The Extraction of Trajectories from Real Texts Based on Linear Classification

Hanjing Li
Harbin Institute of Technology MOE-MS Key Laboratory of Natural Language Processing and Speech, Harbin, 150001

hjlee@mtlab.hit.edu.cn

Tiejun Zhao
Harbin Institute of Technology MOE-MS Key Laboratory of Natural Language Processing and Speech, Harbin, 150001

tjzhao@mtlab.hit.edu.cn

Sheng Li
Harbin Institute of Technology MOE-MS Key Laboratory of Natural Language Processing and Speech, Harbin, 150001

lisheng@hit.edu.cn

Jiyuan Zhao
Harbin Institute of Technology MOE-MS Key Laboratory of Natural Language Processing and Speech, Harbin, 150001

jyzhao@mtlab.hit.edu.cn

Abstract

Text-to-scene conversion systems need to share the spatial descriptions between natural language and the 3D scene. Such applications are the ideal candidates for the extraction of spatial relations from free texts, in which the extraction to trajectories that are focus objects in spatial descriptions is an essential problem. We present an analysis of how the space relations are described in Chinese. Based on this study, we propose a method where the extraction of trajectories is modeled as a binary classification problem and resolved based on a linear classifier with syntactic features. Moreover, experimental results are analyzed in detail to demonstrate the effectiveness of the linear classifier to the extraction problem of the trajectory concept.

1 Introduction

Text-to-scene conversion systems are systems where a static 3D scene or a cartoon is generated from text. How to extract spatial relations from texts is an essential problem for sharing the spatial information between natural language and 3D scene in such systems. A spatial relation is composed of a spatial expression and a trajectory according to cognitive linguistics. Furthermore, a trajectory is the focus object to be described in a spatial relation. Therefore, it is an essential problem to extract trajectories from a text for a text-to-scene system.

Until now, there are some researches done to resolve the extraction of the trajectory. It is resolved based on dependency relationship in *Wordseye* system (Coyne and Sproat, 2001), and case grammar in *SWAN* system (Lu and Zhang, 2002). But two points must be clarified in these two systems. First, the input of them was restricted to a simple subset of English or Chinese consisting of simple sentences without complex grammar phenomena such as clauses, ellipsis of the object or subject, and spatial focus shifts. Also, only one or zero spatial relation is and should be described clearly in one simple sentence, such as *The store is under the large willow*. (Coyne and Sproat, 2001). Second, they do not refer to the term trajectory and tasks of the extraction of the spatial relation and the trajectory. Although the *Carsim* system, which is another text-to-scene system, refers to the term trajectory and the task of the extraction of the trajectory on IDA method (Johansson, 2006), a trajectory means the route of every event in the animation which does not refer to the same meaning as in this paper.

The task considered here is to acquire the corresponding trajectories of a spatial expression at the text level, and then every trajectory and the spatial expression can compose a special spatial relation. For example, see the fable in Figure 1, "鹰" (the eagle) and "乌龟" (the tortoise) that should be extracted, match exactly with the spatial expression of "到高空" (in the clouds) separately and result in two spatial relations, "鹰到高空" (the eagle in the clouds) and "乌龟在高空" (the tortoise in the clouds) that can be acquired exactly. In the paper, we focus on the extraction from real texts. Based on an analysis of spatial descriptive Chinese language, we have developed a binary classifier that can identify trajectories and compose spatial relations.
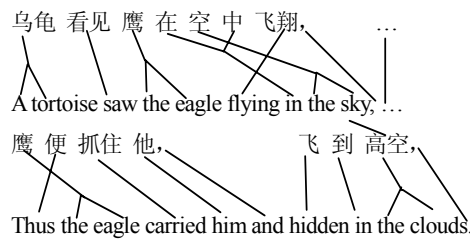
乌龟 看见 鹰 在 空 中 飞翔,   …

A tortoise saw the eagle flying in the sky, …

鹰 便 抓住 他,   飞 到 高空,

Thus the eagle carried him and hidden in the clouds.

Figure 1: 'The Eagle and the Tortoise' in "Aesop's Fables"

To study the characteristics of descriptive spatial language, we analyzed the descriptions by cataloguing the relations between trajectories and their corresponding spatial expressions respectively on Chinese grammar. The combination style of a trajectory and the corresponding spatial expression is referred to as a descriptive strategy including three types of the mapping, the spatial focus shift and the syntactic location.

