Using Deep Learning to Mine the Key Factors of the Cost of AIDS Treatment

Dong Liu[1,2], Zhidong Cao[2(✉)], and Su Li[1]

[1] Beijing Key Laboratory of Big Data Technology on Food Safety, Beijing Technology and Business University, Beijing 100048 People's Republic of China

[2] State Key Laboratory of Complex Systems Management and Control, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 People's Republic of China

zhidong.cao@ia.ac.cn

Abstract. The medical burden of AIDS is a significant public health problem. However, it is affected by the multiple factors, among which there is yet some vague cognition, and further exploration is necessary. Thus, the artificial neural network (ANN) and restricted Boltzmann machine (RBM) be treated as the infrastructure of deep neural networks (DNN), mainly based on the features of demography, pathology and clinical manifestation of AIDS patient's medical records to mine the impact factors of AIDS cost. And the proposed model could bring to light the previously uncharted latent knowledge and concepts. Based on reliable healthcare delivery, to inhibit the number of hospital days, intensive care and hospitalized frequency plus other sensitive factors, and avoid secondary infection and exposure to allergic reactions can obviously reduce the AIDS cost.

Keywords: ANN · RBM · AIDS cost · Impact factors

1 Introduction

AIDS (acquired immune deficiency syndrome, AIDS) is a significant public health problem. The joint assessment for AIDS indicated that as of September 2016 in China the cumulative number of deaths has exceeded 201,000 [1]. The patient life suffers a serious threat, but the biological technology resource of curbing AIDS is very scarce and precious. Thus, the scientific and efficient control in AIDS cost became a hot issue in decision-making of healthcare delivery.

The most widely used methods, such as linear regression, logistic regression(LR), clustering methods and SVM etc., have achieved some results [1–4]. However, the linear regression requires the normal distribution and variance homogeneity, plus a presumed linear relationship between the argument and dependent variables. Thus it performs feebly in nonlinear correlation. And the shallow machine learning models (e.g., SVM, LR, etc.) in complicated structure data is lack of expression ability, and could lead to dimension disaster [5]. Obviously, the traditional methods had been sinked into dilemma. Fortunately, the deep models which simulate the hierarchical structure in human brain can achieve the hierarchical expression of data [6]. This way

H. Chen et al. (Eds.): ICSH 2017, LNCS 10347, pp. 280–285, 2017.
https://doi.org/10.1007/978-3-319-67964-8_28

can not only imitate the procedure of human thinking to improve efficiency overwhelmingly, but can identify and characterize implied knowledge and concepts [7, 8].

2 Materials and Methods

2.1 Data

The data sample in this study is derived from the AIDS-related HIS systems in China. A total of 2618 in hospitalized medical records were randomly selected from 2006 to 2010. The data items features include the age, gender, address, payment methods, hospitalized frequency, etc. More crucially, in this paper we put these features as the independent variable, and the actual AIDS cost was take as the dependent variable.

2.2 Methods

Primitively Hinton and Sejnowki were rooted in statistical mechanics research to propose the Boltzmann machine (BM) [5]. Although BM has excellent unsupervised learning ability, its pre-learning process is over complicated. Thus, the RBM come from the simplified BM, and has a prominent advantage, which has been rigorously proved by Bengio that RBM can fit any discrete distribution if the number of hidden layer units is enough [7]. Assuming that (**x**, **h**) represent the state of visible layer and hidden layer. Based energy function, we can deduce the joint probability distribution and marginal probability distribution (see formula (1)).

$$\begin{aligned} &P(\mathbf{x},\mathbf{h}|\theta) = \frac{e^{-E(\mathbf{x},\mathbf{h}|\theta)}}{Z(\theta)} \textbf{ under s.t. } Z(\theta) = \sum_{\mathbf{x},\mathbf{h}} e^{-E(\mathbf{x},\mathbf{h}|\theta)} \Rightarrow P(\mathbf{x}|\theta) = \frac{1}{Z(\theta)} \sum_{\mathbf{h}} e^{-E(\mathbf{x},\mathbf{h}|\theta)} \\ &\textbf{Energy function}: E(\mathbf{x},\mathbf{h}|\theta) = -\sum_{i=1}^{n} b_i x_i - \sum_{j=1}^{m} c_j h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} x_i W_{ij} h_j \end{aligned} \tag{1}$$

In order to obtain $P(x|\theta)$, it must require 2^{m+n} calculations to determine $Z(\theta)$, which is called as normalized factor. Therefore, the determination of probability distribution of RBM was still very hard. RBM had continued to be progression-free in technology till the invention of contrastive divergence algorithm (CD) and its improved algorithm [8, 9].

