Poetry Generation Using Markov Chain

Adarsh Srivastava
Machine Learning Intern
AI Technology and System
adarsh.srivastava765@gmail.com
www.ai-techsystems.com

***Abstract*— This paper is about generating the text automatically or more precisely we can say generating the poem automatically. The poetry is the thing where a poet writes his beautiful thoughts in the form of text which have deep meaning and with having rhythm. It is a creative work where the poet have to think a lot for creating a creative poem. Now a days, we are focusing on generating the creative text and object.**

Keyword- Markov Chain, Text Generation, Natural Language Processing

INTRODUCTION

In the last years we have seen several attempts to generate creative objects automatically, with the help of computer programs and we have started to accept the computer has an artist. The creation of visual art, the composition of musical pieces or the generation of creative text are just some of the fields that actual creative systems deal with.

Now a days, there is a huge amount of data generating on daily basis. It is present in the form of audio, video, text, time series, sensor data, etc. A special thing about this type of data is that if two events are occurring in a particular time frame, the occurrence of event A before event B is an entirely different scenario as compared to the occurrence of event A after event B.

The generation of natural language is a well-established and promising sub-field of artificial intelligence and computational linguistics. However, in conventional machine learning problems, it hardly matters whether a particular data point was recorded before the other. This consideration gives our sequence prediction problems a different solving approach. The main goal is to develop computer programs that can produce text that can be understood by humans. Among the types of text generated automatically we can find, for example, biographies. Attempts to the automatic generation of creative text include systems capable of generating story, songs, lyrics, raps, jokes and poetry as well.

This paper is focused on the generation of the last referred text genre**.** Text, a stream of characters lined up one after another, is a difficult thing to crack. This is because when handling text, a model may be trained to make very accurate predictions using the sequences that have occurred previously, but one wrong prediction has the potential to make the entire sentence meaningless. However, in case of a numerical sequence prediction problem, even if a prediction goes entirely south, it could still be considered a valid prediction (maybe with a high bias). But, it would not strike the eye.

The automatic generation of poetry involves several levels of language like phonetics, lexical choice, syntax and semantics and usually demands a considerable amount of input knowledge. It is a complex task and though an interesting topic for artificial intelligence research.

In this paper, we stated a text automatic generation of poetry based on the markov chain model. It can automatically generate fluent text.

RELATED WORK

The poetry generation is done by using the poetry foundation data which consists of many poetries with having the poet and title details. In the early stage, in order o ensure that the generated text is consistent with the training sample in the probability distribution of the words.

Markov chain model is very suitable for modeling natural text, in recent years, a large number of works using the Markov chain model for automatic generation of text have appeared. Most of these works use the Markov chain model to calculate the number of common occurrences of each word in the training set and obtain the transition probability. Then the transition probability can be used to predict the words. This kind of method greatly improved the quality of the generated texts compared to the previous methods.

The Markov chain model is used to ensure that the automatic text generation process conforms to the statistical language distribution of the training samples.

APPROACH

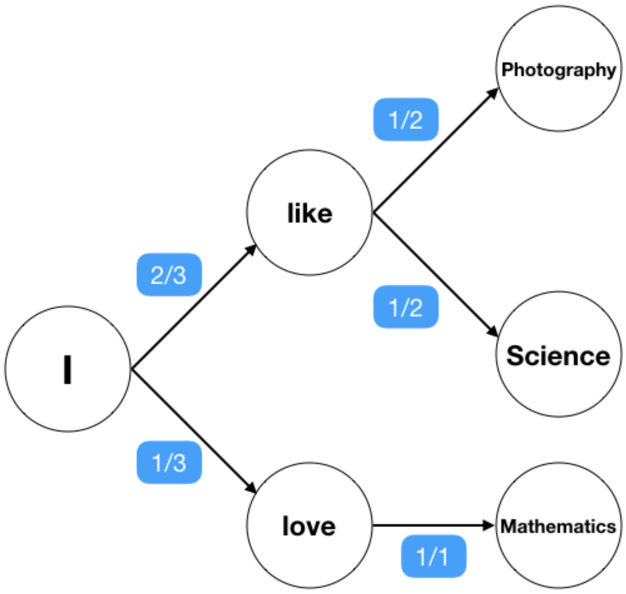
3.1 Automatic Text Generation Based on Markov Chain Model

Markov chains are an important mathematical tool in stochastic processes. The underlying idea is the Markov Property, in order words, that some predictions about stochastic processes can be simplified by viewing the future as independent of the past, given the present state of the process. This is used to simplify predictions about the future state of a stochastic process.

