City-state size prediction based on deep learning quantitative modeling

Jixuan Wang

Keystone Academy, Beijing, China

oliverus2024@163.com

Keywords: Deep learning; LSTM; quantization models; city-state size prediction

Abstract: The Lu County Old City Site is a Western Han Dynasty site discovered during the pre-construction archaeological exploration of the urban sub-center in 2016.By studying the case of the Lu County Old City Site, we can explore and discover the factors that determine the size and development of ancient cities. In this paper, we will quantify the factors affecting the scale and development of ancient cities and construct a model of the important variables and their correlations that determine the scale of ancient city-states. The research methodology includes collecting data and quantitatively analyzing the city size and a variety of related variables (e.g., political system, social culture, economic level, agricultural level, foreign trade, and local climate, etc.), and interpreting the size of ancient city-states as a model accompanying the development of the important variables based on the modern theory of urban economy to illustrate how the size of the city was key variables such as political system, social culture, economic level, and agricultural level Correlation. We adopt the method of constructing a knowledge graph to build a Bert-BiLSTM-Attention-CRF model that incorporates the mechanism of attention, thus accurately and efficiently quantifying the multiple indicators associated with the prediction of city-state size. The model is divided into four layers: the first layer is the BERT layer, which pre-trains the input text to obtain word vectors; the second layer is the BiLSTM layer, which takes the word vectors obtained from the first layer as the input and performs bi-directional training to capture the contextual features of the text; the third layer is the Attention layer, which assigns weights to the different contextual information, and extracts the features that are crucial for knowledge recognition; the fourth layer is the CRF layer, which decodes and annotates the output of the previous layer to obtain the global optimal sequence. Finally, the accuracy of the proposed quantization model and the feasibility of the city-state size prediction model are demonstrated through several sets of experiments.

1. Introduction

The ruins of Lu County Old City is a Western Han Dynasty site discovered in 2016 during the pre-construction archaeological exploration of the city's sub-center, and its discovery advances the history of Tongzhou's founding to at least 2,200 years ago at the beginning of the Western Han Dynasty, and it is the earliest and only known Qin-Han Dynasty city in the Tongzhou area. The Western Han Dynasty was a period of remarkable growth and development in China, marked by political stability, economic prosperity, and cultural expansion. The Tongzhou Lu County provides invaluable data on the urban development patterns of the period, reflecting the complex interplay between socio-political dynamics, economic conditions, and cultural influences.

Archaeology provides a unique perspective on the evolution of ancient cities, i.e., the size and growth of these settlements can be traced by analyzing variables such as floor space, economic development, and other relevant aspects. Specifically, the size of cities in various historical periods is influenced by many factors and variables. By studying the case of the Lu County Old City site, we can explore and discover the factors that determine the size and development of ancient cities, based on the underlying theoretical basis of the Settlement Size Theory, i.e., the number of social interactions, the total socio-economic outputs in a city, and the size of the built-up area of a city are correlated to, and quantitatively predictable by, the size of its population.

In this paper, we will quantify the factors affecting the size and development of ancient cities and

DOI: 10.25236/iwmecs.2023.007

construct a model of the important variables and their correlations that determine the size of ancient city-states. The research methodology includes collecting data and quantitatively analyzing the city size and a variety of related variables (e.g., political system, socio-culture, economic level, agricultural level, foreign trade, and local climate, etc.), and interpreting the size of ancient city-states as a model that accompanies the development of important variables based on modern urban economic theories to illustrate how the size of the city is related to key variables such as political system, socio-culture, economic level, and agricultural level. Correlation.

This study provides a new approach to understanding the size and function of ancient cities by emphasizing the correlation between key variables and city size, providing a model that can be applied to other ancient city sites, which is of great importance to archaeologists and historians, providing a quantitative means of predicting the size of ancient city-states in a given historical period and region. This quantitative tool can be used not only for city-state archaeology, but is also applicable to contemporary studies of urban development.