The initial design for DNN highlighted its hybrid strategy in model building, and Deng deeply analyzed this characteristic [10]. Combining with ANN, the unsupervised learning module can construct a deep model, such as Google's cat [5, 10]. Considering the existence theorem of mapping, that any continuous function can be approximated at a desired accuracy by a three layer perceptron network (ANN with one hidden layer) [11]. Moreover, the more layers in networks the more slow in convergence it will become. In light of this, as shown in Fig. 1, this study by cascading up two layer RBMs and a three layer ANN constructed a DNN. Those RBMs at the bottom of networks could deeply mine the information hidden in raw data, and then form the feature expressions which can act as the input of ANN. The top layer in ANN has regression forecast function, and can calculate the sensitivity of impact variables. In order to

reveal how significant those variables, the variable sensitivity is defined as the important degree of input variables of networks. Namely the sensitivity analysis is to change a certain part of model, for observing the corresponding changes in network, so as to determine the important degree of this part for networks (formula (2)).

$$S_m = \frac{1}{n}\sum_{j=1}^{n}\frac{\max\limits_{i=1\to n}\{f(x_i)\} - \min\limits_{i=1\to n}\{f(x_i)\}}{\max\limits_{i=1\to n}\{f(x_i)\}} \tag{2}$$

In the training phase, we conduct layer-by-layer. Namely when it was RBM we adopted a modified CD algorithm [12] to train, while about ANN we take BP algorithm for pre-learning. In this way our networks can obtain better initial weights. In the fine-tuning stage, we used BP algorithm to fine tune the whole networks.

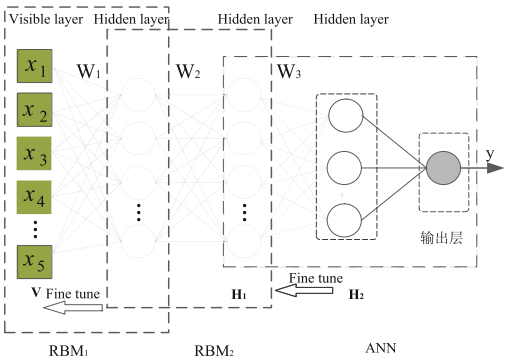

Fig. 1. Schematic diagram of model structure

3 Experimental Results

We primarily carried out a statistical description, which are shown in Tables 1 and 2. In this way, we can get the bottom feature operators. It indicated that the variables of AIDS cost affected each other. Among them there are a lot of complicated Non-linear relationships.

For eliminating the impact of different improved BP algorithms on the DNN performance and efficiency, this paper would use the LM algorithm, variable scale conjugate gradient algorithm. And the experiments noted that the LM algorithm is evidently best among them regardless of fitting and generalization ability. In Fig. 2, it indicated that: the sensitivity of hospital days, days of intensive care, days of sickness in danger, and hospitalized frequency are evidently exceed 0.6, plus rescue times, admission condition etc. are between 0.44 and 0.66, but ethnic, gender and whether follow-up are below 0.4.

Table 1. Discrete variables qualitative description

Var.	Categories	Qty.	Pct./%	Var.	Categories	Qty.	Pct./%
Gender	Man	1645	62.83	Admission Condition	Danger	116	4.43
	Woman	973	37.17		Urgent	867	33.11
Age	≤ 30	309	11.80		General	1640	62.64
	30 ~ 60	1497	57.18	Exchange department	Yes	627	23.95
	60 ~ 70	713	27.23		No	1991	76.05
	>70	99	3.78	Follow up	Yes	231	8.82
Marital status	Married	1725	65.89		No	2387	91.18
	Unmarried	307	11.80	Allergic drug categories	Chinese herb	1309	50
	Divorce	57	2.18		HIV antitoxin	491	18.75
	Widowed	529	20.21		Broad-spectrum antibiotics	572	21.85
Ethnic groups	Han	1797	68.64		Sulfonamides	157	6.01
	Minority	821	31.36		Penicillin	89	3.41
Work categories	Worker	312	11.92	Regions	North China	148	5.65
	Farmer	1184	45.23		South China	330	12.61
	Cadre	115	4.39		Central China	178	6.79
	Students	61	2.33		Northwest	669	25.55
	Children	10	0.38		Southwest	1075	41.06
	Other	936	35.75		Northeast	53	2.02
Payment methods	Basic medical insurance	1373	52.44		East China	165	6.30
				Admission methods	Outpatient clinic	1513	57.79
	Commercial insurance	20	0.76		Emergency	321	12.26
	Free	236	9.01				
	Self-funded	249	9.51		Referral	784	29.95
	Other	460	17.57				

Table 2. Continuous variables qualitative description

Var.	Min	Max	AVG.	SD.
Days of sickness in danger	0	43	15.17	48.37
Hospital days	1	88	23.16	26.54
Days of intensive care	1	56	6.39	27.61
Hospitalized frequency	1	10	4.27	4.13
Rescue times	0	12	0.83	0.821
Age	1	83	54.62	12.413

