

Deep learning algorithm and location big data mining

Gao Faqin^{1, a}

¹ School of Information Science, Zhejiang Sci-Tech University, Zhejiang, Hangzhou, 310018, China

^agfqzjlg@126.com

Keywords: Machine learning; Data mining; Location big data (LBD); Deep learning algorithm (DLA); Deep neural network (DNN).

Abstract. In recent years, location big data (LBD) has become an important resource for intelligent transportation systems. LBD mining faces many serious questions, in order to tackle these questions, we introduce deep learning algorithm (DLA) into LBD mining. First, we explore the current research on deep neural networks (DNN) and DLA. Then, we design a framework that uses DLA in LBD mining from the view of probability. Finally, we provide some research trends of DLA after analysis.

Introduction

Location big data (LBD) generally consists of geographical data, vehicle trajectories, personal navigation data, and attendance records in social media networks, among other types of data. With the continuous growth of location based services (LBS), such as intelligent transportation systems, LBD has recently become a very important resource in building smart cities, such as mining the mobility rules of human groups or predicting your turns at a cross road [1-2]. For example, the mapping of social media feeds has helped governments worldwide gauge public sentiment during significant events in real time, such as in elections and uprisings [1].

LBD has different characteristics such as spaciality, temporality, mobility, and diversity, which makes LBD modeling very complex. Data dimensionality reduction technology is usually used to reduce the complexity of LBD mining. The dimensions of LBD should be reduced as much as possible while still ensuring a minimum of information loss to enhance the efficiency of LBD mining.

With the widespread applications of LBS and the rapid development of big data analysis technology, LBD mining has become an important tool for building smart cities and intelligent transportation systems. When compared to the requirement for practical application, big data analysis services and LBS productions are relatively lacking; much work is needed to create these productions or services. Moreover, LBD mining has become one of the technical bottlenecks of LBS development, and DLA as its core technology, is worthy of in-depth study.

Big Data Mining

LBD is complex, but it can be sparse [3]. It is worthwhile to investigate how useful information can be learned from LBD. The most commonly used technology to solve this problem is big data mining, and extensive work has been done in this area [3]. For example, some famous international journals such as “Nature” and “Science” have published special issues that discuss big data mining [4]. Gartner released the technology maturity curve of big data technology in July 2013, and pointed out that big data analysis and its mining technology is far from mature [3].

LBD grows rapidly and continuously, especially with the popularization of mobile internet and personal electronic devices that contain a position unit, such as smart phones [5]. We need to efficiently gain implicit high-order knowledge from LBD, which has gradually attracted the attention of many experts. Now, there is a general consensus that, in order to mine high-order abstract knowledge, we should introduce deep learning algorithms (DLA) [6]. A deep learning model consists of multilayer nonlinear operation units, where the outputs of low-level units serve as the inputs of high-level units between adjacent layers. As a result, a deep learning model can learn complex relationships from a large amount of input data, and it is a good method for data mining. Not unexpectedly, deep learning about big data analysis has also received a great deal of attention at some international meetings.

Development of DNN

In 2006, Hinton et al. [7] published an article in “Science,” in which they proposed a kind of unsupervised greedy algorithm named the contrastive divergence (CD) algorithm to train a DNN layer by layer, and they used CD algorithm to train a deep belief net (DBN). Almost in the same year, Bengio and Ranzato quickly followed the research for DNN and DLA, and their research started a trend for studying DLAs. For example, Kiwaki et al. have proposed a method to design a covariance matrix of initial GRBM distribution [8]. They analyzed the impact of initial distribution covariance on prediction accuracy. Moon et al. [9] and Bengio et al. [10] studied an unsupervised transfer learning algorithm; Lopes et al. [11] and Hinton et al. [12] have studied an adaptive learning rate adjustment method in order to solve the convergence problem.

