisfy the following equations.「Bai, 1975」, 「Jang, 1981」

$$\hat{P}_{01} = n_{01} / (n_{00} + n_{01})$$
$$\hat{P}_{00} = 1 - P_{01}$$
$$\hat{P}_{10} = n_{10} / (n_{10}+n_{11})$$
$$\hat{P}_{11} = 1 - P_{10}$$

〈Table 2〉 show the probability of transition every 100 transitions by M.L.E.

〈Table 2〉 Probability of Transition by M.L.E.

No.of Transitons	$\hat{P}_{00}$	$\hat{P}_{01}$	$\hat{P}_{10}$	$\hat{P}_{11}$
100	0.457	0.543	0.463	0.537
200	0.643	0.357	0.392	0.608
300	0.665	0.335	0.449	0.556
400	0.556	0.444	0.477	0.523
500	0.541	0.459	0.498	0.502
600	0.573	0.427	0.518	0.482
700	0.507	0.493	0.569	0.431
800	0.460	0.540	0.582	0.418
900	0.440	0.560	0.578	0.422
1000	0.436	0.564	0.590	0.410
1100	0.451	0.549	0.610	0.390
1141	0.457	0.543	0.616	0.384

3. Almost unbiased estimation (A.U.E.)

Let N : Number of the Sample Size

$$S = \sum_{i=0}^{N-1} Xi$$

$$R = \sum_{i=0}^{N-1} Xi \cdot \sum_{i=0}^{N-1} Xi-1 \cdot Xi$$

Then, the almost unbiased estimators satisfy the following equations.

$$\sim P_{01} = (R-X_0-X_{N-1}) / (N-S-1)$$
$$\sim P_{01} = (R-1) / (S-1)$$

and

$$\sim P_{00} = 1 - P_{01}$$
$$\sim P_{11} = 1 - P_{10}$$

We analyze the almost unbiased estimators for the purpose of the comparison with the maximum likelihood estimators.

〈Table 3〉 show the factors required for the computation of A.U.E.

〈Table 3〉 Factors Required for the Computation of A.U.E.

N	S	R	X_0	X_{n-1}
101	56	27	1	0
201	76	47	1	0
301	129	59	1	1
401	195	94	1	1
501	245	124	1	1
601	276	145	1	1
701	329	189	1	0
801	390	229	1	1
901	450	263	1	1
1101	544	343	1	1
1142	558	354	1	0

and 〈Table 4〉 show the probability of transition by A.U.E.

〈Table 4〉 the Probability of Transition by A.U.E.

N	$\hat{P}_{00}$	$\hat{P}_{01}$	$\hat{P}_{10}$	$\hat{P}_{11}$
101	0.409	0.591	0.473	0.527
201	0.629	0.371	0.613	0.387
301	0.667	0.333	0.453	0.547
401	0.552	0.448	0.479	0.521
501	0.522	0.478	0.504	0.496
601	0.559	0.441	0.524	0.476
701	0.493	0.507	0.573	0.427
801	0.446	0.554	0.586	0.414
901	0.420	0.580	0.584	0.416
1001	0.425	0.575	0.595	0.405
1101	0.387	0.613	0.630	0.370
1142	0.395	0.605	0.634	0.366

From 〈Table 1〉 to 〈Table 4〉, we know that the more sample we have, the more maximum likelihood estimators equal to almost unbiased estimators.

4. the result of Hamlet case

Now, we have matrix of transitions probability. One is maximum likelihood estimation.

$$\text{with} \begin{bmatrix} 0.457 & 0.543 \\ 0.616 & 0.384 \end{bmatrix}$$

and the other is almost unbiased estimations.

$$\text{with} \begin{bmatrix} 0.395 & 0.605 \\ 0.634 & 0.366 \end{bmatrix}$$

Both matrices present that diagonal elements are less than the other elements. In other words, we are able to know that the author of hamlet or his works are inconsitent and fickle. Like this way, we'll discuss some other plays of Shakespear on section 4. And also, we'll find out what other people analyze the Shakespear's is right.

3. Application of some other play's

This part is touched with computer.

1. OTHELLO (1604, ed.)
2. TEMPEST (1611)
3. Julius Caesar (1599)
4. Merchant of Venicce (1596)

5. Macbeth (1606)

6. King Lear (1605-06)

Let's see the results of data.

1. Othello

M.L.E.	0.432	0.568
	0.556	0.444
A.U.E.	0.426	0.574
	0.560	0.440

2. Tempest

M.L.E.	0.793	0.207
	0.254	0.746
A.U.E.	0.793	0.207
	0.252	0.748

3. Julius Caesar

M.L.E.	0.559	0.441
	0.474	0.526
A.U.E.	0.532	0.468
	0.495	0.505

4. Merchant of Vanice

M.L.E.	0.662	0.338
	0.576	0.424
A.U.E.	0.662	0.338
	0.576	0.424

5. Macbeth

M.L.E.	0.708	0.292
	0.327	0.673
A.U.E.	0.708	0.292
	0.325	0.675

6. King Lear

M.L.E.	0.770	0.230
	0.589	0.441
A.U.E.	0.771	0.229
	0.584	0.416

** REFER TO DATA OUTPUT

4. CONCLUSION

By the tales of Shakespear, it indicate that Shakespear wrote his play with fairly distinguishing between righteous and bad character, and he want the happy end play. Section 4 prove this. That is the reason why the transition matrix of Macbeth, Tempest Julius Caesar, ed. al are diagonal is that Shakespear distinguish betweengood and bad surely, and the transition matrix of King Lear and Merchant of Venice indicate his happy end.

After the author knew the talensof Shakespear, the author wanted to analyze the Shakespear's play by mathematicca process. So, the author apply markov chain to analize literalic works but the author has not talent at

literature, and so is like to exist baised judgement in diffemtiating between categories. This paper is first attempt to apply Markov chain theory to analize literary works.

References

[1] Cinlar, An Introduction to Stochastic process, Prentice Hall, 1979.

[2] Peter G. Polson and David Huizinga, Statistical Methods for Absoring Markov-Chain Models for Learning : Estimation and Identification. Psychometrika. Vol. 39, No.1 March, 1974, pp. 3-22.

[3] S.R. Broadbent, The Inspection of A Markov Process, Journal of Royal Statistics Socieity, 1958, pp. 111-119.

[4] James Lynch, Sign Changes of the Difference of Convex Functions and Their Application to Large Deviation Rates. The Annals of Probability, 1979, Vol. 7, No.1, pp 96-108.

[5] Violet, R. Cane, Behabior Sequences as Semi-Markov Chains, Journal of Royal Statistics Socieity, 1956, pp. 36-58.

[6] D.S. Bai, "Efficient Estimation of Transition Probabilities in a Markov Chain," Annals of Statistics, 1975, pp. 1305-1317.

[7] J. S. Jang, Estimation of parameters of a Two State Markov Process by Interval Sampling. Unpublished M. A. Degree Thesis, 1981.