International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-5, Issue-2, Feb.-2017
http://iraj.in

AN OVERVIEW OF HASHING TECHNIQUES USING NEURAL NETWORK

MRINAL JYOTI DEKA

Electronics and Communication Engineering, National Institute Of Technology, Silchar, India
E-mail: mrnl.dk@gmail.com

Abstract- For efficient and strong positive output in any industry, it is necessary to analyze the data involved in it; data analysis becomes a part of success in any projects. Therefore, research on big data is increasing day by day to find and design efficiently any updates related to any field. As in much application like large scale search, comparing and matching pattern, it has become difficult to have a straightforward solution due to limitations of computational complexity and consumption of large space requirement for memory. As a solution for this many approaches like Approximate Nearest Neighbor search based on hashing techniques and other randomized hashing methods have gained popularity, but the interpretation of them in many real-world applications has been proven to be insufficient. In the meanwhile many advanced methods to analyze the big data by in cooperating data-driven learning methods in the development of advanced hash functions have emerged. However, till date, no systematic review has been undertaken to analyse and summarize hashing technique using neural network. This paper mainly covers the trend of hashing techniques and its advanced methods in applications, activities described in the many sectors. The findings indicate that the popularity of this technique, its pros and cons. Overall Paper provides a comprehensive overview of articles published in academic journals. The review provides stakeholders with valuable information on the hashing approaches and about advanced hash functions.

Keywords- Big Data, Hashing techniques, Machine Learning and ANN.

I. INTRODUCTION

Advance in technologies have resulted in providing information in a massive rate. But organizing them and analysing these data is quite a challenge to researchers, industrialist in other word for any professionals. According to Jun Wang et. Alas, there is a dramatic increase in the data size modern technology must deal with such huge database. Here the searching can be visual images and videos, audios in the media sector. It can be similar data samples in a given database which relates to the issues of nearest neighbour. Also, according to Yao et.al, the development in web technology has resulted in increase of research activities in visual search and the similarity search is considered to be one of the problems. As the data involved in the search will be very large.

Deep Neural Networks has drawn maximum attention and many research efforts in various Artificial intelligence areas. Here hashing (mathematical functions are used and a value from string of text or sorting key values in an efficient manner is generated this is called hashing) plays a very important role. According to (Liu.W,2012) deep learning helps in learning robust and strong feature representation for highly complicated data; deep learning is leveraged for finding dense closely packed hash codes. Since one main purpose of deep learning is to learn robust and powerful feature representations for complex data, it is very natural to leverage deep learning for exploring compact hash codes which can be regarded as binary representations of data. In this section, we briefly introduce several recently proposed hashing methods that employ deep learning. The hashing technique‘s basic idea is to construct a series of hash functions so that each example is mapped into a compact binary code, making the hamming distance less and maximizing them for dissimilar examples simultaneously (Yao. et. al, 2016). According to (Zhu et.al, 2016) hashing has been widely installed to approximate nearest neighbor search for large-scale multimedia retrieval and hashing has been the high attention catcher in large data scale vision problems like image retrieval and recognition mainly because of its storage and efficiency in searching.

In spite of an increase in the popularity of hashing techniques using neural network in the field of Deep learning, no article to date has experimented to systemically summarize these applications. Therefore the aim of this paper is to provide details of research by reviewing of hashing technology using neural network through a robust content analysis of published peer-reviewed journal articles. Selected articles are reviewed and information about hashing techniques is compiled. The ultimate list of articles reviewed for this paper covers articles published in reputed referred scholarly journals. This review provides stakeholders with valuable information on different researches in deep learning.

II. HASHING TECHNIQUES

In fields like graphics, geometrical computation, telecommunication hashing methods are used intensively. Locality Sensitive Hashing (LSH) is one of the popular methods used.A important quality in LSH family of proficiency is a hashing function that with high probability returns the similar bit for data points which are nearby data points in the original measurable space according to (Wang J, 2015).