In order to enhance the performance of deep learning technology, a large number of researchers such as Bengio, Hinton, Tieleman, Larochelle, Erhan, Ranzato, and Glorot [13~18] have extensively researched deep learning. Erhan et al. [15] have tried to understand the reason that unsupervised learning can help to enhance the performance of deep learning; and Glorot et al. [16] have researched the reason why the original training process of DNN was a failure. Hinton et al. proposed another DNN modeling method called deep Boltzmann machines (DBM) in 2009, which improved the ability of deep learning. In 2008, Tieleman et al. [17] revised the CD algorithm to maximize its likelihood, and put forward a new algorithm called persistent contrastive divergence (PCD). After that, Tieleman et al. studied the influence of Markov chain Monte Carlo (MCMC) sampling on the learning performance of restricted Boltzmann machine (RBM) [18], which is the basic component of DBM and DBN. Inspired by the research of Hinton and Tieleman, a series of RBM learning algorithms, based on the MCMC sampling algorithm, were proposed between 2010-2014 [19].

Structures of DNN

Besides the DBN model and DBM model that have been proposed by Hinton et al., Huang et al. have proposed a local convolution RBM learning model [20]. Poon et al. designed a new kind of DNN structure after they studied how to select a DNN structure to improve computational efficiency [21]. As for the improvements to the design of DBN model and DBM model, some papers have designed many kind of RBM models, such as mcRBM [22], ClassRBM, [23] and spike-and-slab RBM [24]. Some of these models have been successfully applied in some practical applications. In addition, Schmidhuber [25] and Glorot et al. [26,27] have analyzed the impact of mapping relationships on learning performance, as well as the impact of network depth on DLA performance. Bengio et al. testified that deep learning ability increases exponentially as the number of layers rises [10,28], but that the number of local extrema and calculation complexity also significantly increases. Regarding the neuron-model design of DNN, Nair et al. compared the learning performance of a DNN that used rectified neurons and logistic S neurons, respectively [29]. Glorot et al. designed a DNN deep learning model that uses rectified nonlinear neurons instead of logistic S neurons [27].

Practical Applications of DNN and DLA

In recent years, deep learning has achieved significant progress in many fields, including speech, image and natural language understanding, and its overall application. Work by Dahl et al. for the Microsoft Institute first used a DNN in speech processing in 2009, which has significantly decreased speech recognition error rates and has caused speech processing to become the first successful application field of deep learning [30]. In 2012, Hinton et al. made great breakthroughs in the image field by researching ImageNet evaluation using a deep convolutional neural network (CNN); the error rate with this method is reduced from 26% to 15% [31]. After further improvements, the error rate dropped to 1.2% in 2013 [32]. Taigman et al. have achieved a good result in face recognition using the same kind of DNN as mentioned in the literature [31], where the face recognition rate increased to a human recognition level [33]. Socher et al. of Stanford University have used a recurrent neural network (RNN) in sentiment analysis; the accuracy rate increased from 80% to 85% [34].

In China, Keda Xunfei Co. used DNN technology in its voice cloud platform in 2011 for the first time. Baidu Co. founded the IDL (Deep Learning Research Institute) devoted to the study of deep learning algorithms. Deep learning technology has a profound influence in Baidu, and has achieved remarkable achievements in speech recognition, OCR recognition, face recognition, image search,

and in other applications. In addition, other domestic companies, such as Sogou, have started to use deep learning technology in their products.

This section summarizes the current research situation of big data technology and deep learning technologies. Section 2 gives a special method that applies a DLA to LBD mining. Section 3 analyzes the research trend of LBD mining technology and deep learning technology.

Methodologies for Using DLA in LBD Mining

First, we extract features from the LBD to build the LBD model. The extracted features include the characteristics of selected road networks (such as the length of highways versus ordinary roads, the number of road intersections, and the bending degree of the road), points of interest (POI), the characteristics of individual movements (such as individual mobile uniqueness, mobile randomness [35], mobile periodicity [36], mobile expectation[37] and the similarity of mobile trajectories over a period of time), the characteristics of regional dynamics (such as regional comprehensive [38], regional mobility [36], and regional agglomeration characteristics, among others).

Then, we preprocess the LBD model with either matrix analysis or clustering analysis to reduce its dimension. After that, the parameters of simplified LBD model are put into the DNN, then the DNN takes these parameters as some kinds of deeper features related to specific applications.