We propose a set of computational mechanisms that correspond to the most commonly used descriptive strategies. The related terms are formally defined first. Also the extraction of trajectories is modeled as a binary classification problem. Based on those formalizations, we deal with the extraction with Winnow being a linear classification algorithm on syntactic features at the text level.

To evaluate the method, we created an evaluation corpus of "Aesop's Fables", and used three evaluation measures including precision P (the percentage of all correct results in the all results identified by the method), recall R (the percentage of all correct results identified by the method in all really correct results), and F-score ($F=(P*R*2/(P+R))$).

The method has three advantages. First, it showed that the shallow syntactic features are effective for the extraction of the cognitive concept of a trajectory. Second, it did work in a range of linguistic phenomena that make use of really complex compositions of spatial semantics. Last, it was effective without a parser which is domain dependence for Chinese.

The paper is organized as follows. In Sect. 2, we analyze descriptive strategies in Chinese, and then give the definitions of terms. Also, we review SNoW. In Sect. 3, we first formalize the problem considered in this paper, and then we discuss the linear classifier to extract the trajectory. Experimental results are given in Sect. 4. Finally, we make some concluding remarks in Sect. 5.

2 The Resource Creation

2.1 The Definitions of Relevant Terms

We define formally terms including trajectory, landmark, spatial expression and spatial description. Trajectory and landmark are all from Langacker's Cognitive Grammar Framework (Langacker, 1987). Trajectory *TR* enclosed by a particular bound presents the focus object represented in a spatial description. The bound is called Landmark *LM*.

A spatial expression *SE* limits the location of the trajectory, and is formalized as ([*Pre*], *Loc*, *LM*), in which [] means that the part can be omitted, *Pre* denotes preposition, and *Loc* denotes localizer (explained in (Zou, 1989)). For example "在空中" (in the sky) is formalized as (在, 中, 空) (in, ,sky).

A Spatial description describes the spatial configuration of a trajectory with respect to a landmark. In this paper, we define spatial description formally as a spatial relation. A spatial relation is composed of two parts including a spatial expression and trajectory and taken as a binary (([*Pre*], *Loc*, *LM*), *TR*), which represents relative or absolute position and orientation. For example, the description of "鹰在空中" (the eagle in the sky) in Figure 1 is formalized as ((在, 中, 空), 鹰) ((in, , the sky), the eagle).

2.2 The Corpus Creation

As there are no publicly available evaluation corpora, we selected "Aesop's Fables" as a corpus for this task where 10 volumes and 434 texts in Chi-

nese are available at http://www.white-collar.net/child/fable/yisuo/index.html.

The "Aesop's Fables" was annotated with spatial expressions, landmarks and trajectories successively, and turned into the evaluation database for the extraction of spatial relations. The texts annotated separately by every annotator were merged into the final corpus according to the all-passed rule. There were six annotators including 2 undergraduate students, 3 graduate students and 1 PhD student, who all take part in the project "Research on Visualization of Spatial Descriptions in text".

The "Aesop's Fables" was first segmented and part-of-speech (POS) annotated with the tool described in (Lv, 2003), anaphora resolved by hand, (for example, [AR乌龟/ng] is an anaphora resolution of "他/r" (it) in Figure 2), and then called initial corpus. Afterwards, the initial corpus was annotated on spatial relations by hand with the rules presented following.

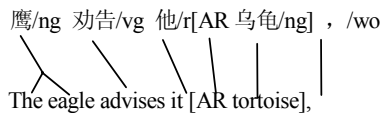鹰/ng 劝告/vg 他/r[AR 乌龟/ng] ，/wo

The eagle advises it [AR tortoise],

Figure 2: The sample of initial corpus with POS and anaphora resolution

Spatial Expression: Participants recognized a spatial expression according to the definition presented in Sect. 2.1, and annotated it with the following notations: the beginning symbol of a spatial expression #l, the end symbol #r, the spatial expression notation SE, and the sequence number of the spatial expression. For example in Figure 3, the string of "#l 空中/s #r[SE1]" means "空中" (in the sky) is the first spatial expression in the text. In addition, when there were two spatial expressions in one phrase of the definition they were all annotated separately at the same time like "#l在/p #l空中/s #r[SE1]#r[SE2]" in Figure 3.