2. Related Work

2.1. Quantitative modeling of city size

Foreign research on urban scale mainly starts from efficiency maximization, and the research in this regard can be traced back to the early theory of agglomeration economy; Marshall's theory of scale economy puts forward the concepts of internal economy and external economy, and Weber puts forward the "agglomeration economy" for the first time, which emphasizes the scale of the economy from the spatial point of view. Subsequently, the rise of new economic geography not only focuses on the promotion of agglomeration economy for economic development, but also begins to pay attention to the trade-off mechanism between scale economy and crowding effect, explains the motivation of spatial aggregation to continuously strengthen itself, and constructs the landmark "core and edge" theory. Mills established a general equilibrium model for the optimal size decision of a city by comparing the marginal benefit and marginal cost. Mills builds a general equilibrium model of optimal city size decision by comparing marginal benefits and marginal costs, in which marginal benefits arise from urban agglomeration effect and marginal costs are due to the increase of commuting costs. dixst [1] considers external diseconomies to be equivalent to congestion effect, which is reflected in the degree of congestion on commuting, and Henderson [2] believes that the optimal city size is a result of the joint effect of urban agglomeration economy and agglomeration diseconomies. Sullivan [3] and others have since similarly introduced agglomeration economy effects into the production function model to explore in depth the growth of the urban economy and the optimal size of the city that may exist. Overall, a large number of studies on agglomeration economy and diseconomies together show that the economic efficiency of cities may show an inverted "U" shape change with the expansion of city size, which will first increase and then decrease.

While the theory of benefit maximization serves as the basic principle of the traditional measure of urban scale efficiency, the emergence of some new ideas has also attracted attention in the academic community. For example, Fu-chen and Kamal [4] believe that the efficiency of urban scale is not only the relationship between cost and efficiency, but also should pay attention to the functional links within the city, that is to say, in different stages of development, manufacturing or service industry to provide services mainly when the function, has a different appropriate city size; Bertinelli [5] that the reasonable size of the city should be the result of the coordinated development of the three aspects of the economy, society and the environment. and environment; Combes [6] studied the optimal city size by maximizing the real income and found that the optimal city size is also related to the trade cost of agricultural products. In addition, the urban network theory criticizes the traditional theory of optimal city size and believes that urban efficiency should not be a static process based on population, but a dynamic process based on urban network. Kim [7] analyzes the optimal size efficiency of South Korea under this viewpoint, breaks the traditional functionalist study of urban hierarchy, and emphasizes a new type of city formed by factor flow dominated by capital, labor, and information. Factor flows formed a new type of urban economic linkage, which provides us with a new research perspective.

A large number of theories and practices at home and abroad have repeatedly confirmed that city scale affects the efficiency of urban economy and the role of specialized division of labor on urban development, and the studies are full of content and have certain guiding significance and reference value. However, there also exists research space in the following three aspects: first, most of the existing studies on urban scale and urban economy examine the play of urban scale efficiency from the perspectives of population scale, land scale, and changes in scale structure, and few literatures combine urban scale with specialization and division of labor to conduct research; second, almost all studies on specialization and urban agglomerations have examined the impact of specialization and diversification on urban agglomerations, and few studies have examined the contribution of functional specialization to the development of urban agglomerations; third, there are a limited number of studies that place city size, specialization and urban productivity under a unified analytical framework. Such studies lack empirical research on real-life examples and have not examined much the possible impact of city size, specialization and division of labour on urban productivity.

2.2. Multifactor quantitative modeling

Multi-factor modeling refers to the comprehensive consideration of multiple factors to determine the upward or downward trend of a particular model, mostly used in financial markets, scale forecasting and so on. The proposal and development of multi-factor model theory initially started from abroad. 20 century Markowitz (Markowitz) portfolio theory, capital asset pricing model (CAPM), arbitrage pricing theory (APT), Fama-French three-factor model together constitute the four major theoretical foundations of multi-factor model.