Units of DNNs generally model input data as a probability graph; that is, the DNNs build the probabilistic model of the LBD through sample-based learning. The main function of the DNN is to use input data samples to learn the implicit distributions of θ and ϕ .

In this case, θ is the society semantic probability that some regional area is satisfied. For example, $\theta_{r_i} = \langle \theta_{r_i,1}, \theta_{r_i,2}, \dots, \theta_{r_i,\gamma} \rangle$ is the nonnegative normalized probability vector, which is composed of every social semantic variable in area r_i . On the one hand, θ can be used as the final result of social sensing; on the other hand, θ can also be used as a refined expression of location characteristics to participate in further analysis.

ϕ is a conditional probability distribution matrix. It consists of probabilities of every kind of mobile behavior in every social semantic context. By matching mobility patterns with some special social semantics, human society can be analyzed through LBD mining. For example, the social function of an urban area can be determined by analyzing the agglomeration characteristics of mobile events in this area.

Let social movement behavior after the occurrence of an event, E, be described as the trajectory $S = \langle r_1 \rightarrow \dots \rightarrow r_i \rightarrow \dots \rightarrow r_d \rangle$. There is a characteristic vector $\varphi_{r_i} = \{f_{i,1}, \dots, f_{i,j}, \dots\}$ of the location area r_i that has internal relationships with population transfer, so that the cost function of trajectory S is:

$$N(S, \varphi) \equiv N(S | V_w) = \sum_{r_i \in S} V_{w_i}^T \cdot \varphi_{r_i} = \sum_{r_i \in S} (w_{i,1} \cdot f_{i,1} + w_{i,2} \cdot f_{i,2} + \dots). \quad (1)$$

where $V_{w_i} = (w_{i,1}, \dots, w_{i,j}, \dots)$ is the unknown weight vector. It reflects the costs that occur when the crowd moves from location area r_i to any other location area.

Assume that the population's complete initiatory mobile trajectory $S_{r_i \rightarrow r_i}$ is observed when event E occurs. Then the probability that it reaches location area r_d satisfies the Bayesian formula:

$$P(r_d | S_{r_i \rightarrow r_i}, V_w) = P(S_{r_i \rightarrow r_i} | r_d, V_w) / P(r_d). \quad (2)$$

where $P(r_d)$ indicates the proportion of the population that moved to area r_d after event E occurs, which is a priori. This value is required to get $P(S_{r_i \rightarrow r_i} | r_d, V_w)$ and V_w before solving the above

formula. In this project, we would design the DNN first, and then the DNN would be trained to learn and obtain $P(S|V_w)$, V_w , and then calculate $P(S_{r_i \rightarrow r_j} | r_d, V_w)$

$$P(S_{r_i \rightarrow r_j} | r_d, V_w) = \frac{\sum_{S_{r_i \rightarrow r_d}} \exp N(S|V_w)}{\sum_{S_{r_i \rightarrow r_d}} \exp N(S|V_w)}. \quad (3)$$

$P(r_d | S_{r_i \rightarrow r_j}, V_w)$ can easily be calculated using formula (2)-(3), then we can obtain the probability of the appearance of trajectory S after event E happens.

Based on the above analysis, we can obtain modeling ideas about mining human community movement rules after some particular event E happens. Without a loss of generality, we can use these modeling ideas for other data analysis tasks, such as predicting the turns of a given driver at a given intersection according to historical driving data.

Prospects for Future Research

In general speaking, since 2006, the research on deep learning and its applications in big data analysis has made great progress, but there are still many problems yet to be solved, such as overfitting to the training data, online learning, the effective learning of multi-modal data. In this paper, we will discuss some research trends and solutions about these problems.