Landmark: The landmark was annotated with *LM* and the same sequence number as its corresponding spatial expression in every annotated spatial expression. For example, the string of "#l空[LM1]中/s #r[SE1]" in Figure 3 shows that the landmark of "空" (the sky) corresponds to the first spatial expression of "空中" (in the sky). Moreover, when a landmark matched with two or more spatial expressions, it was annotated twice or more times. So, "空" (the sky) in Figure 3 was annotated with [LM1] and [LM2], meaning it is the

landmark not only for "空中" (in the sky) but also "在空中" (in the sky) at the same time.

Trajectory: The definition in Sect. 2.1 was not sufficient to annotate a trajectory for the participants, because there may be more than one word expressing the same object in the context. Thus, we proposed the nearest and right-most rule, which selects the sixth "乌龟" (the tortoise) as the corresponding trajectory of the third spatial expression in the text in the Figure 3. The trajectory notation is TR, and its sequence number is the same as its corresponding spatial expression.
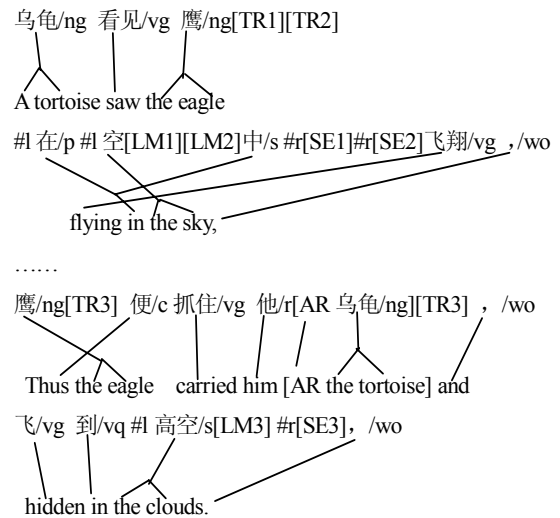
乌龟/ng 看见/vg 鹰/ng[TR1][TR2]

A tortoise saw the eagle

#l 在/p #l 空[LM1][LM2]中/s #r[SE1]#r[SE2]飞翔/vg ，/wo

flying in the sky,

……

鹰/ng[TR3] 便/c 抓住/vg 他/r[AR 乌龟/ng][TR3] ，/wo

Thus the eagle   carried him [AR the tortoise] and

飞/vg 到/vq #l 高空/s[LM3] #r[SE3]，/wo

hidden in the clouds.

Figure 3: The sample of final corpus with spatial relations annotations[1]

Spatial Relation: All spatial relations were explicit as soon as the spatial expressions, landmarks and trajectories were all annotated. A spatial expression, a landmark and a trajectory with the same sequence number compose a specific spatial relation, like "空[LM1] 中/s #r[SE1]", "空[LM1]" and "鹰/ng[TR1]" in Figure 3 compose a spatial relation of ((在, 中, 空), 鹰) ((in, , the sky), eagle).

2.3 Analysis of the Descriptive Strategies

We distinguish two subsets of our development corpus, 1) a set containing 325 fables selected randomly from the corpus as the 'training database', 2) a set that containing the rest 109 fables of the corpus as the 'testing database'. Based on this corpus, how the spatial relations are expressed in Chinese is analyzed in detail.

---

[1] Some sentences are omitted because of the paper space, where four tortoise words are included.

Mapping: The mapping between trajectories and their corresponding spatial expression is the quantity ratio between them. There are four possible mappings in the corpus. 0:1 means that there is zero trajectory matching one spatial expression. 1:1, 2:1 or 3:1 means there are one, two or three trajectories matching one spatial expression which result in one, two or three spatial relations. The statistical data are shown in Table 1.

Spatial Focus Shift: The spatial focus is the current entity or group of entities (and its/their associated spatial location) that the reader is attending to in space (Maybury, 1990). In Figure 4(a), there is a spatial relation "农夫在墙下" (the farmer under a wall) between the trajectory "鹰" (the eagle) and its spatial expression "朝下" (down). In this case the current spatial focus shifts from "鹰" (the eagle) to "农夫" (the farmer), and back to "鹰" (the eagle).

