A SURVEY OF MARKOV CHAIN MODELS IN LINGUISTICS APPLICATIONS

Fawaz S. Al-Anziand Dia AbuZeina

Department of Computer Engineering, Kuwait University, Kuwait City, Kuwait

fawaz.alanzi@ku.edu.kw,abuzeina@ku.edu.kw

ABSTRACT

Markov chain theory isan important tool in applied probability that is quite useful in modeling real-world computing applications.For a long time, rresearchers have used Markov chains for data modeling in a wide range of applications that belong to different fields such as computational linguists, image processing, communications,bioinformatics, finance systems, etc. This paper explores the Markov chain theory and its extension hidden Markov models (HMM) in natural language processing (NLP) applications. This paper also presents some aspects related to Markov chains and HMM such as creating transition matrices, calculating data sequence probabilities, and extracting the hidden states.

KEYWORDS

Markov chains,Hidden Markov Models, computational linguistics, pattern recognition, statistical

1. INTRODUCTION

Markov chains theory is increasingly being adopted in real-world computing applications since it provides a convenient way for modeling temporal, time-series data. At each clock tick, the system moves into a new state that can be the same as the previous one. A Markov chain model is amathematical tool that capture the patterns dependencies in pattern recognition systems. For this reason, Markov chain theory is appropriate in natural langue processing (NLP) where it is naturally characterized by dependencies between patterns such as characters or words.

Markov chains are directed graphs (a graphical model) that are generally used with relatively long data sequences for data-mining tasks. Such tasks include prediction, classification, clustering, pattern discovery, software testing, multimedia analysis, networks, etc. Reference [1] indicated that there are two reasons of Markov chains popularity; very rich in mathematical structure and work well in practice for several important applications. Hidden Markov models (HMM) is an extension of Markov chains that used to find the hidden system's states based on the observations.

In order to facilitate the research in this direction, this paper provides a survey of this so popular data modeling technique. However, because of the wide range of the research domains that use this technique. We specifically focuson the linguistics related applications. Reference [2] list some domains that utilize Markov chains theory which include: physics, chemistry, testing, speech recognition, information sciences, queueing theory, internet applications, statistics, economics and finance, social sciences, mathematical biology, genetics, games, music, baseball,

David C. Wyld et al. (Eds) : ICAITA, CDKP, CMC, SOFT, SAI - 2016
pp. 53– 62, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61305

Markov text generators, bioinformatics. Reference [3] lists the five greatest applications of Markov chains that include Scherr's application to computer performance evaluation, Brin and Page's application to PageRank and Web Search, Baum's application to HMM, Shannon's application to information theory, and Markov's application to Eugeny Onegin.

This paper is organized as follows. The next section presents a background of Markov chains theory. Section 3 highlights the main concepts of HMM followed by a literature review of Markov chains and HMM in section 4. Finally, we conclude in section 5.

2. MARKOV CHAINS

Markov chains are quite useful in modeling computational linguistics. A Markov chain is a memorylessstochastic model that describes the behaviour of an integer-valued random process. The behaviour is the simple form of dependency in which the next state (or event) depends only on the current state. According to [4], a random process is said to be Markov if the future of the process, given the present, is independent of the past. To describe the transitions between states, a transition diagram is used to describe the model and the probabilities of going from one state to another. For example, Figure 1 shows a Markov chain diagram with three states (Easy, Ok, and Hard) that belong to exam cases(i.e. states). In the figure, each arc represents the probability value for transition from one state to another.

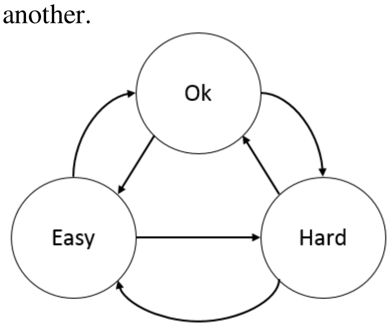

Figure 1. A Simple Markov chain with three states

The Markov chain diagrams are generally represented using state transition matricesthat denote the transition probabilities from one state to another. Hence, a state transition matrix is created using the entire states in the system. For example, if a particular textual application has a training datathat contains N states (e.g. the size of lexicon), then the state transition matrix is described by a matrix A= {aij} of size N*N. In matrix A, the element aij denote the transition probability from a state i to a state j. Table 1 shows how the state transition matrix used to characterize the Markov diagram shown in Figure 1. That is, the matrix carries the state transitions probabilities between the involved states(Easy, Ok, and Hard). For illustration, the P(E|H) denote to the probability of the next exam to be Easy given that the previous exam was Hard.