The field of statistical natural language processing, they usually use statistical language model to model a sentence. A language model is a probability distribution over sequences of words, it can be expressed by the following formula:

$$p(S) = p(w_1, w_2, w_3, \dots, w_n) \\ = p(w_1)p(w_2|w_1) \dots p(w_n|w_1, w_2, \dots, w_{n-1}) \quad (1)$$

where S denotes the whole sentence with a length of n and wi denotes the i-th word in it. $p(S)$ assigns the probability to the whole sequence. It is actually composed of the product of n conditional probabilities, each of the conditional probability calculates the probability distribution of the n-th word when the first $n-1$ words are given, that is $p(wn|w1, w2, \dots, wn-1)$. Therefore, in order to automatically generate high quality texts, we need to obtain a good estimate of the statistical language model of the training sample set. The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.



State transition diagram for our sample data

In probability theory, a Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event only depends on the state attained in the previous event. The Markov chain model is suitable for modeling time series signals. For instance, suppose there is a value space $\chi = \{x1, x2, x3, \dots, xm\}$,and $Q = \{q1, q2, q3, \dots, qn\}$ is a stochastic variable sequence, whose values are sampled from χ. For the convenience of the following description, we will record the value of t-th state as $x\ t$, that is $qt = x\ t$, $x\ t \in \chi$. If we think that the value of the state at each moment in the sequence is related to the state of all previous moments, that is $p(qt\ |q1, q2, \dots, qt-1)$, then the Markov chain model can be expressed as follows:

$$P(q_t = x^t) \\ = f\left(P(q_{t-1} = x^{t-1}), P(q_{t-2} = x^{t-2}), \dots, P(q_1 = x^1)\right), \\ \text{s.t.} \sum_{i=1}^{m} P(q_t = x_i) = 1, \forall x_i \in \chi \quad (2)$$

where f is the probability transfer function. Then the probability of the whole sequence can be expressed as follows:

$$p(Q) = p(q_1, q_2, q_3, \dots, q_n) \\ = P(q_1 = x^1)P(q_2 = x^2) \dots P(q_n = x^n) \\ = p(q_1)p(q_2|q_1) \dots p(q_n|q_{n-1}, q_{n-2}, \dots, q_1) \quad (3)$$

Compare formula (3) with formula (1), we find that if we consider the signal xi at each time point in formula (3) as the ith word in the sentence, it can exactly represent the conditional probability distribution of each word in the text, which is $p(wn|w1, w2, \dots, wn-1)$, and then it can perfectly model the statistical language model of the text. It is because of this commonality that the Markov chain model is very suitable for modeling text and is widely welcomed in the field of natural language processing, especially in the field of automatic text generation. Generally, in actual situations, the influence of the signal at each moment in the sequence signal on the subsequent signal is limited, that is, there exists a influence domain, and beyond the influence domain, it will not continue to affect the subsequent time signal. Therefore, we assume that for a time-series signal, the value of each time signal is only affected by the first few finite moments. If the value of the signal at each moment is only affected by the signals of the previous m moments, we call it the m-order Markov model and can be expressed as follows:

$$P(Q_t = x^t | Q_{t-1} = x^{t-1}, Q_{t-2} = x^{t-2}, \dots, Q_1 = x^1) \\ = P(Q_t = x^t | Q_{t-1} = x^{t-1}, Q_{t-2} = x^{t-2}, \dots, Q_{t-m} = x^{t-m}), \\ \text{s.t.} \quad n > t > m \quad (4)$$

