ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.2, February 2006, pp.259−266
DOI: 10.1360/jos170259

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于次范畴化的汉语多义动词模糊聚类*

韩习武+, 赵铁军

(哈尔滨工业大学 计算机学院,黑龙江 哈尔滨 150001)

Inducing Fuzzy Classes for Chinese Polysemic Verbs via Subcategorization Information

HAN Xi-Wu+, ZHAO Tie-Jun

(Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: Phn: +86-451-86609036, E-mail: hxw@mtlab.hit.edu.cn, http://mtlab.hit.edu.cn

Han XW, Zhao TJ. Inducing fuzzy classes for Chinese polysemic verbs via subcategorization information. *Journal of Software*, 2006,17(2):259−266. http://www.jos.org.cn/1000-9825/17/259.htm

Abstract: This paper describes the application of Fuzzy *k*-Means, a derivant of k-Means that may assign an item to more than one cluster, in the task of inducing fuzzy classes for Chinese polysemic verbs. The probability distributions over subcategorization frames of 60 Chinese verbs, among which there are 40 polysemic ones and 20 monosemic ones are first acquired, and then these verbs are clustered into fuzzy classes. Evaluation and post-hoc analysis show that a combined measure of purity and pairwise precision can better estimate the clustering performance, and although to a certain extent syntactic behaviors of verbs have their counterparts of meaning components underlying, syntactic behaviors of verbs cannot be easily predicted from a single semantic level, at least for Chinese polysemic verbs.
Key words: Chinese; polysemic verb; subcategrization; fuzzy *k*-means; cluster

摘 要: 描述了应用模糊 *k* 均值方法聚类汉语多义动词的实验,共涉及到 60 个汉语动词,40 个多义词,20 个单义词.首先,自动获取每个动词的次范畴化框架的概率分布,然后,导出这些动词的模糊聚类.结果表明,纯洁度和对精确度的综合量度较好地反映了聚类性能,尽管动词的句法行为在一定程度上体现了深层语义,但汉语动词的句法行为不易从单一的语义层预测出来.
关键词: 汉语;多义动词;次范畴化;模糊 *k* 均值;聚类
中图法分类号: TP391 文献标识码: A

An almost axiomatic linguistic hypothesis is that lexical meanings of a word, to a certain extent, determine its respective syntactic behaviors. And by means of clustering, many researches have explored and supported such connection between meaning components of a verb and its subcategorization distribution, e.g. Ref.[1] for English, and Ref.[2] for German.

Using the Information Bottleneck and nearest neighbor methods, Ref.[2] clustered a set of 78 polysemic and 32 monosemous English verbs into mono-classes, i.e. one verb belongs to only one class. Via a special evaluation

* Supported by the National Natural Science Foundation of China under Grant No.60373101 (国家自然科学基金)
Received 2004-05-01; Accepted 2005-07-28

scheme of checking the resulting mono-classes against a polysemic gold standard, they reported that many polysemic verbs do have some predominating sense in corpus data, and for verbs with flat sense-distributions there are limitations of clustering undisambiguated subcategorization information.

But in Chinese, most common verbs are more or less polysemic. It is impossible and unpractical to assign one of them to an exact class. In this paper, we suggest using Fuzzy *k*-Means[3], a derivant of k-Means that may assign an item to more than one clusters, to induce fuzzy classes for Chinese polysemic verbs. And as far as we know, this is the first attempt to form fuzzy clusters for verbs automatically. We first acquired the subcategorization frames and relevant frequencies for the test verbs from the output of a cascaded HMM parser, then filtered the raw subcategorization frames statistically, and at last clustered the verbs via the final normalized SCF information. Our evaluation and post-hoc linguistic analysis show that a combined measure of purity and pairwise precision can better estimate the clustering performance, and although some specific syntactic behaviors of verbs have their counterparts of meaning components underlying, the connections between pure meanings and syntax can not be simply taken for granted, at least for Chinese polysemic verbs.

In the rest of this paper, we discuss our test verbs and gold standard in Section 1, describe our subcategorization frames for Chinese verbs and data preparation in Section 2, illustrate the Fuzzy k-Means method in Section 3, evaluate and analyze the clustering results in Section 4, and conclude with possible directions for future work in Section 5.