Even though LSH provides guaranteed performance, randomized techniques which are based on LSH declined due to many critical drawbacks. In order to overcome these drawbacks, many other hashing methods were proposed recently to grasp techniques of machine learning to produce more effective hash codes (Cayton and Dasgupta, 2008). To achieve this many knowledgeable tools of machine learning and algorithms have been adopted for designing hash function, including algorithm boosting according to (Shakhnarovich G, 2005). This paper attempts to summarize different basics of the hash and few popular methods are discussed.

The following are common terms used in hashing techniques

2.1 Script for Hashing

One can adopt a set of hash functions as

H = {h1.....hk} to find out a k-bit binary code

y={y1.....yk} for x as

y= {h1(X)... h2(X)..., h k(X)} for a given sample point X € RD

Kth bit is calculated as yk = h k(X)

The hash function can be mapped in different ways as

• h k: RD□ B

• as a hamming space:

H: x□{h1(x),…, hk (x)}

Also with given set of hash functions all the items in the database can be mapped to the corresponding binary codes. Once the binary codes are obtained one can perform ANN search in hamming space with minimum computation. Hamming distance is calculated in an efficient manner as a bitwise logic operation, according to (Wang J, 2015)

2.2 Basic Steps in Hashing based ANN search

There are basically three steps they are:

• Designing Hash functions

• Indexing the database items

• Online Querying with Hashing

2.3 Designing Hash functions

• Random projections or permutations are used in randomized hashing approaches

• Data distribution and various stages of supervised information(includes point wise labels, pair wise relationships, and ranking orders) are exploited in case of emerging learning to hash framework to find out optimal parameters of the hash functions.

Popular and commonly used hash functions are a generalized linear projection due to their efficiency

hk(x) = sgn (f(wkTx+bk)).

.In this equation f(.) is a non linear prespecified function

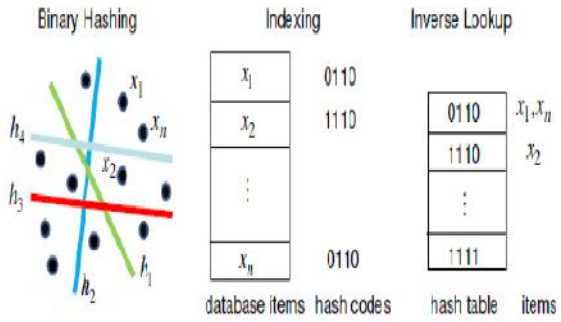
{wk , bk}K where k =1,wk is a projection vector and corresponding intercept bk

The data X is used to estimate these parameters along with supervised information

Along with different choices of f(.) leads to a broader range of hashing ways like LSH retains f(.) as identity function, in spectral hashing f(.) as cosine or sinusoidal function and in kernel based hashing as shift-invariant according to (WeissY,2008) and (Raginsky M, 2009)

2.4 Indexing Using Hash Tables

With a learned hash function, one can compute the binary codes Y for all the items in a database. In practice, the hash codes of the database are organized as an inverse-lookup, resulting in a hash table or a hash map as shown in Figure1.



Source: Wang J, 2015, Learning to Hash for Indexing Big Data - A Survey

Figure1.Demonstration of linear projection based binary hashing, indexing, and hash table construction for fast inverse lookup.

2.5 Online Querying with Hashing

The objective of this step is to find the nearest database items in a given query. The following steps are followed

• With the same hash functions that mapped the database items to code query is converted into a code

• By Computing the Hamming distance between the query code to all the database codes nearest neighbors of the query can be found.

