Learning Ostensive Knowledge of Shaft Parts based on CRF model

Wei Xiong (✉ w_xiongs@usst.edu.cn)

University of Shanghai for Science and Technology <https://orcid.org/0000-0003-1163-2161>

Haolin LI

University of Shanghai for Science and Technology

Research Article

Keywords: Ostensive Knowledge, Knowledge Learning, CRF, Shaft Parts, Dimension Tolerance

Posted Date: March 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1154200/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title page

Learning Ostensive Knowledge of Shaft Parts based on CRF model

Wei Xiong, born in 1982, is currently a PhD candidate at *School of Mechanical Engineering, University of Shanghai for Science and Technology, China*. He received his master degree from *University of Shanghai for Science and Technology, China*, in 2013. His research interests include Intelligent manufacturing and knowledge mining from big data.
Tel: +86-021-55271656; E-mail: w_xiongs@usst.edu.cn

Haolin Li, born in 1961, is currently a professor and a PhD candidate supervisor at *School of Mechanical Engineering, University of Shanghai for Science and Technology, China*. His main research interests include Intelligent manufacturing, precision grinding.
E-mail: haolin61@163.com

Corresponding author: Haolin Li, E-mail: haolin61@163.com

ORIGINAL ARTICLE

Learning Ostensive Knowledge of Shaft Parts based on CRF model

Wei XIONG¹ • Haolin LI¹

Abstract: During the process of mechanical design, there exist a large amount of knowledge and information that needs to be obtained by consulting and referencing past cases. This type of knowledge is called ostensive knowledge in philosophy, and belongs to a typical relational tacit knowledge. On the one hand, ostensive knowledge cannot be expressed through words, numbers, scientific formulas, and coding procedures. On the other hand, knowledge is parasitic in cases and cannot exist independently from them. This paper introduces ostensive knowledge to the mechanical design field for the first time, and points out that the learning model of ostensive knowledge must have the memories of cases and the ability of case-based reasoning. On this basis, this paper focuses on the shaft parts, and proposes an ostensive knowledge learning model for shaft parts based on the conditional random field model. Finally, an experiment is carried out using actual engineering cases to analyze and verify the model, which shows that the model has good memory and reasoning capabilities for shaft parts cases.

Keywords: Ostensive Knowledge; Knowledge Learning; CRF; Shaft Parts; Dimension Tolerance

1. Introduction

During the process of mechanical design, there exist a large amount of knowledge and information that needs to be obtained by consulting and referencing past cases. Take the dimensional tolerance selection of mechanical parts for example, it is almost impossible for engineers to annotate reasonable dimensional tolerances without referring to any successful cases. This type of knowledge is called ostensive knowledge in philosophy, which is the knowledge that can be learned only by pointing to some

object or practice [1]. Collins [1] defines it as a typical relational tacit knowledge, which cannot be expressed through words, numbers, scientific formulas, and coding procedures on the one hand. On the other hand, ostensive knowledge is parasitic in the cases and cannot exist independently from them.

With the advent of the big data era, the use of machine learning technology to learn and extract knowledge from big data has aroused increasing interest among scholars. Google put forward the concept of knowledge graph in 2012, which has driven many scholars to carry out research on knowledge learning and extraction methods [2]. It is generally believed that natural text data contains a large amount of knowledge, so the natural text mining has become the research hot spot spontaneously. In order to build an intelligent search engine based on text semantics, the knowledge graph was constructed to a knowledge network which interconnected by "entity-relation-entity" triples [3]. So the knowledge learning and extraction based on text mining have been segmented into two parts which are named entity recognition and relationship recognition. In the early time, many researches mainly focused on the traditional rule methods [4-5]. In recent years, with the development and maturity of artificial intelligence technology, many artificial intelligence technologies have been applied to the named entities and relations recognition, which include maximum entropy Markov model(MEMM) [6-7], conditional random field(CRF) [8-10], convolutional neural network(CNN) [11-13], recurrent neural network(RNN) [14-16] and its improvement long short term memory(LSTM) model [17-19], as well as LSTM-CRF [20-21] model based on the combination of LSTM and conditional random field, etc.