1 Test Verbs and Their Semantic Classes

To our knowledge this is the first work to cluster Chinese verbs via pure subcategorization information, thus we randomly constructed a task-oriented set of 60 test verbs (See Table 1). For each of the verbs there are 400 citations as a minimum and 3600 citations as a maximum from the balanced corpus of People Daily 1998, and among them there are 40 polysemic ones (the first 2 columns in Table 1) and 20 monosemic ones (the last column in Table 1) for the purpose of comparison. For possible major meanings of the verbs please refer to Table 2, where we give the first three English labels from Hownet V. 1.5.

Table 1 The test verbs and their semantic classes

Verbs	Thesaurus classes	Verbs	Thesaurus classes	Verbs	Thesaurus classes
翻	FA,FB,HD,HG	吃	FC,HB,HE,JE	帮助	HI
给	HI,JE	开	FA,HB,HC,HE,HF,HG,IA,IB,IE,IG	保持	JD
控制	GB,HC,JE	拉	FA,HE,HH,HI,JE	采访	HJ
表现	HJ,JD	落	HE,ID,IE,JA,JD,JE	出版	HD
跑	FB,HJ,IA	上	FA,HG,HH,HI,HJ,IH,JD,JE	促进	JE
超过	JB,JE	拿	FA,HC,HI,HM	集中	IE
决定	GA,HC,IE,JE	打	FA,HB,HD,HE,HG,HH,HI,HJ,IC	加强	IH
落实	GA,IE	起	HC,HG,HJ,IG,IH,JD	调查	HC
处理	HC,HE	去	HF,HG,HH,HJ,ID,JD	进行	IG
说明	HG,HI,JA	放	FA,HB,HD,HF,HI,HJ,HM,IB,ID,IH	贯彻	IE
听	FC,HC,HI	下	FA,HB,HC,HH,HJ,IB	关心	GB
代表	HI,JA	压	FA,GB,HN,ID,IH,JE	负责	HJ
等	HI,HJ,JB	干	GB,HC,HJ,HN,IB,IH,JE	分析	HG
结合	HJ,IE	过	HF,HJ,IE,JB	体现	JA
换	HE,HI,IH	看	FC,GB,HG,HI,HM	站	FB
发表	HC,HI,HJ,JD	作	HC,HG,HJ,IA,IC,JA,JD	通过	HF
反映	HC,JA	做	HC,HD,HG,HH,HJ,IE	投资	HE
支持	GB,HI,JD	走	FB,HF,HI,IA,ID	希望	GB
分	HC,HJ	抓	FA,HJ,HM	战	HB
准备	GB,HJ	发	GB,HB,HC,JD	学习	HG

Table 2 The first 3 Hownet English labels for the testing verbs

Verbs	Hownet English meanings	Verbs	Hownet English meanings	Verbs	Hownet English meanings
翻	LookFor, add, cross	吃	eat, depend, destroy	帮助	help
给	CauseToDo, give, provide	开	CauseToLive, StateChange, TurnOn	保持	keep
控制	control	拉	attract, attract, attract	采访	investigate
表现	ShowOff, show	落	MoveItDown, ResultFro, decline	出版	publish
跑	engage, flee, run	上	GoForward, GoUp, LeaveFor	促进	urge
超过	surpass	拿	CauseToDo, MakeTrouble, believe	集中	assemble, control, gather
决定	ResultIn, decide	打	TakeOutOfWater, associate, beat	加强	MakeBetter
落实	AtEase, decide, fulfil	起	PickOut, appear, arise	调查	investigate
处理	handle, remove, sell	去	RegardAs, fro, go	进行	GoForward, GoOn, conduct
说明	explain	放	abandon, add, adjust	贯彻	conduct
听	listen, obey, perception	下	GiveBirth, GoDown, GoOut	关心	PayAttention
代表	mean, replace	压	approach, delay, frighten	负责	manage
等	equal, wait	干	IllTreat, do, fight	分析	analyze
结合	GetMarried, merge	过	cross, pass, surpass	体现	mean
换	exchange, exchange, replace	看	TakeCare, cure, depend	站	CeaseSelfMove, stand
发表	publish	作	RegardAs, compile, do	通过	accept, cross
反映	mean, tell	做	RegardAs, be, compile	投资	provide
支持	endorse, endure	走	SelfMoveInManner, function, leak	希望	expect
分	distinguish, issue, separate	抓	PayAttention, catch, catch	战	fight, fight
准备	plan, prepare	发	announce, become, express	学习	study