There are three types of spatial focus shift. The first type is *typeSI* when there is no other spatial relation between a trajectory and its corresponding spatial expression. The second type is *typeSII* when there are other spatial expression(s), but no complete spatial relation between them, which means there is not any spatial focus shift too. In Figure 4(b), there is a spatial expression "画板上" (in the wall) between the trajectory "鸽子" (the pigeon) and its another corresponding spatial expression "地上" (onto the ground), which is the grammar ellipsis of a trajectory raised by the ellipses of the object or subject. Lastly, there is type *typeSIII*, in which there are complete spatial relation(s) between a trajectory and its corresponding spatial expression, such is the case in Figure 4(a). This type means there is a semantic interruption by the spatial focus shift. We show the statistical data for the spatial focus shift for 1:1 spatial relations in Table 2.

Syntactic Location Relation: There are three types of syntactic location relations between a trajectory and its corresponding spatial expression depending whether they are in the same sentence ended by full stop, question mark or exclamatory mark, or not. The first is called *typeLI* that means the trajectory and its spatial expression are in the same sub-sentence that is the sentence without punctuations. The second is *typeLII* – they are in the same full sentence, not sub-sentence. The last is *typeLIII* meaning they are in the different sen-

tences. The statistic data on the corpus is shown in Table 3.

Data base	Amount of 0:1	Amount of 1:1	Amount of 2:1	Amount of 3:1
Training	5 (8%)	539 (86.94%)	58 (9.35%)	18 (2.90%)
Testing	2 (8.85%)	203 (89.82%)	18 (7.96%)	3 (1.46%)

Table 1: Mapping between *TR* and *SE*

Data base of 1:1	Amount of *TypeSI*	Amount of *TypeSII*	Amount of *Type-SIII*
Training	523(97.03%)	14 (2.6%)	2 (0.37%)
Testing	190(93.60%)	11(5.42%)	2(0.99%)

Table 2: Spatial Focus Shift in 1:1 of *TR* and *SE*

Data base	Amount of *TypeLI*	Amount of *TypeLII*	Amount of *Type-LIII*
Training	439(71.38%)	158(25.69%)	18(2.93%)
Testing	159(70.98%)	63(28.13%)	2(0.89%)

Table 3: Syntactic Location Relations

鹰/ng[TR3] 看见/vg 农夫/nc[TR1][TR2] 坐/vg 在/d, /wo

The eagle saw the farmer   sitting

#l 将要/vz 倒塌/vg 的/ sed #l 墙/ng[LM1][LM2] 下/f #r

under a wall that was on the verge of collapsing,

[SE1]#r[SE2]就/d 立刻/d #l 朝/p 下/f #r[SE3]飞/vg 去/vg ，/wo

and flew down at once,

(a) A Sentence from 'The Eagle and the Farmer'

他/r[AR 鸽子/ng][TR2][TR3][TR4][TR5]

She [AR the pigeon]

立刻/d 呼呼/o 地/usdi 猛/d 飞/vg 过去/vq ，/wo

flew   right   ,

不料/d 一头/d 碰撞/vg #l 在/p #l 画板/ng[LM2][LM3] 上/f

but collided   with the wall

#r[SE2]#r[SE3]，/wo 折断/vg 了/ut 翅膀/ng ，/wo

, broke her wings and

摔/vg #l 在/p #l 地[LM4][LM5]上/s #r[SE4]#r[SE5]，/wo

fell   onto the ground,
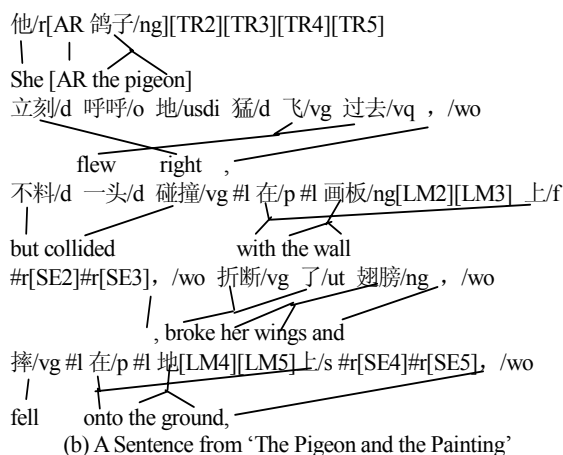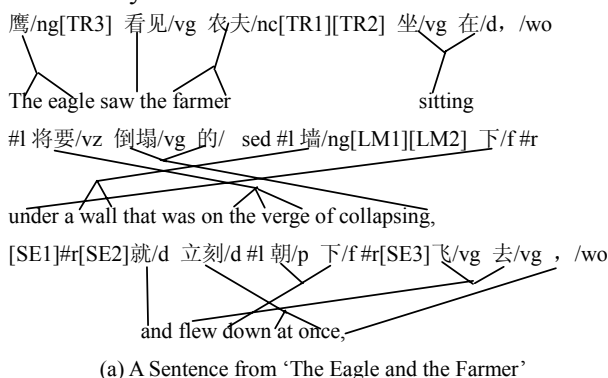
(b) A Sentence from 'The Pigeon and the Painting'