In mechanical design, Cheong et al. [22] from AUTODESK proposed a method to automatically extract function knowledge from natural language text. Their

✉ Haolin LI
haolin61@163.com

¹ School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

method uses syntactic rules to extract subject-verb-object triples from parsed text, and then classify the triples as artifact-function-energy flow knowledge. Based on this triples Cheong et al. [23] also compose a knowledge graph to extract the structure knowledge of mechanical systems. Siddharth et al. [24] propose a large, scalable engineering knowledge graph, comprising sets of real-world engineering “facts” as “entity-relationship-entity” triples that are found in the patent database.

However, because the ostensive knowledge is parasitic in the cases, it cannot exist independently from the cases. Therefore, the structure of triples cannot learn and extract the ostensive knowledge from documents. For example, the tolerance of the hole must be determined with reference to the specific part cases, and cannot simply be represented by the triple structure of “sth-is-sth”. Therefore, it is necessary to study an ostensive knowledge learning and extraction method based on the mechanical part cases, which is widely used during the mechanical design process. Related reports in the literature, however, rarely talked about the learning and extraction of ostensive knowledge. Wen et al. [25] used conditional random fields to identify semantic information of closed loops from two-dimensional drawings of shaft parts. It, in a sense, can be considered as an exploration of ostensive knowledge learning and extraction. However, their purpose is to identify semantics information of shafts like hole, key and out diameter from closed loops in the 2D engineering drawings, and then reconstruct the 3D models.

Based on the characteristics of ostensive knowledge, we point out that the learning model of ostensive knowledge should have two characteristics. Firstly, it must have the memories of cases. Secondly, it must have the ability of case-based reasoning. We find that the conditional random field (CRF) model have both of these characteristics. At the same time, we also observe that shaft parts have a relatively simple structure which called linear-chain structure. Therefore, the linear-chain structure of shaft parts can be easily applied to the linear-chain CRF model which is the simplest CRF structure. So, in this paper, we choose shaft parts to propose an ostensive knowledge learning method based on CRF model, and the ostensive knowledge of shaft parts learnt in this paper is the out diameter tolerances.

The rest of the paper is organized as follows. An introduction of CRFs model is presented in section 2, with the history, characteristics and its definition. The linear-chain structure and out diameter tolerance selection of shaft parts is then described in section 3. Our method is introduced in section 4 which include graph representation,

feature sets composition and parameter estimation and prediction. Section 5 analyses the experimental results of implementation from real cases. Finally, a summary and outlook of our work is presented in Section 7.

2. Conditional random field (CRF) model

Conditional random fields are developed from hidden Markov model(HMM) and maximum entropy Markov model(MEMM), all of which are probabilistic framework for labeling and segmenting sequential data [26]. The MEMM uses the discriminative model to improve the generative model used in HMM which needs to enumerate all possible observation sequences to define a joint probability over observation and label sequence. This improvement makes the MEMM very easy for engineering applications. Furthermore, CRF improves the MEMM by adding a single exponential model for the joint probability of the entire sequence of labels given the observation sequence, while the MEMM uses per-state exponential models for the conditional probabilities of next states given the current state [8]. Therefore, on the one hand, by adding the joint probability of the entire sequence of labels given the observation sequence in the CRF, the weights of different features at different states can be traded off against each other, which can improve the ability of case-based reasoning. On the other hand, the training of the joint probability can make the model to have the memories of cases.

Definition of CRF [8]: Let $G=(V,E)$ be a graph such that $Y=(Y_v)_{v \in V}$, so that Y is indexed by the vertices of G . Then (X,Y) is a **conditional random field** in case, when conditioned on X , the random variables Y_v obey the Markov property with respect to the graph: $p(Y_v|X, Y_w, w \neq v) = p(Y_v|X, Y_w, w \sim v)$, where $w \sim v$ means that w and v are neighbor in G .

Fig. 1 Structure of linear-chain CRF

Fig. 1 shows a graph of linear-chain CRF model structure. The random variables Y in the graph $G=(V,E)$ is a linear chain structure. According to the definition, the linear-chain conditional random field is a distribution $p(y|x)$ that takes the form [27]:

$$p(y|x) = \frac{1}{Z(x)} \exp \left\{ \sum_{k=1}^K \lambda_k f_k(y_t, y_{t-1}, x_t) \right\} \quad (1)$$

where $Z(x)$ is an instance-specific normalization function

$$Z(x) = \sum_y \exp \left\{ \sum_{k=1}^K \lambda_k f_k(y_t, y_{t-1}, x_t) \right\} \quad (2)$$

Where: $\Lambda = \{\lambda_k\} \in \mathbf{R}^K$ is a parameter vector, and $\{f_k(y_t, y_{t-1}, x_t)\}_{k=1}^K$ is a set of real-valued feature functions.