Dealing with the Overfitting Problem

In order to deal with the ‘overfitting’ problem, many researchers have put forward many improved methods, such as stopping training immediately when learning performance began to deteriorate, introducing weight penalties of various kinds (such as L1/L2 regularization and normalized-weight penalty factor [39]), Bayesian model averaging [40], and regularized Fisher mapping [41], but not all of these methods are suitable for deep learning of big data, due to its complexity. and these methods are only suitable for small and simple applications and are not suitable for complex nonlinear modeling tasks, such as DLA. To tackle this problem, experts are studying how to effectively tackle the “overfitting” problem. This paper argues that it is important to improve the generalization ability of deep learning algorithms to overcome the "overfitting" problem based on the idea of using a model averaging method to overcome the “overfitting” problem.

Online Deep Learning

When the distribution of a test set is different from that of the training set, the DLA will encounter a large error rate. Therefore, online learning or online adjustment technology is required in deep learning to guarantee the performance of big data analysis. However, it has been previously reported that online learning algorithms are likely to fall into local extremes, as long as they go into an adjustment stage [43]. As a result, online deep learning becomes more and more difficult when the depth of DNN increases.

Based on the research of our group, this paper argues that the inevitable development trend of online deep learning algorithm is based on the idea of neural network sampling and model averaging method. In order to overcome the influence of the changing model parameters on the precision of the prediction model, and to appropriately reduce the computational complexity, a specific task-independent and unsupervised online deep learning algorithm should be designed, which can be used to add the online learning function to the commonly used deep learning algorithms.

Learning from Multiple Modalities

Many recent papers have studied multimodal learning algorithms. But they cannot be used to effectively analyze unlabeled data, and they cannot easily handle the problem of missing data. Further study is needed to fully realize multimodal learning for big data. The best way to do this is to study how to communicate between different modalities to mine the relationships among them.

Through the analysis of both domestic and foreign research, this paper argues that a multiple-modality deep learning model should be designed based on a layered design idea, and its deep learning algorithm is designed based on the approximate maximum likelihood criterion. Also, in

order to further improve learning performance, study of the joint optimization method of each layer's parameters is required, as well as the method for generating lost modal data and the method of learning for the joint statement of multimodal data, which can also help to resolve some questions that may be caused by uncertainty factors, such as noisy or missing data.

Acknowledgements

This material is based upon work funded by Zhejiang Provincial Natural Science Foundation of China under Grant No. LY14F030025, and it is also financially supported by National Natural Science Foundation of China under Grant No. 61402417.

References

- [1]Information on <http://www.esri.com/products/arcgis-capabilities/big-data>,2015.
- [2]Information on <http://www.cio.com/article/2685967/big-data/big-data-analytics-gets-smarter-with-location-based-services.html>,2014.09
- [3]J N Liu,Y Fang, C Guo, et al: Geomatics and information science of Wuhan University, Vol. 39(2014), 4, p. 379-385.
- [4]X Q Cheng, X L Jin, Y Z Wang, et al:Journal of software, vol.25 (2014),9,p. 1889 - 1908
- [5]H B Ma, K Wang, T X Ma: geomatics & spatial information technology, vol.37 (2014),7,p. 19-22.
- [6]J W Liu, Y Liu, X L Luo: application research of computers, vol.31 (2014),7,p. 1921-1930,1942.
- [7]G Hinton,Osindero S, Y The: Neural Computation, vol.18 (2006),7,p. 1527-1554
- [8]Kiwaki T, Aihara K: Artificial Intelligence Research, Vol.4 (2015),1,p.53-59.
- [9]Moon S, Kim S, Wang H: arXiv preprint arXiv:1412.3121, 2014.
- [10]Bengio Y:Unsupervised and Transfer Learning Challenges in Machine Learning, 2012, 7,p.19.
- [11]Lopes N, Ribeiro B: Pattern Recognition, Vol.47 (2014),1,p.114-127.
- [12]Hinton G E. A practical guide to training restricted Boltzmann machines. In:Neural Networks: Tricks of the Trade. Springer Berlin Heidelberg, 2012.
- [13]Bengioy,Lamblinp,Popovicid,et al: Advances in Neural Information Processing Systems. Cambridge:MIT Press,2007,p.153-160.
- [14]Ranzatom, Poultneyc, Chopras,et al: Advances in Neural Information Processing Systems. Cambridge: MIT Press,2007,p.1137-1144.
- [15]Erhand, Bengio Y,Courvillea,etal: Journal of Machine Learning Research, Vol.11 (2010),2,p. 625-660.
- [16]Glorotx,Bengio Y,in: Proc of the 13th Int. Conf. on Artificial Intelligence and Statistics. 2010,p.249-256.
- [17]Tieleman T,In: Proc of the 25th Int. Conf. on Machine learning. ACM,2008,p.1064-1071.
- [18]Tieleman T, Hinton G,In: Proc of the 26th Annual Int. Conf. on Machine Learning. ACM, 2009, p.1033-1040.
- [19]Fischer A, Igel C: Pattern Recognition, Vol.47 (2014),1,p.25-39.
- [20]Huang G B, Lee H, et al. In: Proc of IEEE Conf. on Computer Vision and Pattern Recognition. 2012, p.2518-2525.
- [21]Poon H, Domingos P. In: Proc of IEEE Int. Conf. on Computer Vision Workshops. 2011, p.689-690.