2.6 Randomized Hashing Methods

Locality sensitive hash family is one of the examples of this. It uses maximum as hashing techniques due to its simplicity also it has interesting proximity preserving properties. In this Random Projection Based Hashing and Random Permutation-based Hashing, Learning Based Hashing Methods are categorized as follows:

• Data-Dependent vs. Data-Independent

• Unsupervised, Supervised, and Semi-Supervised

• Pointwise, Pairwise, Triplet-wise and List wise

• D. Linear vs. Nonlinear

• E. Single-Shot Learning vs. Multiple-Shot Learning

• F. Non-Weighted vs. Weighted Hashing

This paper aims in overviewing the work done on hashing networks by many researchers in a later section.

III. RESEARCH METHOD

Content analysis is considered to be a suitable method to make systematically and objectively correct interferences by literature review based on data collected according to Downe W, 1992

And it is considered as the suitable method because it quantifies specific phenomena and describes it. An important task in content analysis is the collection of

samples and determination of content analysis forms the present study. Figure.2shows the content analysis procedure used in

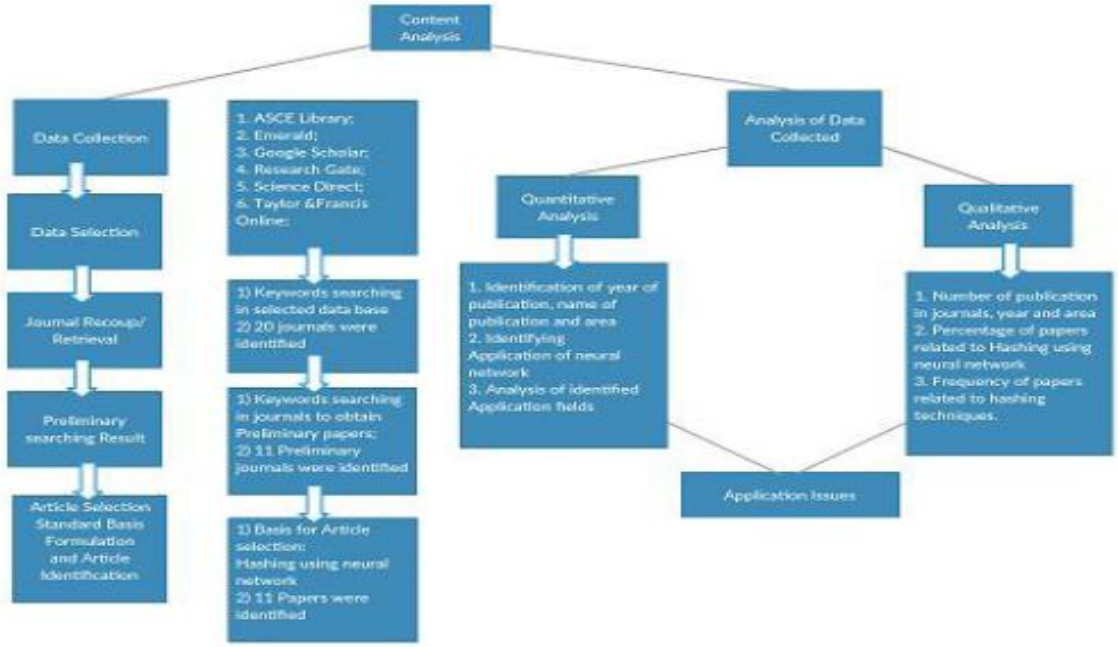


Figure2. Content analysis procedure for the study

Step1: Identification of Academic databases. Among the academic databases ASCE Library, Emerald, Google Scholar, Research Gate, Science Direct and Taylor &Francis Online, are used. Databases used here covers basics of "Deep learning and how to handle big data using Hash"

Step 2: Identification of the journals to be used for searching article. By implementing planned strategy academic journals relevant to this paper were found out, like "Hashing Techniques using neural network". This resulted in identifying 20 journals related to Hashing.

Step 3: By preliminary search, the websites of 20 journals were searched using "Application of Hashing Techniques".