Table 1. A state transition matrix of three states

State		**Next Exam**		
		Easy (E)	**Ok (O)**	**Hard (H)**
Previous Exam	**Easy (E)**	P(E|E)	P(O|E)	P(H|E)
	Ok (O)	P(E|O)	P(O|O)	P(H|O)
	Hard (H)	P(E|H)	P(O|H)	P(H|H)

In Table 1, the sum of the probability values at each row is 1 as the the sum of the probabilities coming out of each node should be 1. Hence,P(E|E)+P(O|E)+P(P(H|E) equal 1. Markov chain is a worthy topic that has many details. For examples, it contains discrete-time, continuous-time,

time-reversed, reversible, and irreducible Markov chains. The case shown in Figure 1 is irreducible case, also called ergodic, where it is possible to go from every state to every state.

To illustrate a simple Markov chain data model, a small data set contains two English sentences used to create a transition matrix based on the neighbouringcharacters sequences. The sentences are inspirational English quotes picked from [5]:

(1) Power perceived is power achieved. (2) If you come to a fork in the road, take it.

Figure 2 shows the transition matrix of these quotes by counting the total number of occurrences of the adjacent two character sequences. It is a 19 × 19 matrix where the value 19 is the total number of unique characters appeared in thesentences (i.e the two quotes). In this example, creating transition matrix is case insensitive where D is same as d, as an example. In addition, a space between two words discarded and not considered in the transition matrix. Figure 2 also shows that the maximum number in the matrix's entries is 3 (a highlighted underlined value) which means that moving from character e to r (e➔r) is the most frequently sequence appeared in this small corpus. The words that contains this sequence are :{ Power (two times) andperceived}.

	a	c	d	e	f	h	i	k	m	n	o	p	r	s	t	u	v	w	y
a	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
c	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e	0	0	2	0	0	0	1	0	0	0	0	0	**3**	0	0	0	1	0	0
f	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
h	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
i	0	0	0	1	1	0	0	0	0	1	0	0	0	1	1	0	1	0	0
k	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
o	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	2	0
p	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
r	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
u	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
v	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
w	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
y	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Figure 2. A transition matrix of two characters sequences

Based on the information provided in the transition matrix shown in Figure 2. It is possible to answer some questions related to the given data collection. Among inquires, what is the total number of the two characterssequences appeared in the given data set?What are the two characters sequences that did not appear in the data collection?What is the least frequently two characters sequences in the data set? Accordingly, Markov chains are used as prediction systems such as weather forecasting. Therefore, it is possible to predict the tomorrow's weather according to the today's weather. For example, if we have two states (Sunny, Rainy), and the requirement is to find the probability P(Sunny|Rainy), Markov chains make it possible based on the information provided in the probability transition matrix. Another example of the using Markov chains is banking industry. A big portfolio of banks is based on loans. Therefore, Markov chainsare used to classify loans to different states such as Good, Risky, and Bad loans.

For simplicity, the information presented in Figure 2 shows the transition matrixbased on total number of occurrences. Figure 3 shows the same information but using probabilities instead of the number of occurrences. That is, it contains the probability of moving from one character to another. As previously indicated, the sum of entries at each row is equal 1. In Figure 3, any matrix entry that has 0 means that there is no transition at that case. Similarly, if the matrix entry is 1, it means that there is only one possible output of that state. For example, the character “o” comes after “y”, and this is the only possible arc of the state “y”.

	a	c	d	e	f	h	i	k	m	n	o	p	r	s	t	u	v	w	y
a	0	0.33	0.33	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0
c	0	0	0	0.33	0	0.33	0	0	0	0	0.33	0	0	0	0	0	0	0	0
d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e	0	0	0.29	0	0	0	0.14	0	0	0	0	0	0.43	0	0	0	0.14	0	0
f	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
h	0	0	0	0.5	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0
i	0	0	0	0.17	0.17	0	0	0	0	0.17	0	0	0	0.17	0.17	0	0.17	0	0
k	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
o	0.17	0	0	0	0	0	0	0	0.17	0	0	0	0.17	0	0	0.17	0	0.33	0
p	0	0	0	0.33	0	0	0	0	0	0	0.67	0	0	0	0	0	0	0	0
r	0	0.33	0	0	0	0	0	0.33	0	0	0.33	0	0	0	0	0	0	0	0
s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t	0.33	0	0	0	0	0.33	0	0	0	0	0.33	0	0	0	0	0	0	0	0
u	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
v	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
w	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
y	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Figure 3. A probability transition matrix of two characters sequences