CRFs model provides a powerful framework for representing, learning and reasoning for structure data prediction. Its discriminative model and the calculation of joint probability are very easy to solve engineering problems. It has been widely used in computer vision [28-29], natural language processing [30-31] and web information extraction [32-33], etc.

3. Structure of shaft part and out diameter tolerance selection

The typical mechanical parts can be categorized into shaft-type parts, disc-type parts, box-type parts and gear parts. The characteristic of the shaft-type parts is that their structure is composed of a series cylindrical feature. As shown in **Fig. 2**, **Fig. 2(a)** is a 3D model of typical shaft part, which can be decomposed into 16 cylindrical constructive solid geometry(CSG) [34] according to the boundary planes normal to the axial direction, as shown in **Fig. 2(b)**. **Fig. 2(c)** is an undirected graph composed of the decomposed CSG in linear-chain, in which each node annotates the geometric information of CSG. So, finally, the structure of shaft part can be transformed into an undirected linear-chain graph with the geometric information of CSGs.

Fig. 2 the structure of typical shaft part

a).3D model of typical shaft part, b).feature decomposition of shaft part
c). undirected graph of decomposed features

The selection of outer diameter tolerance of shaft parts is an important task during the design process of shaft parts. In the real design process, the method used by engineers is

analogy [35] which is a comparison of one thing with another thing that has similar features. So the knowledge required for the selection of outer diameter tolerance of shaft parts is a typical ostensive knowledge. The standard tolerance and fitness system have a long history of use in the mechanical engineering. The standard tolerance is categorized into 20 grades, and 28 basic deviations are designed for different types of fitness. The 20 standard tolerances and 28 basic deviations can form 560 tolerance zones. However, not all the tolerance zones will be used in the actual design of shaft parts, the standard also provide 119 general tolerance zones, 59 popular tolerance zones and 13 priority tolerance zones, as shown in **Fig. 2**

Fig. 3 general, popular, priority tolerance zones of shaft parts

4. Ostensive knowledge learning with CRF

In order to utilize the linear-chain CRF model for the linear-chain structured shaft parts, three steps need to be considered, which are graph representation, feature sets and parameter estimation and prediction.

4.1. Graph representation

Fig. 3(c). Then the learning model of ostensive knowledge for shaft parts can be represented as a graph structure of a linear-chain CRF, as shown in **Fig. 4(c)**.

According to the analysis of real engineering cases and the standard tolerance and fitness system, it is found that there are 53 tolerance zones in the standard tolerance and fitness system used for shaft parts. Therefore, the out diameter tolerance selection problem of shaft parts is transformed into a sequence labeling problem, and the

states of output Y is 53.

Fig. 4 CRF graph representation for shaft parts
a).3D model of typical shaft part, b).CSG decomposition of shaft part
c).CRF graph representation

4.2. Feature sets

The most significant advantage of CRF is that its input data can integrate rich features, and these features can participate in the calculation of the joint probability distribution.

We extract a set of features from the geometric information of the decomposed CSG of shaft parts, which can be divided into three categories: shape features, dimension features and relation features. The total number of features is 70, as list in Table 1

In order to facilitate the construction of the loss function during training, the feature functions use indicator function, take the feature function of ‘if there is a straight hole in the shaft center’ in the Table 1 for example, it takes the form:

$$f_H(y, x) = \begin{cases} 1 & \text{If H is True} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The indicator function is a discrete function. Therefore, the input of features needs to be discretized while applied to the features which input is continuous. Several features in the Table 1 are explained as follow:

- **Basic shapes:** Basic shapes include 7 features, which is cylinder, cone, spline, disk, cone disk, concave arc column, and convex arc column. These are all the basic shape features in shaft parts design. As for shape recognition of CSG, the author adopts the method of deep learning, and its discussion is out of the scope of this paper. The author will publish it separately.
- **If there is a straight hole in the CSG center:** Whether there is a straight hole in the center of the CSG, if there is, the feature function will be equal to 1, if there is not, it will be 0. The total number of features is 1.
- **Length-diameter ratio of CSG(l/d):** It is the ratio between length and diameter of CSG, which can take the form l/d . The ratio is a continuous feature. Therefore, when constructing the feature function, the ratio is discretized into 11 ranges which are $\{(0,0.1], (0.1,0.3], (0.3,0.5], (0.5,0.75], (0.75,1], (1,1.5], (1.5,2], (2,2.5], (2.5,3], (3,4], (4,\infty)\}$. As a result, the total number of features is 11. We observed that there is a certain relation between the length-diameter ratio and the tolerance grade. If the ratio is below 0.5 or above 2.5, the possibility of ‘free tolerance’ will be very high.
- **Diameter ratio of CSG to the maximum diameter of shaft(d/D_{max}):** it is the ratio of the diameter of each CSG to the maximum diameter of the shaft. Similar to the length-diameter ratio, the diameter ratio