- [22]Ranzato M, Hinton G E. In: Proc of IEEE Conf. on Computer Vision and Pattern Recognition. 2010,p.2551-2558.
- [23]Louradour J, Larochelle H: arXiv preprint arXiv:1103.4896, 2011.
- [24]Courville A, Bergstra J, Bengio Y. In: Proc. of the 28th Int. Conf. on Machine Learning (ICML-11).2011,p 1145-1152.
- [25]Schmidhuber J:Neural Networks, 2015, 61: 85-117.
- [26]Glorot X,Bengio Y. In: Proc of the 13th Int. Conf. on Artificial Intelligence and Statistics. 2010,p.249-256.
- [27]Glorot X,Bordes A,Bengio Y. In: Proc of the 14th Int. Conf. on Artificial Intelligence and Statistics. 2011,p.315-323.
- [28]Bengio Y, Delalleau O. In: Proc of Algorithmic Learning Theory. Springer Berlin Heidelberg, 2011, p.18-36.
- [29]Nair V, Hinton G E. In: Proc of the 27th Int. Conf. on Machine Learning.2010,p.807-814.
- [30]Dahl GE, Yu D, Deng L, Acero A:IEEE Trans. on Audio, Speech, and Language Processing, Vol.20 (2012),1,p.30–42.
- [31]Krizhevsky A, Sutskever I, Hinton GE. In: Advances in Neural Information Processing Systems. 2012.p. 1097–1105.
- [32] Information on <http://www.image-net.org/challenges/LSVRC/2013/results.php>
- [33]Taigman Y, Yang M, Ranzato M, Wolf L. In: Proc. Of the IEEE CVPR. 2014.
- [34]Socher R, Perelygin A, Wu J Y, Chuang J, Manning CD, Ng AY, Potts C. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing (EMNLP). 2013,p.1631–1642.
- [35]Song C, Qu Z, Blumm N, et al:Science, Vol.327 (2010),5968,p.1018-1021.
- [36]Liu SY, Liu YH, Ni LM, et al. In:Proc of the 16th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. New York:ACM, 2010. P.919-928.
- [37]Noulas A, Scellato S, Lathia N, et al. In: Proc of IEEE 12th Int. Conf. on Data Mining (ICDM). IEEE, 2012,p. 1038-1043.
- [38]Cranshaw J, Toch E, Hong J, et al. In:Proc of the 12th ACM Int. Conf. on Ubiquitous Computing. New York: ACM, 2010.p. 119-128.
- [39]Zeiler M D, Fergus R:arXiv preprint arXiv:1301.3557, 2013.
- [40]Xiong H Y, Barash Y, Frey B J:Bioinformatics, Vol.27 (2011),18,p.2554-2562.
- [41]Wong W K,Sun M:IEEE Trans on Neural Networks, Vol.22 (2011),10,p.1668-1675.
- [42]Hinton G E, Srivastava N, Krizhevsky A, et al:arXiv preprint arXiv:1207.0580, 2012.
- [43]Weston J,Ratle F,Collobert R.In: Proc of the 25th Int. Conf. on Machine Learning. New York: ACM Press,2008,p.1168-1175.