3. HIDDEN MARKOV MODELS

Hidden Markov models (HMM) is an extension to Markov chains models as both used fortemporal data modeling. However, the difference is that the states in Markov chain models are directly observed while they are hidden in the case of HMM.We explain the concept of HMM based on Figure 1 that shows athreeexam’s states Markov diagram. As a very simple example, supposed that a student’s parents want to know the levels (i.e the difficulty) of theirson’s exams, naturally, it is possible to recognize the exam as Easy or Ok if the son feels Fine. Similarly, it is possible to recognize the exam as Hard if the son looks Scared. From the parents’ point of view, the required states (i.e. Easy, Ok, or Hard) are hidden. However, they directly observe the student’s reaction or feeling. Hence, the parents might use the observed reactionas an indication to know the hidden states. HMM is described using three matrices: the initial probability matrix, the observation probability matrix, and the state transition matrix. Figure 4 shows a HMM diagram that shows the states and the observations. In the figure, each arc represents the probability between the states and between the states and the observations.

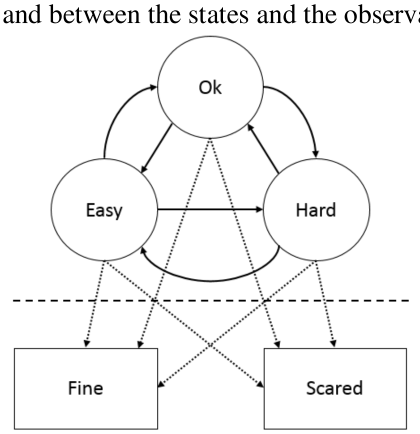

Figure 4. A HMM diagram with the transition and the observation arcs

Based on the information provided in the matrices, either Baum-Welch (also called any path) or Viterbi (also called best path) algorithms used to find the probability scores during recognition phase. Figure 5 shows the trellis diagram forexam states HMM. While Baum-Welch algorithm is used to compute the recognitin probability of a sequence, Viterbi is used to find the best-state sequence associated with the given observtatin, this procoss is also known as back-trakcing. Hence, after computing the observations sequence probability and finding the maximumprobability (supposed the star in Figure 5), the Viterbi algorithm leads the process back to identify the states (sources) from which the observations sequence have been emitted. In Figure 5, the maximum probalities supposed to be achieved at the states shown using the dotted lines: Ok, Easy, Hard, respectively.

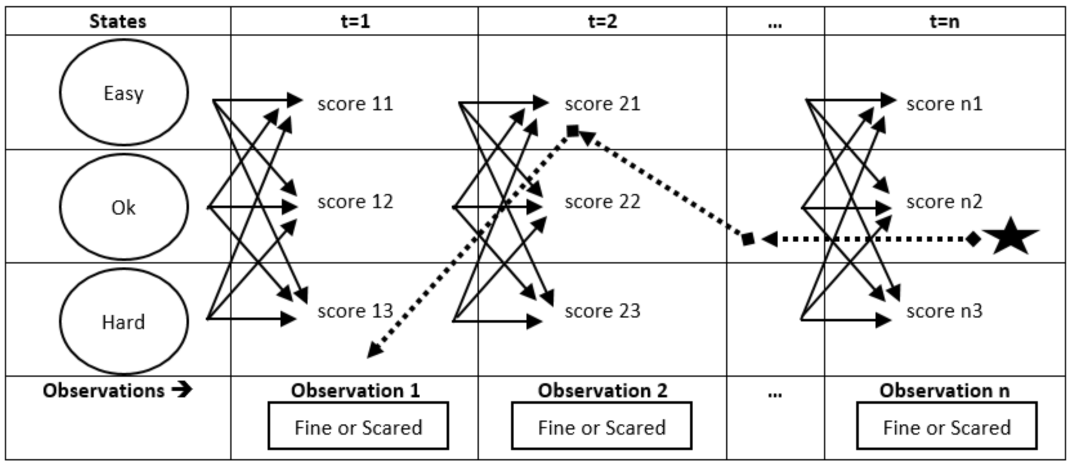

Figure 5.Trellis diagram of three states HMM

4. LINGUISTIC APPLICATIONS

In the literature, there are quite many works on modelingcontent dependencies for linguistics applications. Markov chain models and HMMs are of great interest to linguistic scholar who primarily work on data sequences. Even though this study focuses on linguistic applications, however, Markov chains used to model a variety of phenomena in different fields. The following are some of studies employed Markov chains. We intentionally ignored the references as the literature hastoo many studies employed Markov chains:

recasting,
iformatics,
er quality,
nagement,
, business

The following two subsections include some of the linguistic studies that utilized Markov chain theory. Linguistic applications topics mainly include (but not limited) speech recognition,speech emotion recognition,part-of-speech tagging, machine translation, text classification, text summarization, optical character recognition (OCR), named entity recognition, question answering,authorship attribution, etc.For the reader who interested in NLP, Reference [6] is a good reference as it demonstrates a thorough study of NLP (Almost) from Scratch.

4.1. Markov chains based research

The literature has a large number of studies that employ Markov chains for NLP applications. The following are some linguistic related applications. Reference [7] proposed a word-dividing algorithm based on statistical language models and Markov chain theory for Chinese speech processing. Reference [8] presented a semantic indexing Markov chains algorithm that uses both audio and visual information for event detection in soccer programs. Reference [9] investigated the use of Markov Chains and sequence kernels for the task of authorship attribution. Reference [10] implemented a probabilistic framework for support vector machine (SVM) that allows for automatic tuning of the penalty coefficient parameters and the kernel parameters via Markov chain for web searching via text categorization. Reference [11] demonstrated an automatic video annotation using multimodal Dirichlet process mixture model by collecting samples from the corresponding Markov chain. Reference [12] used a linguistic steganography detection method based on Markov chain models. Reference [13] showed how probabilistic Markov chain models can be used to detect topical structure in large text corpora.

Reference [14] proposed a method of recognizing location names from Chinese texts based on Max-Margin Markov Network. Reference [15] utilized Markov chain and statistical language models in a linguistic steganography detection algorithm. Reference [16] proposed a Markov chain based algorithm for Chinese word segmentation. Reference [17] presented two new textual feature selection methods based on Markov chains rank aggregation techniques. Reference [18] proposed a Markov chain model for radical descriptors in Arabic Text Mining. Reference [19] presented statistical Markov chain models for the distributions of words in text lines. Reference [20] proposed a method for handwritten Chinese/Japanese text (character string) recognition based on semi-Markov conditional random fields (semi-CRFs). Reference [21] presented a Markov chain method to find authorship attribution on relational data between function words. Reference [22] utilized a probabilistic Markov chain model to infer the location of Twitter users. Reference [23] proposed a Markov chain based technique to determine the number of clusters of a corpus of short-text documents.Reference [24] proposed a Markov chain based method for digital document authentication. Reference [25] used Markov chain for authorship attribution in Arabic poetry.

4.2. Hidden Markov modelsbased research

Linguistic HMM based research has been for long an active research area due to the rapid development in NLP applications. The literature has many studies as follows. Reference [26] proposed to extract acronyms and their meaning from unstructured text as a stochastic process using HMM. Reference [27] proposed a morphological segmentation method with HMM method for Mongolian.Reference [28] employed HMM for Arabic handwritten word recognition based on HMM. Reference [29] presented a scheme for off-line recognition of large-set handwritten characters in the framework of the first-order HMMs. Reference [30] proposed the use of hybrid HMM/Artificial Neural Network (ANN) models for recognizing unconstrained offline handwritten texts. Reference [31] used HMMs for recognizing Farsi handwritten words.

Reference [32] describes recent advances in HMM based OCR for machine-printed Arabic documents.Reference [33] proposed a HMMbased method fornamed entity recognition. Reference [34] combined text classification and HMM techniques for structuring randomized clinical trial abstracts. Reference [35] employed HMM for medical text classification. Reference [36] propose text (sequences of pages) categorization architecture based on HMM.Reference [37] described a model for machine translation based on first-order HMM.Reference [38] introduced speech emotion recognition by use of HMM.Reference [39] presented a HMMbased method for speech emotion recognition. Reference [40] discussed the role of HMM in speech recognition. Reference [41] indicated that almost all present day large vocabulary continuous speech recognition (LVCSR) systems based on HMMs.Reference [42] presented a text summarization method based on HMM. Reference [43] presented a method for summarizing speech documents using HMM. Reference [44] used HMM for part-of-speech tagging task. Reference [45] presented a second-order approximation of HMM for part-of-speech tagging task.