When we use the Markov chain model for automatic text generation, we actually hope to use the Markov chain model to obtain a good statistical language model estimate through learning on a large number of text sets. For a big training corpus which contains multiple sentences, we first build a big dictionary D that contains all the words appeared in the training set, that is

$$D = \{word_{D_1}, word_{D_2}, word_{D_3}, \dots, word_{D_N}\}$$

where $wordDi$ indicates the i-th word in the dictionary D and N is the number of the word. Dictionary D corresponds to the value space χ described above. As we have mentioned before, each sentence S can be regarded as a sequential signal and the i-th word in S can be viewed as the signal at the time point i, that is

$$S = \{wordS1, wordS2, wordS3, \dots, wordSL\}, \\ \text{s.t.} \ \forall\ wordSi \in D$$

where $wordSi$ indicates the i-th word in sentence S and L is the length of it. In the automatic text generation process, we need to calculate the transition probability of each word. For the Markov chain model, according to the big number theorem, we usually use the frequency of each phrase in the data set to approximate the probability. For example, for a second-order Markov chain model, the calculation formula is as follows:

$$\begin{aligned} & p(word_{S_n} = word_{D_i} | word_{S_{n-2}}, word_{S_{n-1}}) \\ & \approx \frac{count(word_{S_{n-2}}, word_{S_{n-1}}, word_{D_i})}{count(word_{S_{n-2}}, word_{S_{n-1}})}, \\ & \text{s.t.} \sum_{i=1}^{N} p(word_{S_n} = word_{D_i} | word_{S_{n-2}}, word_{S_{n-1}}) = 1, \end{aligned} \tag{6}$$

where $count(wordSn{-}2\ ,\ wordSn{-}1\ ,\ wordDi\ )$ is the number of occurrences of this phrase $\{wordSn{-}2\ ,\ wordSn{-}1\ ,\ wordDi\ \}$ in the training set. If we don't need to embed information but just generate natural text, we usually choose the word with the highest probability as the output at each iteration.

GOAL

For the purpose of his report, Manurung defines that poetic text must hold all the following three properties:

• Meaningfulness: convey a conceptual message, which is meaningful under some interpretation.

• Grammaticality: obey linguistic conventions prescribed by a given grammar and lexicon.

• Poeticness: exhibit poetic features.

He describes a taxonomy for poetry generation systems, based on the goals they try to achieve and how their output texts embody the previously referred properties:

• Word salad: systems that simply concatenate random words together, without following any grammatical rules (none of the properties are embodied).

• Template and grammar-based: words are selected from a lexicon in order to fill gaps in sentence templates (property of grammaticality).

• Form-aware: the choice of words follows a pre-defined text form, like the haiku1 or the sonnet.

• Poetry generation: the generated text embodies all the three properties of grammaticality, poeticness and meaningfulness.

CONCLUSION

As we can notice, Markov models do provide decent results. Hence, Markov models should not be completely written off. It is advisable to try Markov models before jumping into much complex models such as LSTMs. It would be much more interesting to see how the combination of Markov models and LSTM would play out together.

REFERENCE

[1] K. Binsted. Machine humour: An implemented model of puns. PhD thesis, University of Edinburgh, Scotland, 1996.

[2] L. Bourbeau, D. Carcagno, E. Goldberg, R. Kittredge, and A. Polguere. Bilingual generation of weather forecasts in an ` operations environment. In Proceedings of the 13th conference on Computational linguistics, pages 318–320, Morristown, NJ, USA, 1990. Association for Computational Linguistics.

[3] S. Bringsjord and D. A. Ferrucci. Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling Machine. Lawrence Erlbaum Associates, Hillsdale, NJ., 1999.

[4] B. D´ıaz-Agudo, P. Gervas, and P. A. Gonz ´ alez-Calero. Po- ´ etry generation in colibri. In ECCBR '02: Proceedings of the 6th European Conference on Advances in Case-Based Reasoning, pages 73–102, London, UK, 2002. Springer-Verlag.

[5] P. Gervas. Wasp: Evaluation of different strategies for the ´ automatic generation of spanish verse. In Wiggins, G. (Ed.). Proceedings of the AISB00 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science, Birmingham, UK, 2000.

[6] Yan Zhongliang,Jin Shuyu , Huang Yongfeng, Zhang Yujin and Li Hui3 , Automatically Generate Steganographic Text Based on Markov Model and Huffman Coding

[7]] T. Konstantopoulos, "Introductory lecture notes on MARKOV CHAINS AND RANDOM WALKS", Autumn 2009,[online], http://www2.math.uu.se/~takis%20/L/McRw/mcrw.pdf

[8] Poetry Foundation., https://www.poetryfoundation.org/