Step 4: In order to select the relevant articles for study particular selection criteria have been adopted as follows

1. Research method "Hashing Techniques",
2. Application of Hashing Techniques in neural network".

Two round article selection procedures were followed as narrated by (Yang et al. 2011). In first stage articles were checked by its 'title, abstract, and keywords' to find out whether they meet the selection criteria. Next stage involved reading and analyzing the whole article to make a final selection of articles. Totally 11 articles were finalized among 20 academic articles. The analysis adopted in this work mainly includes both qualitative and quantitative content analysis. Qualitative content analysis is used to find the meaning of the data, recognize application fields, in-depth analysis, year of publication in journals and the area of publication.

The quantitative content analysis is used to find out the frequency. Mainly it is adopted to find the numbers of publications in journals; Also numbers with respect to years and geographical areas, activities involved in the development of the Hashing techniques in terms of percentage were found and they are temporarily represented in the form of tables.

IV. OVERVIEW OF HASHING TECHNIQUES USED IN NEURAL NETWORK AND IT'S APPLICATION

4.1 Trends

4.1.1 Distribution of Articles

Table 1 shows the distribution of the 11 articles. All these 11 articles are from a different application of hashing techniques.

Table1. Distribution of the articles

Journal title	Number of articles
ACM	1
CV paper	1
Elsevier, Science Direct Neural Networks	1
International Conference on Learning Representations	1
Internet resources	1
Springer	1
Proceedings of the 32nd International Conference on Machine Learning	1
IEEE	2
Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence	1
Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence	1

4.1.2 Publications Activities in the development of the Hashing in Neural networks over years

Figure.3 explains the trend in Hashing in Neural networks publications over time. It can be observed that a maximum number of publications during the year 2015, followed by the year 2016.Which indicates the popularity of Hashing in Neural networks in recent years.

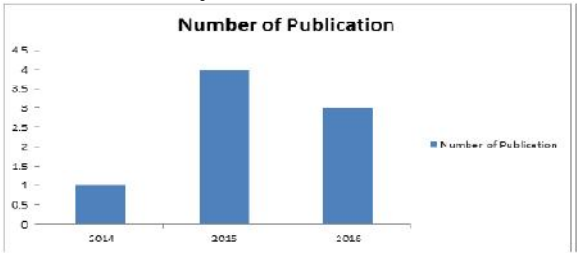


Figure.3Publication trend over years

4.1.3 Applications in Geographical Areas

Table 2 shows the distribution of hashing in neural networks publications by countries/regions. Totally there are 13 articles published about Hashing in Neural networks, among them USA context, accounting for 42% of all articles, followed by China (25%), Japan, Switzerland, Singapore and Canada accounting about (8%).

Country or Region	Number of publications	Percentage of publications (%)
China	3	25
Japan	1	8
USA	5	42
Switzerland	1	8
Singapore	1	8
Canada	1	8

Table 2. Distribution by countries or regions

4.2 Applications

Totally 7 different conference sectors in which Hashing in Neural networks are applied were identified as shown in Table.3 some of them are presented in International Conferences on Learning Representations, Machine Learning, and Artificial Intelligence and one of them presented in proceedings of the IEEE, remaining papers were highlighted Hashing for Compact Binary Codes Learning, Neural Networks, Journal of Big Data and lastly one of them from Internet resources. Details of these are described in the following subsection.

Application Fields	References
International Conference on Learning Representations International Conference on Machine Learning	(Bengioy.Y.et.al,2014) (Chen. W.et.al,2015)
Neural Networks	(Jürgen.S,2015), (Yao.T.et.al,2014)
International Joint Conference on Artificial Intelligence	(Li.W.et.al,2016),(Yao.T.et.al,2016),(Zhu.H.et.al,2016)
Hashing for Compact Binary Codes Learning	(Liong. V.et.al,2015)
Journal of Big Data	(Najafabadi.M.et.al,205)
Proceedings of the IEEE	(Wang.J.et.al,2015)

4.2.1 Deep Learning

Najafabadi et al. Journal of Big Data (2015) mainly focuses on two topics one of them is to find out how deep learning helps in Big Data for solving specific problems. And another one is to find out how specific areas of Deep Learning must be improved to solve and focus on the challenges associated with big data analytics.