5. CONCLUSIONS

This work demonstrates the potential and the size of Markov chains research. The study reveals that the Markov chain and HMM is of high important for linguistic applications. Similarly, Markov chains are also widely used in many other applications. For future work, it worthy to explore the power of Markov chain in new linguistic and scientific directions with more details.

ACKNOWLEDGEMENTS

This work is supported by Kuwait Foundation of Advancement of Science (KFAS), Research Grant Number P11418EO01 and Kuwait University Research Administration Research Project Number EO06/12.

REFERENCES

[1] Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." Proceedings of the IEEE 77.2 (1989): 257-286.

[2] Markov_chain. (2016, August). Retrieved from https://en.wikipedia.org/wiki/Markov_chain

[3] Von Hilgers, Philipp, and Amy N. Langville. "The five greatest applications of Markov Chains." Proceedings of the Markov Anniversary Meeting, Boston Press, Boston, MA. 2006.

[4] Leon-Garcia, Alberto, and Alberto. Leon-Garcia. Probability, statistics, and random processes for electrical engineering. Upper Saddle River, NJ: Pearson/Prentice Hall, 2008.

[5] California Indian Education. (2016, August). Retrieved from http://www.californiaindianeducation.org/inspire/world/

[6] Collobert, Ronan, et al. "Natural language processing (almost) from scratch."Journal of Machine Learning Research 12.Aug (2011): 2493-2537.

[7] Bin, Tian, et al. "A Chinese word dividing algorithm based on statistical language models." Signal Processing, 1996., 3rd International Conference on. Vol. 1. IEEE, 1996.

[8] Leonardi, Riccardo, PierangeloMigliorati, and Maria Prandini. "Semantic indexing of soccer audio-visual sequences: a multimodal approach based on controlled Markov chains." IEEE Transactions on Circuits and Systems for Video Technology 14.5 (2004): 634-643.

[9] Sanderson, Conrad, and Simon Guenter. "On authorship attribution via Markov chains and sequence kernels." 18th International Conference on Pattern Recognition (ICPR'06). Vol. 3. IEEE, 2006.

[10] Lim, Bresley Pin Cheong, et al. "Web search with text categorization using probabilistic framework of SVM." 2006 IEEE International Conference on Systems, Man and Cybernetics. Vol. 4. IEEE, 2006.

[11] Velivelli, Atulya, and Thomas S. Huang. "Automatic video annotation using multimodal Dirichlet process mixture model." Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on. IEEE, 2008.

[12] Chen, Zhi-li, et al. "Effective linguistic steganography detection." Computer and Information Technology Workshops, 2008. CIT Workshops 2008. IEEE 8th International Conference on. IEEE, 2008.

[13] Dowman, Mike, et al. "A probabilistic model of meetings that combines words and discourse features." IEEE Transactions on Audio, Speech, and Language Processing 16.7 (2008): 1238-1248.

[14] Li, Lishuang, Zhuoye Ding, and Degen Huang. "Recognizing location names from Chinese texts based on max-margin markov network." Natural Language Processing and Knowledge Engineering, 2008. NLP-KE'08. International Conference on. IEEE, 2008.

[15] Meng, Peng, et al. "Linguistic steganography detection algorithm using statistical language model." Information Technology and Computer Science, 2009. ITCS 2009. International Conference on. Vol. 2. IEEE, 2009.

[16] Baomao, Pang, and Shi Haoshan. "Research on improved algorithm for Chinese word segmentation based on Markov chain." Information Assurance and Security, 2009. IAS'09. Fifth International Conference on. Vol. 1. IEEE, 2009.

[17] Wu, Ou, et al. "Rank aggregation based text feature selection." Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on. Vol. 1. IET, 2009.

[18] El Hassani, Ibtissam, AbdelazizKriouile, and Youssef BenGhabrit. "Measure of fuzzy presence of descriptors on Arabic Text Mining." 2012 Colloquium in Information Science and Technology. IEEE, 2012.

[19] Haji, Mehdi, et al. "Statistical Hypothesis Testing for Handwritten Word Segmentation Algorithms." Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on. IEEE, 2012.

[20] Zhou, Xiang-Dong, et al. "Handwritten Chinese/Japanese text recognition using semi-Markov conditional random fields." IEEE transactions on pattern analysis and machine intelligence 35.10 (2013): 2413-2426.