Table 1 Feature sets

Categories	Description of features	Notations	Total number of features
Shape	Basic Shapes	BS	7
	If there is a straight hole in the CSG center	H	1
	If there is a taper hole in the CSG center	T	1
	If there exists single slot	S	1
	If there exist multiple slots	MS	1
	If there exit a thread.	Th	1
Dimension	Length-diameter ratio of CSG(l/d)	LD	11
	Diameter ratio of CSG to the maximum diameter of shaft(d/D_{max})	D	9
	Length ratio of CSG to the overall length of shaft(l/L_{OA})	L	11
	Position ratio of CSG in the shaft (Pos/L_{OA})	Pos	11
Relation	Compared to the length on the left	LL	4
	Compared to the length on the Right	RL	4
	Compared to the diameter on the left	LD	4
	Compared to the diameter on the right	RD	4

TOTAL 70

is discretized into 9 ranges. Therefore, the total number of features is 9.

- **Compared to the length on the left:** it is the relation between the left node and the current node in the length direction, which include 4 features: ‘left node longer’, ‘left node shorter’, ‘same length’ and ‘no left node’.

4.3. Parameter estimation and prediction

As shown in Eqs.(1) and Eqs.(2), the target for of CRF model training is to estimate the parameter $\Lambda = \{\lambda_k\} \in \mathbf{R}^K$. However, before training the model, the first step is to construct a loss function. The linear chain CRF uses the maximum likelihood function as the loss function during training, which can be written in the form as [27]:

$$L(\theta) = \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^K \lambda_k f_k(y_t^{(i)}, y_{t-1}^{(i)}, x_t^{(i)}) - \sum_{i=1}^N \log Z(x^{(i)}) - \sum_{k=1}^K \frac{\lambda_k^2}{2\sigma^2} \quad (4)$$

Where: N is the number of shaft parts, T is the number of CSG for each shaft part, K is the total number of features, and the last term is a regularization term to prevent overfitting.

The partial derivatives of the Eqs.(4) are:

$$\frac{\partial L(\theta)}{\partial \lambda_k} = \sum_{i=1}^N \sum_{t=1}^T f_k(y_t^{(i)}, y_{t-1}^{(i)}, x_t^{(i)}) - \sum_{i=1}^N \sum_{t=1}^T \sum_{y, y'} f_k(y, y', x_t^{(i)}) p(y, y' | x^{(i)}) - \sum_{k=1}^K \frac{\lambda_k}{\sigma^2} \quad (5)$$

The partial derivative Eqs.(5) is used to optimize the loss function Eqs.(4). The function $L(\theta)$ is concave, which implies that it has exactly one global optimum. Here, we employ the limited-memory BFGS [36], which only use the derivative of the loss function.

Since $\{f_k(y_t, y_{t-1}, x_t)\}_{k=1}^K$ is an indicator function, the calculation of Eqs.(5) is to calculate $p(y, y' | x^{(i)})$. Using the forward-backward algorithm [37], $p(y, y' | x^{(i)})$ can be calculated by Eqs.(6)

$$p(y_{t-1}, y_t | x) \propto \alpha_{t-1}(y_{t-1}) \Psi_t(y_{t-1}^j, y_t^i, x_t) \beta_t(y_t) \quad (6)$$

Where: α is the forward variables, β is backward variables. For more details about forward and backward variables, see [27]. Ψ is the weight on the transition from state i to state j when the current observation is x, which is defined as:

$$\Psi_t(y_t^j, y_{t-1}^i, x_t) = p(y_t = j | y_{t-1} = i) p(x_t = x | y_t = j) \quad (7)$$

After all the parameters are obtained through training, it is necessary to predict the out diameter tolerance of new shaft parts. The prediction method is to use the Viterbi algorithm [37] to find the most likely tolerance zone label assignment $\mathbf{y}^*$ based

for the input $\mathbf{x}$ of new shaft parts: $\mathbf{y}^* = \arg \max_{\mathbf{y}} p(\mathbf{y} | \mathbf{x})$

5. Case study

5.1. Case application

In the learning model proposed in this paper, the number of features is 70, and the number of output states of the tolerance zone is 53. Consequently, the K of parameters $\Lambda = \{\lambda_k\} \in \mathbf{R}^K$ is 6625.