In this paper, the main discussion concentrates on learning from huge volume data, semantic indexing, discriminative tasks, and data tagging and from improvement perspective, it focuses on learning from streaming data, handling high dimensionality of the data, expandable of models distributed and parallel computing.

The final paper concludes that deep learning has potential to provide solutions for the following

• To address the analysis of data and problems that occur in handling huge volumes of input data.

• More specifically, it aids in extracting complex data from unsupervised data due to this feature it is considered as the most valuable tool for Big Data Analytics.

• The hierarchical learning and extraction of different levels of the complex, data abstractions in Deep Learning provides a certain degree of simplification for Big Data Analytics tasks, especially for analyzing massive volumes of data, semantic indexing, data tagging, information retrieval, and discriminative tasks such a classification and prediction.

In this paper Deep learning (DL) in neural networks: An overview,(Schmidhuber.J,2015),has reviewed the deep supervised learning ,unsupervised learning, reinforcement learning & evolutionary computation, and short programs encoding deep and large networks by indirect searching. The main goal of this is to attribute credit to scientists who have contributed to the present state of the art. Here DL research is considered as continually progressing, a deep network, which has been influenced each other in a complicated way.

Here the research work done from 1960 to 2013 is an overview and highlights of it are as follows

In conclusion (Schmidhuber.J, 2015) says that for supervised and unsupervised learning and reinforcement learning deep learning in the neural network is very much required. To diminish the problems with deep Credit Assignment Paths Unsupervised learning smoothens sequences and stationary patterns of both supervised learnings and Reinforcement Learning (RL).

For both supervised and RL with deep NNs learning dynamic programming is very important.

4.3.2 Efficient Similarity Retrieval

In this paper, Deep Hashing Network (DHN) architecture for supervised hashing is proposed, which has ventured to make end user learn a good image representation combined to hash codes and to control error due to quantization is presented. Bayesian learning framework is used

There are four important components in the Deep Hashing Network (DHN) model mainly as follows:

• To capture image representations a sub-network with multiple convolution-pooling layers is adopted.
• To generate compact binary hash codes a fully-connected hashing layer is used
• For similarity-preserving learning, a pairwise cross entropy loss layer is implemented.
• For controlling hashing quality a pairwise quantization loss is adopted.
• Updated state-of-the-art hashing methods are used

To summarize this paper mainly focuses on learning different hash methods that built data dependent hash coding for efficient image retrieval focus on learning to hash methods that build data-dependent hash coding for efficient image retrieval, which has shown better performance than data-independent hashing methods.

According to (Wang.J.et.al, 2015) as there is a huge growth in big data more prominence started increasing in designing efficient indexing and search methods. Even though large-scale search and pattern matching has been done in many critical applications wherein finding the nearest neighbors to any problem is considered to be the fundamental research till now no straightforward solution is not feasible by large comparison due to the prohibitive computational complexity and memory requirement.

V. ISSUES WITH HASHING TECHNIQUES

Many issues were recognized while over viewing hashing techniques, among them some prominent issues are listed below as follows:

1. There will be a lack of theoretical assurance in many learning based hashing techniques on the returned neighbor quality according to (Wang.J.et.al,2015)
2. Even though many techniques have presented theoretical concepts of collision probability majority of them are based on randomized hash functions according to (Kulis, P. Jain, and Grauman K, 2009), (Jain.P.et al, 2010) and (Liu.W.et al, 2012)
3. (Mu.Y et al, 2014) says that as compact hash codes have a large potential in many other data modeling tasks like nonlinear kernel SVM classifiers and rapid kernel approximation the doubt that arises here is whether it is possible to use directly compact codes to achieve efficiency in unsupervised or supervised learning without affecting the accuracy.