The portfolio theory proposed by Markowitz [8] is considered to be the beginning of modern financial economics, which falls the measure of risk of stock returns on the variance or standard deviation of the rate of return and argues that when constructing a portfolio, an investor is mainly balancing the interrelationships between the expected return and the extreme variance of the expected return. Fama and French [9] conducted an analysis of the Fama and French [9] analyzed the U.S. stock market data from 1929-1963, and found that the traditional CAPM model for small-capitalization, low price-to-book ratio companies have higher average returns, and the stock market's β-value alone can not adequately explain this phenomenon. Beginning of the 21st century, multifactor modeling has been widely used in the field of quantitative stock selection.Asness et al. [10] explored the Fama-French three-factor model in depth, and found that the Fama-French three-factor model is the most important model in the quantitative stock selection. Asness et al [10] explored the Fama-French three-factor model in depth and found that the "value + momentum effect" model has greater advantages, thus giving another three-factor model composed of market, value, and momentum factors that can describe the pattern of changes in asset returns in the global market. Fama and French [11] carried out a deeper excavation of the three-factor model, and added an innovative model of earnings, which can be used for stock selection. Fama and French [12] in the three-factor model to deepen the mining, and innovatively added the profit and investment class factors, and then constructed a five-factor stock selection model, and the use of historical data to test the model, and ultimately determine the validity of the factors.

3. Modeling Algorithm

3.1. City-state size prediction model

In this chapter, we first identify the characteristic indicators affecting city size and construct a prediction formula for city size based on these indicators, thus emphasizing the correlation between key variables and city size. It should also be noted that these indicators are not all numerical indicators that have been quantified, such as political system, social culture, economic level, agricultural level, foreign trade and local climate, etc., so we need to construct a deep learning network for the quantification of the corresponding indicators.

The selected indicators related to city size prediction are as follows: Urban agricultural level (TFP): mainly obtained by quantifying the information related to the area of arable land and the number of working population; The level of urban economy and foreign trade (POP, POP2): mainly obtained through the quantification of information related to the types and quantities of commodities, transaction frequency, and production efficiency; The level of transportation development (RFI): mainly obtained through the quantification of information related to the number and length of roads; Social and cultural level (HUM): mainly obtained by quantifying information related to the number and value of cultural and artistic works; Governmental role (GOV): mainly obtained by quantifying the information about political system and policies;

The prediction model of city size can be expressed as:

$$City = [w_1, ..., w_6]\begin{bmatrix} f_1 \\ ... \\ f_6 \end{bmatrix} \tag{1}$$

Deep learning NLP models to analyze financial public opinion ultimately serve to design quantitative factors for quantitative investment strategies. Here, the predicted stock review sentiment of each stock on each trading day of the day is used as a quantization factor according to the ratio of the number of POSITIVE and NEGATIVE entries to the total number of comments on that day, calculated as:

The financial opinion factor values calculated from the above can be used as trading signal components for subsequent quantitative investment strategies and participate in further analysis.

3.2. Indicator quantification model

In this paper, the BERT model incorporating the attention mechanism is proposed to be used to engage in financial opinion text analysis [13]. The model is divided into four layers in total: the first layer is the BERT layer, which pre-trains the input text to obtain word vectors; the second layer is the BiLSTM layer, which takes the word vector information obtained from the first layer as the input, and performs bi-directional training on it to capture the contextual features of the text; the third layer is the attention layer, which assigns weights to the different contextual information, and extracts the feature information that is crucial for knowledge recognition; the fourth layer is the CRF layer, which decodes and annotates the output of the previous layer to obtain the global optimal sequence and classify the text information.

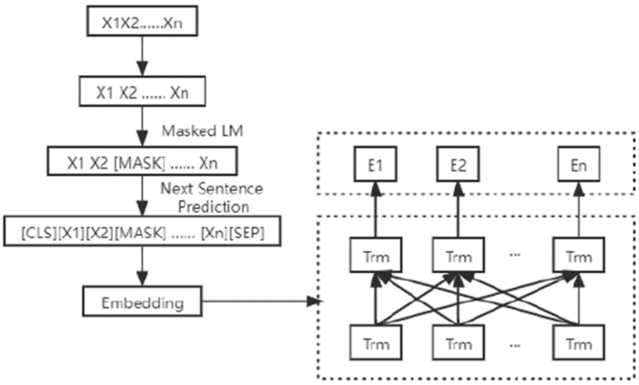

Figure 1 Bert Model structure.