This article is supported by Shanghai Machine Tool Works Ltd., which provided 46 shaft parts from their actual production for training and testing of our model. The total number of decomposed CSG from all the 46 shaft parts is 1040, which is far less than the number of 6625 parameters. As a result, this data sets cannot cover all the 6625 parameters very well. However, the discriminative model which used in CRF was originally proposed based on small data sets, and the maximum entropy method in CRF can effectively solve the problem of insufficient parameter coverage of a small data set. In fact, even if the number of CSG in the data set exceeds the number of parameters, it is not guaranteed that the data set can cover all parameters. However, the entropy increase method will naturally keep the untrained parameters with a very low entropy value.

Four of the shaft part drawings are shown in Fig. 5. Due to the commercial confidential information involved, we only display the outline of the shaft parts, and all the technical parameters have been erased. It can be found from Fig. 5 that the structure of part a and part b is the same except for difference in size. And part c is similar to part a and part b with the same local structure. Part d is a more special one of the 46 parts, its structure is not similar

Fig. 5 Part of the shaft part drawings

to other parts, and the size is much smaller than the other 46 parts.

In this paper, Matlab is used for training and prediction. We selected parts b, c and d in Fig. 5 as test parts in turn, and conducted three training and test sessions respectively. In each training and test session, we successively selected one part from these three parts as the test part, and the remaining two parts together with the other 43 parts are used for training parts. The results of the three training and test sessions are list in Table 2

Table 2 Test error of three training and testing session

Seq.	Test part	Test Error
1 st	b	0
2 nd	c	0.0455
3 rd	d	0.2667

From the data in Table 2, it can be seen that the structure of part b is exactly the same as part a which is in the training data, as a result, its test error is 0. Part c and part a have the same local structure and are similar in global, therefore, the test error is only 0.0455. The part d is quite different from the parts in the training set, and the test error is 0.2667.

5.2. Case analysis

The critical difference between CRF and MEMM is that CRF consider the joint probability of the entire sequence of labels given the observation sequence which can be mainly reflected by constructing the instance-specific normalization function $Z(\mathbf{x})$ in the Eqs.(2), while the MEMM only consider the relation between the current state and the next state. This difference lies in the fact that CRF treats each input shaft part as a whole, while the MEMM just connect all the shaft parts back and forth in series to calculate the connected CSG locally, which lacking the consideration of the shaft part as whole. This eventually determines that CRF has memories of shaft parts, because the CRF will always take the maximum entropy of the shaft part as a whole when a shaft part is trained. The testing result of part b in Table 2 also verify this.

Whether it is CRF, HMM or MEMM, they all naturally have local memory and powerful reasoning ability. In Fig. 5, part c and part a have the same local structure and are similar in global, while part d is quite different from the parts in the training set. It can be seen from the results in Table 2 that the test error of the shaft part with similar structure is significantly lower than that of the part without similar structure. Meanwhile, although the structure of shaft part has not appeared in the training set, it can still be inferred based on the shaft parts data in the training set. This can be verified by the testing results of part c and part

d in Table 2, which test errors are 0.0455 and 0.2667

6. Conclusions

This paper studies and analyzes the process of acquiring knowledge from cases during the mechanical design process. For the first time, the concept of ostensive knowledge from tacit knowledge is introduced into the mechanical engineering to distinguish the knowledge of triples structure proposed in the knowledge graph. On the one hand, ostensive knowledge cannot be expressed through words, numbers, scientific formulas, and coding procedures. On the other hand, ostensive knowledge is parasitic in the cases and cannot exist independently from them. Therefore, we proposed that the learning model of ostensive knowledge should have two characteristics. Firstly, it must have the memories of cases. Secondly, it must have the ability of case-based reasoning. This paper also proved through case analysis that the CRF-based ostensive knowledge learning model has both of these two characteristics.