6. CONCLUSIONS

According to (Schmidhuber J, 2015), deep learning is applicable for supervised, unsupervised and reinforcement learning for both deep supervised and reinforcement learning dynamic programming is essential. According to (Wang J, 2015) through designing a certain indexing structure, the ANN search with hashing methods can be even more efficient. (Zhu.H.et. al, 2016) says that Deep Hashing Network optimizes the pairwise cross-entropy loss on semantic similarity pairs and the pairwise quantization loss on compact hash codes simultaneously. Extensive experiments on standard image retrieval datasets show the DHN architecture yields substantial boosts over the state-of-the-art hashing methods.

REFERENCES

[1] Bengioy Yoshua, Courville Aaron, and Vincenty Pascal, “Representation Learning: A review and New Perspectives”, International Conference on Learning Representations
[2] Chen Wenlin, Wilson James T,Tyree Stephen,Weinberger_Kilian Q and Chen Yixin “Compressing Neural Networks with the Hashing Trick”, Proceedings of the 32 nd International Conference on MachineLearning, Lille, France, 2015. JMLR: W&CP volume 37.
[3] Jürgen Schmidhuber, “Deep learning in neural networks: An overview”, Elsevier Ltd. Neural Networks 61 (2015) 85–117.
[4] L.. Cayton and S. Dasgupta, “A learning framework for nearest neighbor search,” in Advances in Neural Information Process- ing Systems 20, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. Cambridge, MA: MIT Press, 2008, pp. 233–240.
[5] Li Wu-Jun , Wang Sheng, and Kang Wang-Cheng ,"Feature Learning Based Deep supervised Hashing with Pairwise Labels" Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)1711
[6] Liong Venice Erin, Lu Jiwen , Wang Gang , Moulin Pierre , and Zhou Jie ,“Deep Hashing for Compact Binary Codes Learning”,IEE explore
[7] M. Raginsky and S. Lazebnik, “Locality-sensitive binary codes from shift-invariant kernels,” in Advances in Neural Information Processing Systems 22, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I.Williams, and A. Culotta, Eds. MIT Press, 2009, pp. 1509–1517
[8] Najafabadi Maryam M, Villanus M tre Flavio , Khoshgoftaar, Taghi M, WaldRandall and magicuhare Edin “Deep learning applications and challenges in”, Journal of Big Data (2015) 2:1,DOI 10.1186/s40537-014-0007-7
[9] P. Jain, S. Vijayanarasimhan, and K. Grauman, “Hashing hyperplane Jain queries to near points with applications to large active learning,” in Advances in Neural Information Processing Systems 23, J. Lafferty, C. K. I. Williams, J. Shawe-Taylor,R. Zemel, and A. Culotta, Eds. MIT Press, 2010, pp. 928–936.
[10] W. Liu, J. Wang, Y. Mu, S. Kumar, and S.-F. Chang, “Compact hyper plane hashing with bilinear functions,” in Proceedings of the 29th international conference on Machine learning, 2012.
[11] Wang Jun , Liu Wei , Kumar Sanjiv, and Chang Shih-Fu,”Learning to Hash for Indexing Big Data - A Survey”, PROCEEDINGS OF THE IEEE PROCEEDINGS OF THE IEEE
[12] Y. Mu, G. Hua, W. Fan, and S.-F. Chang, “Hash-svm: Scalable kernel machines for large-scale visual classification,” in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, USA,2014, pp. 446–451.
[13] Yao Ting, LonSeg Fuchen, Mei, -Based Hashing for Image Tao and Rui Yong † “Deep mantic-Preserving and Ranking Retrieval”, Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)
[14] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng and Grégoire Mesnil “Learning Semantic Representations Using Convolutional Neural Networks for Web”, WWW’14 Companion, April 7–11, 2014, Seoul, Korea
[15] Zhu Han , Long, Mingsheng , WangJianmin,and Cao Yue ,“ Deep Hashing Network for Efficient Similarity Retrieval”, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence.

★★★