The gold standard of semantic classes is generally obtained from Chinese Synonym Thesaurus[4] with references from other Chinese dictionaries, such as Ref.[5] and the Modern Chinese Classification Dictionary. Words in the thesaurus are organized hierarchically according to their meanings, e.g. classes beginning with letter F denoting acts of body, while class FA denoting acts of upper limbs, and class FB denoting acts of lower limbs; classes beginning with letter H denoting social activities, while class HD denoting production activities, and class HH denoting acts in sports (See Table 1). It is one of the most applied lexical resources in Chinese NLP. However, the thesaurus doesn't define a predominant sense or class for a polysemic verb. Test verbs in our gold standard are grouped into 27 classes, which in turn fall in 5 hypernym sets. Thus our gold standard is purely based on meanings, and it differs from a Levin style standard, which is formed in a syntactic-semantic way.

2 Our Subcategorization Frames

According to Ref.[6], SCF for Chinese verbs can be described as a quintuple grammar $\langle \boldsymbol{V}, \boldsymbol{T}_A, \boldsymbol{N}_A, \boldsymbol{P}_A, \boldsymbol{C}_L \rangle$, in which

i) $\boldsymbol{V}$ is a set of verbs capable of filling in the predicate slot;

ii) $\boldsymbol{T}_A$ is a set of argument types, and $\boldsymbol{T}_A$={NP, VP, QP, BP, PP, BAP, BIP, TP, MP, JP, S}(for respective definitions see Table 3);

iii) $\boldsymbol{N}_A$ is a set of numbers of argument slots;

iv) $\boldsymbol{P}_A$ is the set of positions for argument slots;

v) $\boldsymbol{C}_L$ is the set of constant labels that may be added to some particular SCF, and $\boldsymbol{C}_L$={"zhe"(着), "le"(了), "guo4"(过), "mei2"(没), "bu4"(不)}, where the first three furnish SCF aspects and the last two offer negation options.

Since this paper mainly focuses on more appropriate classification of polysemic verbs, we leave out the $\boldsymbol{C}_L$ information that is not fine-grained enough.

Table 3 Types of argument in SCF

T_A	Definition	T_A	Definition
NP	Nominal phrase	BAP	Phrase headed by "ba3" (把)
VP	Verbal phrase	BIP	Phrase headed by "bei4" (被) or other characters with passive sense
QP	Tendency verbal complement	TP	Temporal phrase
BP	Resulting verbal complement	MP	Quantifier complement
PP	Positional phrase	JP	Adjective or adverb or "de" (得) headed complement
S	Clause or sentence		

We obtained our SCF data using the subcategorization acquisition system described in Ref.[7], which employs a cascaded HMM parser and a predefined SCF set of 137 frames, a mixed set of 126 auto-acquired and 11 hand-constructed ones. The resulting lexicon of the People Daily 1998 Corpus was evaluated against manually processed data before and after filtering of a binomial hypothesis test. Before filtering, the type precision is 58.65% and the type recall is 61%, while after that the former increases to 70.6% and the latter declines to 54.3%. Since the filtered data outperformed the unfiltered, we used the normalized filtered SCF information for our clustering experiment.

3 Clustering Methodology

In order to uncover an inherent natural structure and generalize with this knowledge, researchers almost always use clustering as a standard procedure in multivariate data analysis. In our case, the data objects are represented by verbs, and the relevant features are realized by a probability distribution of SCF for a concerned verb.

Previous clustering methods used in nature language processing tend to assign one object to an exact class, e.g. Refs.[1,2]. In fact, the boundaries between language units, at least between those of polysemic verbs, are often somewhat indeterminate, so we used Fuzzy k-Means for inducing fuzzy classes for Chinese polysemic verbs. Fuzzy k-means minimizes the within-class sum square errors functional under the following conditions[3]:

$$\sum_{k=1}^{k} m_{ik} = 1,\ i = 1,2,...,n; \qquad \sum_{i=1}^{n} m_{ik} > 0,\ k = 1,2,...,c \tag{1}$$

where m_{ik} is the membership of the ith item in class k, $m_{ik}\in[0,1]$, i=1,2,...,n; k=1,...,c; and n is the number of data, c is the number of classes. The objective function is defined as the following:

$$J = \sum_{i=1}^{n}\sum_{k=1}^{c} m_{ik}^{\phi} d^2(x_i c_k) \tag{2}$$

where n is the number of data, c is the number of classes, c_k is the vector representing the centroid of class k, x_i is the vector representing the ith item and $d^2(x_i,c_k)$ is the squared distance between x_i and c_k according to a chosen definition of distance, which for simplicity is further denoted by d^2_{ik}. And in this paper, we choose Euclidean distance. ϕ is the fuzzy exponent and ranges from $(1,\infty)$. It determines the degree of fuzziness of the final solution, which is the degree of overlap between groups. With ϕ=1, the solution is a hard partition. As ϕ approaches infinity, the solution approaches its highest degree of fuzziness.

The minimization of the objective function $\boldsymbol{J}$ provide the solution for the membership function[3]:

$$m_{ik} = d_{ik}^{2/(\phi-1)} \Big/ \sum_{j=1}^{c} d_{ij}^{2/(\phi-1)},\ i = 1,2,...,n; k, j = 1,...,c \tag{3}$$

$$c_k = \sum_{i=1}^{n} m_{ik}^{\phi} x_i \Big/ \sum_{i=1}^{n} m_{ik}^{\phi},\ k = 1,2,...,c \tag{4}$$

In this paper we choose the number of classes k=25,26,27,28 and 29, choose ϕ=1.3 for the fuzziness exponent, 0.001 for the stopping criterion, and Euclidean distance to describe the variable-space. Then the fuzzy *k*-means algorithm in use is briefly described as follows:

(1) Initialize with random memberships $M^{(0)}$

(2) At iteration it =1,2,3, …

(re) calculate $C=C^{(it)}$ using Eq.(4) and $M^{(it-1)}$

(3) Re-calculate $M=M^{(it)}$ using Eq.(3) and $C^{(it)}$.

(4) Compare $M^{(it)}$ to $M^{(it-1)}$. If $||M^{(it)}-M^{(it-1)}||<0.001$, then stop; otherwise return to step (2).

The input data to clustering was obtained from the automatically acquired SCF lexicon for our 60 test verbs (see Section 1). The counts and frequencies were extracted from SCF distributions statistically filtered by means of binomial hypothesis testing (see Section 2). Besides the memberships, the output data also include the Confusion Index, which is a measure of the degree of class overlap in the respective attribute space. It is calculated as:

$$\boldsymbol{CI}=1-(\text{the biggest membership}-\text{the second biggest membership}) \tag{5}$$

We set 0.1 as the threshold for the membership of an item in a certain cluster and also 0.1 for the Confusion Index.

4 Clustering Evaluation and Analysis

4.1 Evaluation methods

There are many different strategies for the evaluation of clustering results against a gold standard. In order to avoid bias towards specific numbers of classes or class sizes and cover as much information as possible, we applied the purity, the pairwise precision, and a combined measure. In all the three methods an output value will be a normalized one, i.e. ranges from (0,1).

Our first measure, the purity, is a global measure, which evaluates the mean precision of the clusters, weighted according to the cluster size[8]. We form our own equation as follows:

$$Purity=\frac{\sum_i |A_i|+\sum_j |B_j|}{2}\times\sum_i\sum_k\frac{|AB_{ij}|}{||A_i|-|B_j|+1|} \tag{6}$$

such that it takes into account the difference between the sizes of counter classes in the clustering results and the gold standard. Here, $0<\text{purity}\le 1$, $|A_{\boldsymbol{i}}|$ is the number of items of the ***i***th class in the clustering results, $|B_{\boldsymbol{j}}|$ is the number of items of the ***j***th class in the gold standard, and $AB_{\boldsymbol{ij}}=\text{Max}j|A_i\cap B_j|$.

Our second measure, the pairwise precision:

$$PP=1/k\sum_{i=1}^{k}\frac{number\_of\_correct\_pairs\_in\_K_i}{number\_of\_pairs\_in\_K_i} \tag{7}$$

is a concise version of Ref.[1]. We adjust the calculation so as to avoid the bias of large cluster sizes and normalize the output value.