Bert is a model for pre-training language, the flow of the model is shown in Fig. 1. The Bert model uses Masked Language Model (MLM) and Next Sentence Prediction (NSP) for pre-training.During the training process, 15% of the words are randomly masked using the [MASK] flag, and the prediction of the masked words is made on the basis of the contextual content. On this basis, two random sentences in the text are judged to determine whether there is a continuous relationship or not, and the beginning of the sentence is labeled with the [CLS] flag, and the middle or the end of the sentence is labeled with the [SEP] flag. The sequence vectors obtained from the above process are input into the bi-directional Transformer for feature extraction, and finally the sequence vectors with rich semantic features are obtained.

Transformer consists of two parts, Encoder and Decoder, both of which have a 6-layer structure. The input sequence flows into Encoder module after word embedding and position information encoding; the resulting sequence flows into Decoder module after word embedding and position information encoding. Finally, the output sequence of Decoder module is processed by Linear layer and Softmax layer to get the final sequence vector. The flow of Transformer is shown in Fig. 2.

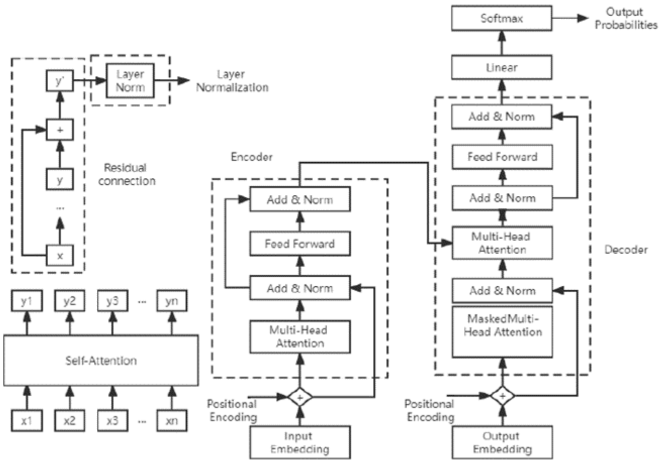

Figure 2 Transformer Flowchart.

In natural language processing, input sentences are encoded by a model and eventually represented as fixed-length vectors. When the input sentence is short, the model is usually able to encode a reasonable vector for it. When the input sentences are long, it is more difficult for the model to learn. In practice, what has a significant impact on the results or what is really needed is usually only part of the data in the input sentences. To address this problem, adding the attention mechanism to the model enables the neural network to focus on a specific subset and thus capture the most effective data. The main role of the attention mechanism in the whole model is to assign weights to the bi-directional semantic feature vectors obtained from the BiLSTM layer, so as to highlight the features that can play an important role in knowledge recognition while ignoring irrelevant features, and the structure of the BiLSTM layer is shown in Figure 3.

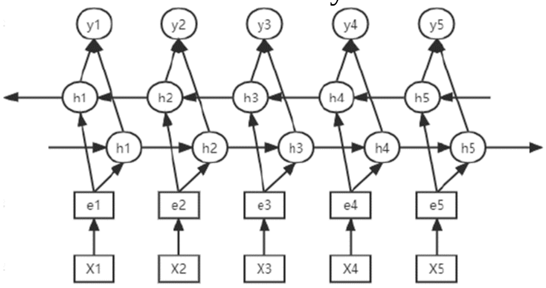

Figure 3 BiLSTM layer structure.