The main contributions of this paper are summarized as follows:

1. As far as we know. It is the first time that introduce the concept of ostensive knowledge from tacit knowledge into the mechanical engineering. In addition, two basic requirements for the learning model of ostensive knowledge are proposed, which are memory of cases and reasoning based on cases.
2. Proposed ostensive knowledge learning model of shaft parts based on CRF, and proved through case analysis that the model have both characteristics.

At present, there are only 46 parts in the data set used in this paper, and the number of CSGs obtained after decomposition is only 1040. The small data set has a certain impact on the training accuracy of the model, especially the insufficient training for reasoning ability of the model. Meanwhile, although the input feature set we selected is easy to extract, it does not show a relatively linear relation with the outer diameter tolerance selection of shaft parts. One way to solve this problem is to use deep learning to map the input features to a high dimensional space which showing a linear relation between the input features and the output state. For example, the LSTM-CRF model which based on the combination of CRF and LSTM developed from recurrent neural network. But high-dimensional mapping still faces a problem, that is, the data set need to be large.

7. Declaration

Availability of data and materials

Available from the corresponding author upon request

Competing interests

The authors declare no competing financial interests.

Funding

Supported by Shanghai Science and Technology Commission (Grant No. 19060502300).

Authors' contributions

The author's contributions are as follows: Wei Xiong conceived of the study, analyzed the data and wrote the manuscript; Haolin Li was in charge of the project, discussed the results and revised the manuscript.

Acknowledgements

The authors sincerely thanks to Dr. Xiong DU of Shanghai Machine Tool Works Ltd. for providing cases data.

References

- [1] H Collins. *Tacit and Explicit Knowledge*. The University of Chicago Press Chicago and London, 2010.
- [2] H Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web, IOS Press*, 2017, 8: 489–508.
- [3] G. O. Blog. *Introducing the knowledge graph: thing, not strings*. 2012. <http://googleblog.blogspot.com/2012/05/>.
- [4] RAU L F. Extracting company names from text. *Proceedings The Seventh IEEE Conference on Artificial Intelligence Application*, Miami Beach, FL, USA. 1991:29-32.
- [5] Y Gotoh, S Renals, G Williams. Named entity tagged language models. *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*. 1999, 1: 513-516.
- [6] G L Sun, Y Guan, X L Wang, et al. A maximum entropy Markov model for chunking. *2005 International Conference on Machine Learning and Cybernetics, IEEE*, 2005, 6: 3761-3765.
- [7] R. Li, L. Liu, H. Fu, et al. Application Study of Hidden Markov Model and Maximum Entropy in Text Information Extraction. *Artificial Intelligence and Computational Intelligence*, 2009: 399-407.
- [8] J. Lafferty, A. McCallum, F. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *in Proceedings of the Eighteenth International Conference on Machine Learning*, 2001: 282–289.
- [9] A. Álvarez, C. D. Martínez-Hinarejos, H. Arzelus, et al. Improving the automatic segmentation of subtitles through conditional random field. *Speech Communication*, 2017, 88: 83-95.
- [10] C. Nguyen Viet, Y. Nan, L. Wee Sun, et al. Conditional Random Field with High-order Dependencies for Sequence Labeling and Segmentation. *Journal of Machine Learning Research*, 2014, 15(3): 981–1009.
- [11] R. Collobert, J. Weston, L. Bottou, et al. Natural language processing (almost) from scratch. *Journal of machine learning research*, 2011, 12: 2493–2537.
- [12] K. Xu, Y. Feng, S. Huang, et al. Semantic relation classification via convolutional neural networks with simple negative sampling. in *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, 2015: 536-540.
- [13] J. Qiu, Y. Zhou, Q. Wang, et al. Chinese clinical named entity recognition using residual dilated convolutional neural network with conditional random field. *IEEE transactions on nanobioscience*, 2019, 18(3): 306–315.
- [14] X. Zhang, F. Chen, R. Huang. A Combination of RNN and CNN for Attention-based Relation Classification. *Procedia computer science*, 2018, 131: 911–917.
- [15] Z. Lei, S. Zhao, H. Song, et al. Scene text recognition using residual convolutional recurrent neural network. *Machine Vision and Applications*, 2018, 29(5): 861–871.
- [16] D. Tang, B. Qin, T. Liu. Document modeling with gated recurrent neural network for sentiment classification. in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015: 1422–1432.
- [17] S. Gajendran, M. D. V. Sugumaran. Character level and word level embedding with bidirectional LSTM – Dynamic recurrent neural network for biomedical named entity recognition from literature. *Journal of Biomedical Informatics*, 2020, 112(445): 103609.
- [18] S. Ravuri, A. Stolcke. Recurrent neural network and LSTM models for lexical utterance classification. *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
- [19] S. K. Sahu, A. Anand. Drug-drug interaction extraction from biomedical texts using long short-term memory network. *Journal of biomedical informatics*, 2018, 86: 15–24.
- [20] C. LEE. LSTM-CRF Models for Named Entity Recognition. *IEICE transactions on information and systems*, 2017, 100(4): 882-887.
- [21] B. Tang, X. Wang, J. Yan, et al. Entity recognition in Chinese clinical text using attention-based CNN-LSTM-CRF. *BMC medical informatics and decision making*, 2019, 19(3): 89-97.
- [22] H. Cheong, W. Li, A. Cheung, et al. Automatic extraction of function knowledge from text. *Proceedings of ASME 2015 IDETC/CIE (DETC2015- 47541)*, Boston, MA, 2015, 2A: 1-10.
- [23] H. Cheong, W. Li, F. Iorio. Automated extraction of system structure knowledge from text. *Proceedings of ASME 2016 IDETC/CIE (DETC2016-59551)*, Charlotte, NC, 2016, 2A: 1-10.
- [24] L. Siddharth, L. T. M. Blessing, K. L. Wood, et al. Engineering knowledge graph from patent database. *Journal of Computing and Information Science in Engineering*, 2021, 22(2): 021008.
- [25] Y. Wen, H. Zhang, F. Li, et al. Identifying and constructing elemental parts of shafts based on conditional random fields model. *Computer-Aided Design*, 2015, 62: 10–19.
- [26] H. M. Wallach. Conditional random fields: An introduction. *Technical Reports (CIS)*, 2004: 22.
- [27] C. Sutton, A. McCallum. An introduction to conditional random fields for relational learning. *Introduction to statistical relational learning*, 2006, 2: 93–128.
- [28] F. Liu, G. Lin, C. Shen. CRF learning with CNN features for image segmentation. *Pattern Recognition*, 2015, 48(10): 2983–2992.
- [29] L. Yu, J. Xie, S. Chen. Conditional random field-based image labelling combining features of pixels, segments and regions. *IET computer vision*, 2012, 6(5): 459-467.
- [30] J. Diesner, K. M. Carley. Conditional random fields for entity extraction and ontological text coding. *Computational and*