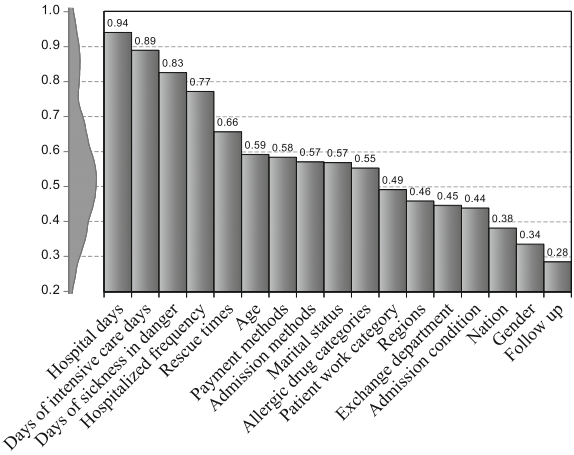

Fig. 2. The sensitivity of impact factors of AIDS cost

4 Conclusion and Discussion

However, the deep model is like a black box,which only focuses on input and output. Thus, we conducted the variables sensitivity analysis to explain the proposed deep model. The sensitivity is greater than 0.5 variables, can be considered very important; otherwise they can be considered more important, because they both are not strictly different. But AIDS is particular than general diseases after all, and here is some evidences: (1) the sensitivity of age was close to 0.6, because AIDS attacks the immune system in human body, and the older, the worse resistance. Once sick easily lead to complication, the medical cost will rapidly increase; (2) as AIDS is fatal, the patients' payment method and occupation categories for HIV aren't as sensitive as other diseases cost; (3) but the sensitivity of marital status is higher than other diseases cost to blame the AIDS's infectious mode; (4) also the allergic drugs categories are more sensitive, because the immune ability of patients is weak or disorder. In case of allergic reaction occurring, it can damage to the normal cell and tissue, and the medical expenses increased; (5) the sensitivity of intensive care is close to 0.9, mainly because of the expensive service. Initially AIDS cannot be spread, the patients with weak immunity who must prevent an infection from other diseases avoid endangering their lives, and thus they need the special care in quarantine. These indicated that the deep model can identify and characterize implied knowledge and concepts.

Based on reliable healthcare delivery, to reduce the hospital days, intensive care and hospitalized frequency is primary path for reducing AIDS's hospitalization cost effectively. Consciously we must commitment to preventing the infection (between patients and others around them, patients and patients with other virus), allergic reactions and related variables, such as age, allergic drug classification, etc. In this way, we can't reduce the unnecessary treatment program and economy burden, but also can improve the efficiency of hospital, even reducing unnecessary waste of medical resources.

Acknowledgement. This work is partially funded through NNSFC Grants #91546112. National Key R&D Program #2016YFC1200702 and #2016QY02D0200.

References

1. Zhang, Y.Q., Qin, X., Zhou, L., et al.: The AIDS epidemic and economic input impact factors in Chongqing, China, from 2006 to 2012: a spatial-temporal analysis. BMJ Open **5**(3) (2015)
2. Zhang, X.L., Zhang, Y.R., Aleong, T.H., et al.: Factors associated with the Household Income of Persons Living with HIV/AIDS in China. Global J. Health Sci. **4**(3), 108–116 (2012)
3. Harmon, T.M., Fisher, K.A., Mcglynn, M.G., et al.: Exploring the potential health impact and cost-effectiveness of AIDS vaccine within a comprehensive HIV/AIDS response in low and middle-income countries. PLoS ONE **11**(1), e0146387 (2015)
4. Stover, J., Bollinger, L., Izazola, J.A., et al.: What Is required to end the AIDS epidemic as a public health threat by 2030? The Cost and Impact of the Fast-Track Approach. PLoS ONE **11**(5), e0154893 (2016)
5. LeCun, Y., Bengio, Y., Hinton, G.: Deep Learing. Nature Mag. **521**(7553), 436–444 (2015)
6. Ryota, S., Shusuke, Y., Yasutaka, M., et al.: Deep learning application trial to lung cancer diagnosis for medical sensor systems. In: 2016 International SoC Design Conference (ISOCC), pp. 191–192 (2016)
7. Le, R.N., Bengio, Y.: Representational power of restricted boltzmann machines and deep belief networks. Neural Comput. **20**(6), 1631–1649 (2008)
8. Hinton, G., Osindero, I.Y.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
9. Tieleman, T., Hinton, G.E.: Using fast weights to improve persistent contrastive divergence. In: Proceedings of the 26th International Conference on Machine Learning, Helsinki, Finland, pp. 1064–1071 (2008)
10. Deng, L.: A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans. Signal Inf. Proces. **3**, 14–43 (2014)
11. Furundzic, D., Djordjevic, M., Bekic, A.J.: Neural networks approach to early breast cancer detection. Syst Architect **44**, 617–633 (1998)
12. Tieleman, T.: Training restricted Boltzmann machines using approximations to the likelihood gradient. In: Proceedings of 25th International Conference on Machine Learning, New York, pp, 1064–1071. ACM (2008)