Our third measure, the combined one, is intended for a comprehensive evaluation, because in the experiment we found that the relationship between the former two measures is nonlinear and thus a straightforward conclusion is difficult to be made. The calculation is simply performed in the way like that of F measure as follows.

$$\boldsymbol{F}=2\times Purity\times PP/(Purity+PP) \tag{8}$$

4.2 Clustering evaluation

Since our gold standard is purely semantic, in evaluation we suppose that a polysemic verb will appear in more than one clusters, and a monosemic verb will appear in only one cluster, and verbs sharing same meanings tend to appear in the same cluster. We chose clustering number k=26,27 and 28 to describe in this paper because the number in the standard is 27 and these results are also the best. For baseline we chose the best one from 10 randomly formed fuzzy sets, clustering numbers of which ranges from (25,29). The complete set and the singleton set were also evaluated. Table 4 shows the results against the standard. It demonstrates that the Fuzzy k-means clustering method performs much better than our best random clustering baseline and the result with k=26 contains the most reasonable clusters. Purity is an intuitively more plausible measure in that it assigns low values to both the complete and our gold standard.

Table 4 Evaluation results

k	*Purity*	*PP*	*F*
(Random) 28	0.101 3	0.210 3	0.136 7
(Complete) 1	0.003 8	0.378 0	0.007 5
(Singleton) 60	0.069 0	0.000 0	0.000 0
(Clustering) 26*	0.165 4	0.415 4	0.236 6
(Clustering) 27	0.134 6	0.362 7	0.196 2
(Clustering) 28	0.155 7	0.298 8	0.204 8

The evaluation results in Table 4 demonstrate that the Fuzzy k-means clustering method performs much better than our best random clustering baseline. Purity is an intuitively more plausible measure in that it assigns low values to both the complete and singleton set, while the relationship between the actual clustering performance and the pairwise precision seems far from being linear because it assigns a large value to the complete set. Thus we know that the pairwise precision itself gives no linear estimation of clustering performances though it does give some reasonable information (for k=26 the clustering result outperforms the complete set). Meanwhile, the combined measure ***F*** takes more aspects into consideration and offers us a comprehensive account.

4.3 Analysis for polysemy

For k=26, we performed both quantitative and qualitative analysis for our syntactically recovered polysemic verbs, i.e. those having been assigned to more than one clusters or those whose Confusion Index is high enough, since we have set 0.1 as threshold for the membership of an item in a certain cluster and also 0.1 for the Confusion Index (Section 3). We calculated the polysemy precision and recall in the way showed by formulas 8 and 9.

Precision=the number of polysemic verbs occurring both in the results and the standard/
the number of polysemic verbs in the results (9)

Recall=the number of polysemic verbs occurring both in the results and the standard/
the number of polysemic verbs in the standard (10)

Thus, for clustering memberships and the Confusion Index, we got quantitative information against our gold standard as in Table 5. The information shows that though the polysemy precision is comparatively plausible, the recall and F measure are too low. This suggests that the connections between pure meanings and syntax should not be simply taken for granted, at least for Chinese polysemic verbs. Our qualitative analysis offers insights that syntactic behaviors of verbs cannot be easily predicted from a single semantic level.

Table 5 Quantitative polysemy analysis

	Precision (%)	Recall (%)	*F*
Membership	85.7	15	25.53
CI	64.7	27.5	38.6

In the clustering output, we have 17 verbs with CI larger than 0.1, and 6 of them are monosemous in the gold standard. The recovered 11 polysemic verbs support the semantic syntax connection, while the 6 monosemous ones seem opposing it. However, we observed in our corpus that 4 of the 6 take even more SCF types than some recovered polysemic verbs.

Of the 29 polysemic verbs that haven't been recovered at all, there are 14 ones generally falling in their hypernym sets, i.e. most of the first letters of their semantic denotations are the same (See Table 6). This may indicate that our gold standard is little too fine-grained, because trivial meaning components usually don't result in syntactic differences.

For the other 15 verbs, there are 8 ones (See Table 7) whose polysemic components more often than not appear in similar syntactic patterns, at least in our corpus. And this also accounts for the semantic denotations with different beginning letters in Table 6.