- Mathematical Organization Theory, 2008, 14(3): 248-262.
- [31] Z. Hao, H. Wang, R. Cai, et al. Product named entity recognition for Chinese query questions based on a skip-chain CRF model. *Neural Computing and Applications*, 2013, 23(2): 371-379.
 - [32] Y. H. Ding, Q. Z. Li, Y. Q. Dong, et al. 2D Correlative-Chain Conditional Random Fields for Semantic Annotation of Web Objects. *Journal of Computer Science and Technology*, 2010, 25(4): 761-770.
 - [33] T. L. Wong, W. Lam. Learning to extract and summarize hot item features from multiple auction web sites. *Knowledge and Information Systems*, 2008, 14(2): 143-160.
 - [34] S. Ghali. Constructive solid geometry. In: *Introduction to Geometric Computing*, Springer, London, 2008: 277-282.
 - [35] Z. Ding. Reasonable selection and correct annotation of tolerances. National Defence Industry Press, 1991. (in Chinese)
 - [36] J. L. Morales. A numerical study of limited memory BFGS methods. *Applied Mathematics Letters*, 2002, 15(4): 481-487.
 - [37] L. R. R. Rabiner. Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*,

1989, 77(2): 257-286.

Biographical notes

Wei Xiong, born in 1982, is currently a PhD candidate at School of Mechanical Engineering, University of Shanghai for Science and Technology, China. He received his master degree from University of Shanghai for Science and Technology, China, in 2013. His research interests include Intelligent manufacturing and knowledge learning from big data.

Tel: +86-021-55271656; E-mail: w_xiongs@usst.edu.cn

Haolin Li, born in 1961, is currently a professor and a PhD candidate supervisor at School of Mechanical Engineering, University of Shanghai for Science and Technology, China. His main research interests include Intelligent manufacturing, precision grinding.

E-mail: haolin61@163.com