CRF is a sequence annotation model that can make full consideration of the interrelationships of the output labels, and effectively model the existence of constraint relationships for the final predicted labels, so as to achieve the purpose of improving the prediction accuracy. The output of the previous layer of the attention mechanism is used as the input of the CRF layer, which is corrected to obtain the optimal predicted sequence through the relationship of neighboring labels. For example, an I tag preceded by a B tag cannot be an O tag, an I-FIE will not appear after a B-SUB, and so on. An example output is shown in Fig. 4.

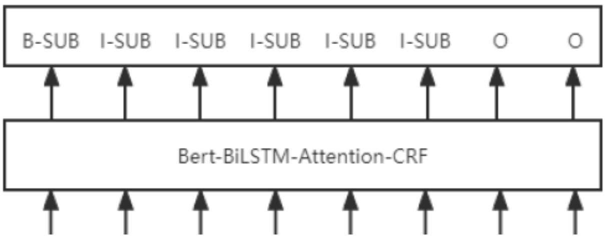

Figure 4 Sample Output Diagram.

4. Experiments and Analysis

4.1. Experimental setup

In this paper, we use the current mainstream deep learning framework Pytorch to build the model and accelerate the computation using NVIDIA2080Ti graphics card. The word embedding dimension is 512, the transformer input dimension is 512, and the batch size input data is 16. The convolutional neural network is fine-tuned, and the learning rate of the convolutional neural network is 1E-4 and the learning rate of recurrent neural network is 1E-4 during the training of the network, and the whole network is trained by the Adam optimizer, and the learning rate is decreased to 0.8 if the evaluation indexes have not been improved in 5 consecutive epochs and the training is terminated after 25 epochs. If the evaluation index is not improved in 5 consecutive epochs, the learning rate will be decreased to 0.8, and the training will be terminated after 25 epochs, which can effectively avoid gradient explosion. dropout regularization method is used to prevent overfitting, which takes the value of 0.1. The test adopts the bundle search method, and the assumption that the associated vocabulary beam size in the vocabulary list is 5.

The data required for the experiment uses textual information related to the ancient city obtained by web crawlers as well as the information required for quantification and organized into vector form as input. The pre-training of the experiments consumes a large amount of machine computing resources and usually takes a long time to train. Currently, most researchers choose to use Bert's pre-training model, and then combine it with their own research content to slightly adjust the model, so as to achieve better results. The relevant parameters of the Bert-BiLSTM-Attention-CRF model are shown in Table 1.

Table 1 Experiment parameters.

Parameter	Value
Max Length	120
Blank Padding	True
Dropout	0.25
Learning Rate	0.001
Batch Size	64

4.2. Results and analysis

In order to comprehensively verify the effectiveness of the model, we first evaluated the proposed quantitative model and analyzed the quantitative results of the six indicators by precision regression, as shown in Figure 5.

As can be seen from the resultant figure 5, the quantization results of the proposed model fluctuate up and down within a reasonable interval of the true value, and the fluctuation trend is very stable, thus indicating that the proposed model has high quantization accuracy and robustness, and can be used in the city size prediction model.

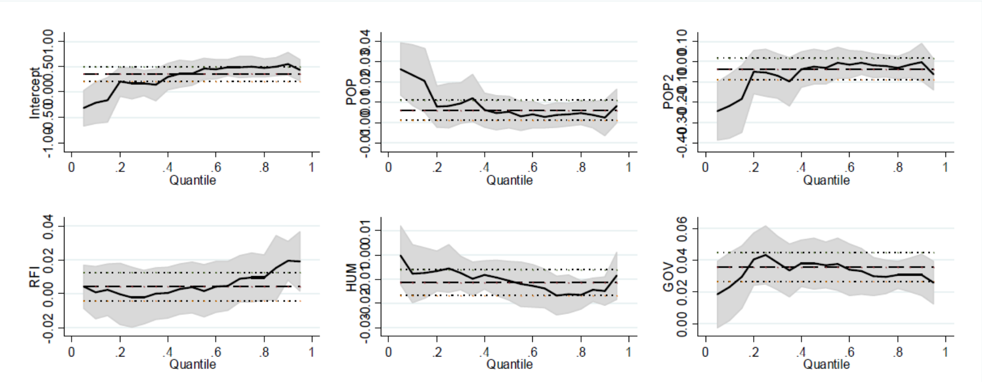

Figure 5 Quantitative results of the six indicators.