Table 6 Verbs generally falling in their hyponym sets

Verbs	Thesaurus classes	Verbs	Thesaurus classes
超过	JB,JE	打	FA,HB,HD,HE,HG,HH,HI,HJ,IC
处理	HC,HE	发表	HC,HI,HJ,JD
分	HC,HJ	看	FC,GB,HG,HI,HM
换	HE,HI,IH	拉	FA,HE,HH,HI,JE
拿	FA,HC,HI,HM	做	HC,HD,HG,HH,HJ,IE
听	FC,HC,HI	去	HF,HG,HH,HJ,ID,JD
说明	HG,HI,JA	抓	FA,HJ,HM

Table 7 Verbs appearing in similar syntactic patterns

Verbs	Thesaurus classes	Verbs	Thesaurus classes
翻	FA,FB,HD,HG	结合	HJ,IE
吃	FC,HB,HE,JE	走	FB,HF,HI,IA,ID
反映	HC,JA	准备	GB,HJ
过	HF,HJ,IE,JB	落实	GA,IE

For instance, "翻" in all its four polysemes means "look for" or "turn over" in English, tending to take SCFs like NP V NP, NP V, and NP V PP; "吃" in its first polyseme means "eat", in its second and third means "depend on", and in its last means "destroy", mostly taking SCFs like NP V NP, and NP V QP NP; "反映" in its first polyseme means "mean", and in its second means "tell", often appearing in SCFs like NP v NP, NP v NP NP, and NP v NP VP.

For another two, "放" and "决定", SCFs for their polysemes are often alternatives of each other. And the errors in the rest five verbs generally come from noisy data, and bad performance of our SCF acquisition system.

5 Conclusion

This paper has described a novel approach to induce fuzzy classification for Chinese verbs. To our knowledge it is the first attempt to cluster language items automatically in a fuzzy way, and the first experiment of clustering polysemic Chinese verbs. This involved applying the Fuzzy *k*-means method to cluster polysemic SCF distributions extracted from corpus data using the system of Ref.[7]. A task-oriented evaluation scheme is introduced which enables us to investigate the connections between polysemes and our clustering classification.

Our investigation reveals that the combined measure of purity and pairwise precision can better estimate the clustering performance, and although to a certain extent syntactic behaviors of verbs have their counterparts of meaning components underlying, syntactic behaviors of verbs cannot be easily predicted from a single semantic level, at least for Chinese polysemic verbs.

In the future, we plan to work on improving the accuracy of subcategorization acquisition for Chinese verbs, constructing better polysemic gold standards, and extending our experiment of fuzzy clustering to larger data and other NLP fields.

Acknowledgement In the experiment we used FuzMe version 3.0 as our clustering tool, which is a publication of the Australian Centre for Precision Agriculture (ACPA).

References:

[1] Korhonen A, Krymolowski Y, Marx Z. Clustering polysemic subcategorization frame distributions semantically. In: Proc. of the 41st Annual Meeting of the Association for Computational Linguistics. 2003. 64−71.

[2] Walde SS, Brew C. Inducing German semantic verb classes from purely syntactic subcategorization information. In: Proc. of the 40st Annual Meeting of the Association for Computational Linguistics. 2002. 223−230.

[3] Minasny B, McBratney AB. FuzME version 3.0. Australian centre for precision agriculture. The University of Sydney, 2002.

[4] Mei JJ, *et al*. Chinese Synonym Thesaurus. Shanghai: Shanghai Lexicology Press, 1983 (in Chinese).

[5] Lin XG, *et al*. Modern Chinese Verb Dictionary. Beijing: Beijing Language College Press, 1994 (in Chinese)

[6] Han XW, Zhao TJ, Yang MY. FML-Based SCF predefinition learning for Chinese verbs, In: Proc. of the IJCNLP 2004. Sanya, 2004. 115−122.

[7] Han XW, Zhao TJ, Qi HL, Yu H. Subcategorization acquisition and evaluation for Chinese verbs. In: Proc. of the COLING 2004. Geneva, 2004. 723−728.

[8] Stevenson S, Joanis E. Semi-Supervised verb-class discovery using noisy features. In: Proc. of the CoNLL 2003. Edmonton, 2003. 95−102.

附中文参考文献:

[4] 梅家驹,等.同义词词林.上海:上海辞书出版社,1983.

[5] 林杏光,等.现代汉语动词词典.北京:北京语言文化大学出版社,1994.

HAN Xi-Wu was born in 1972. He is a Ph.D. candidate at the Harbin Institute of Technology. His current research areas are NLP, AI and machine learning.

ZHAO Tie-Jun was born in 1962. He is a professor at the Harbin Institute of Technology and a CCF senior member. His research areas are NLP, machine translation and AI.