[21] Segarra, Santiago, Mark Eisen, and Alejandro Ribeiro. "Authorship attribution using function words adjacency networks." 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013.

[22] Rodrigues, Erica, et al. "Uncovering the location of Twitter users." Intelligent Systems (BRACIS), 2013 Brazilian Conference on. IEEE, 2013.

[23] Goyal, Anil, Mukesh K. Jadon, and Arun K. Pujari. "Spectral approach to find number of clusters of short-text documents." Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on. IEEE, 2013.

[24] Shen, Jau Ji, and Ken Tzu Liu. "A Novel Approach by Applying Image Authentication Technique on a Digital Document." Computer, Consumer and Control (IS3C), 2014 International Symposium on. IEEE, 2014.

[25] Ahmed, Al-Falahi, et al. "Authorship attribution in Arabic poetry." 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA). IEEE, 2015.

[26] Osiek, Bruno Adam, Geraldo Xexéo, and Luis Alfredo Vidal de Carvalho. "A language-independent acronym extraction from biomedical texts with hidden Markov models." IEEE Transactions on Biomedical Engineering 57.11 (2010): 2677-2688.

[27] He, Miantao, Miao Li, and Lei Chen. "Mongolian Morphological Segmentation with Hidden Markov Model." Asian Language Processing (IALP), 2012 International Conference on. IEEE, 2012.

[28] Alma'adeed, Somaya, Colin Higgens, and Dave Elliman. "Recognition of off-line handwritten Arabic words using hidden Markov model approach." Pattern Recognition, 2002. Proceedings. 16th International Conference on. Vol. 3. IEEE, 2002.

[29] Park, Hee-Seon, and Seong-Whan Lee. "Off-line recognition of large-set handwritten characters with multiple hidden Markov models." Pattern Recognition 29.2 (1996): 231-244.

[30] Espana-Boquera, Salvador, et al. "Improving offline handwritten text recognition with hybrid HMM/ANN models." IEEE transactions on pattern analysis and machine intelligence 33.4 (2011): 767-779.

[31] Imani, Zahra, et al. "offline Handwritten Farsi cursive text recognition using Hidden Markov Models." Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on. IEEE, 2013.

[32] Prasad, Rohit, et al. "Improvements in hidden Markov model based Arabic OCR." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.

[33] Zhou, GuoDong, and Jian Su. "Named entity recognition using an HMM-based chunk tagger." proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002.

[34] Xu, Rong, et al. "Combining Text Classification and Hidden Markov Modeling Techniques for Structuring Randomized Clinical Trial Abstracts." AMIA. 2006.

[35] Yi, Kwan, and JamshidBeheshti. "A hidden Markov model-based text classification of medical documents." Journal of Information Science (2008).

[36] Frasconi, Paolo, Giovanni Soda, and Alessandro Vullo. "Hidden markov models for text categorization in multi-page documents." Journal of Intelligent Information Systems 18.2-3 (2002): 195-217.

[37] Vogel, Stephan, Hermann Ney, and Christoph Tillmann. "HMM-based word alignment in statistical translation." Proceedings of the 16th conference on Computational linguistics-Volume 2. Association for Computational Linguistics, 1996.

[38] Schuller, Björn, Gerhard Rigoll, and Manfred Lang. "Hidden Markov model-based speech emotion recognition." Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on. Vol. 2. IEEE, 2003.

[39] Nwe, Tin Lay, Say Wei Foo, and Liyanage C. De Silva. "Speech emotion recognition using hidden Markov models." Speech communication 41.4 (2003): 603-623.

[40] Juang, Biing Hwang, and Laurence R. Rabiner. "Hidden Markov models for speech recognition." Technometrics 33.3 (1991): 251-272.

[41] Gales, Mark, and Steve Young. "The application of hidden Markov models in speech recognition." Foundations and trends in signal processing 1.3 (2008): 195-304.

[42] Conroy, John M., and Dianne P. O'leary. "Text summarization via hidden markov models." Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2001.

[43] Maskey, Sameer, and Julia Hirschberg. "Summarizing speech without text using hidden markov models." Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers. Association for Computational Linguistics, 2006.

[44] Kupiec, Julian. "Robust part-of-speech tagging using a hidden Markov model." Computer Speech & Language 6.3 (1992): 225-242.

[45] Thede, Scott M., and Mary P. Harper. "A second-order hidden Markov model for part-of-speech tagging." Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Association for Computational Linguistics, 1999.