In the past, many researchers chose to use Word2Vec to generate word vectors, and then used BiLSTM-CRF model for knowledge extraction, and then added the attention mechanism to improve the BiLSTM-CRF model. Here, these two models are chosen to compare with the model used in this paper, and the results shown in Table 2 show that the model in this paper has a certain improvement in performance compared with the first two models, and has a better performance, which can be put into the next step of use.

Table 2 Experiment results.

Model	Accuracy P	Recall R	F1
Word2Vec-BiLSTM-CRF	0.821	0.842	0.831
BERT-BiLSTM-CRF	0.850	0.863	0.856
Bert-BiLSTM-Attention-CRF	0.876	0.894	0.885

5. Conclusion

Finally, the schematic diagram of the city size predicted from the quantitative results is shown in Fig. 6, which proves the feasibility and validity of the proposed model.

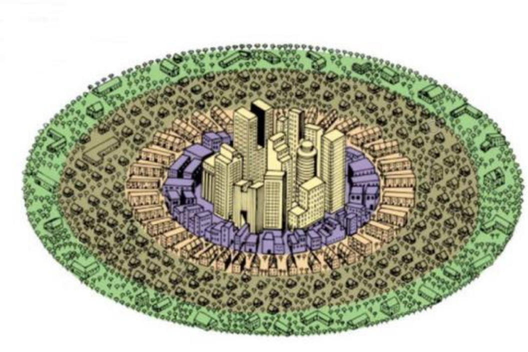

Figure 6 Results of city size projections.

References

[1] Baldi Pierre, Hornik Kurt. Neural networks and principal component analysis: Learning from examples without local minima [J]. Pergamon, 1989, 2(1):53-58

[2] Henderson, J. V., The Sizes and Types of Cities [J]. The American Economic Review, 1974, 64(4): 640-656.

[3] Sullivan, A. M., The General Equilibrium Effects of Congestion Externalities [J]. Journal of Urban Economics, 1983, 14(1): 80-104.

[4] Fu-Chen, Lo, Kamal, S., Growth Pole Strategy and Regional Development Policy: Asian Experiences and Alternative Approaches[R]. Pergamon, 1978.

[5] Bertinelli, L., Black, D., Urbanization and Growth [J]. Journal of Urban Economics, 2004, 56(1): 80-96.

[6] Combes, P. P., Duranton, G., Gobillon, L., et al, The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection [J]. Econometrica, 2012, 80(6): 2543-2594.

[7] Ju-Young Kim, Analysis of City Efficiency Using Urban Network Theory (Korean Language)[J]. The Korea Spatial Planning Review, 2003(38), pp.63-78.

[8] Markowitz H. Portfolio selection [J]. Journal of Finance, 1952, (4):77-91.

[9] Fama E F, French K R. The Cross-Section of Expected Stock Returns[J]. The Journal of Finance, 1992, 47(2): 427-465

[10] CLIFFORD S. ASNESS, TOBIAS J. MOSKOWITZ, LASSE HEJE PEDERSEN. Value and Momentum Everywhere [J]. The Journal of Finance, 2013, 68(3): 929-985

[11] Eugene F. Fama, Kenneth R. French. A five-factor asset pricing model [J]. Journal of Financial Economics, 2015, 116(1):1-22

[12] Zhang M, Yang Z, Liu C, et al. Traditional Chinese Medicine knowledge Service based on Semi-Supervised BERT-BiLSTM-CRF Model[C]//ICSS. : IEEE, 2020: 64-69.

[13] Xu L, Li S, Wang Y, et al. Named Entity Recognition of BERT-BiLSTM-CRF Combined with Self-attention [C]//Lecture Notes in Computer Science: volume 12999 WISA.: Springer, 2021: 556-564.