Figure 4: The sample for spatial focus shift

Those descriptive strategies are not mutually exclusive. Moreover, the phenomena of n:1 mapping, *typeSII*, *typeLII* and *typeLIII* are all produced by

the ellipses of the object or subject. In addition, trajectories of 0:1 and *typeSIII* must be understood from the background of a text, so these two descriptive strategies are not the subjects in this paper.

2.4 Winnow

We choose SNoW as the classifier tool and Winnow as the learning algorithms. SNoW is an architecture consisting of a Sparse Network of linear separators utilizing three learning algorithm including Winnow (Littlestone, 1988). Winnow is used at each target node to learn its dependence on nodes in the first layer. Its input can be formalized as (w_I, α, β, θ): an initial vector $w_I \in Rn$ which is positive, promotion factors α and $\beta \in [1, \infty]$, and a threshold $\theta \in [0, \infty]$. The algorithm proceeds in a series of trials and predicts in each trial according to the threshold function $w_k \cdot X \geq \theta$ given $X \in Rn$. If the prediction is correct, then no update is performed; otherwise the weights are updated as follows:

- On a false positive prediction for all *k* set $w_k \leftarrow \alpha w_k$

- On a false negative prediction for all *k* set $w_k \leftarrow \beta w_k$

3 Extraction of the Trajectory

3.1 The Binary Classification

We model the problem of extracting trajectories from a text as a binary classification problem. The text has been initially tagged using a POS tagger and a spatial expression tagger. The text is taken as the set *Stext* where there are two parts: the spatial expression *sexp_tr* being matched with trajectories and words excluding those in *sexp_tr*, which compose the set *Stext_w*. Every word of the set *Stext_w* is a candidate *ctr*. A predicate *p* taking values in *C* = {-1, 1} asserts whether a *ctr* is a trajectory (*p(ctr)*=1), or is not (*p(ctr)*=-1). The task then is to find the classifier function *h* that maps every word in *Stext_w* to a single value in *C*, *h*: *Stext_w->C*. The classifier *h*, moreover, is found by Winnow.

3.2 Features of the trajectory Classifier

Based on the studies of the descriptive strategies, the features are selected on shallow syntax, and then are modeled as a unit vector *f{pun, pos_ctr$_1$,…, pos_ctr$_m$, pos_verb$_1$,…, pos_verb$_n$,*

dis_verb$_1$,…, dis_verb$_o$, dis_SE$_1$,…, dis_SE$_q$, dis_SEL$_1$,…, dis_SEL$_p$, dis_SEN$_1$,…, dis_SEN$_h$, dis_by$_1$,…, dis_by$_u$, dis_with$_1$,…, dis_with$_i$, hby, hwith}.

The value of the feature *pun* is 1 as the candidate is punctuation and the others are 0, otherwise, the value is 0 and the others are decided on the following discussion.

pos_ctr$_i$ presents the candidate's POS. The used POS tool has 52 part-of-speech tags which are numbered. For example a candidate's POS is *noun*, and then the value of *pos_ctr$_1$* is 1, and the others are 0.

When there is a verb in the same sentence with candidate, *pos_verb$_i$* is assigned 1. The number of *pos_verb$_i$* assigned 1 means the number of verbs which are in the same sentence with the candidate.

There are three classes of distance features. The first *dis_verb$_i$* is distances between a candidate and verbs *ph* which are in the same full sentence. The second is distances between a candidate and phrases *ph*. The phrases include the spatial expression *sexp_tr*, the last spatial expression and the next which are relative to *sexp_tr* according to the sequence in the context. These features are denoted by *dis_SE$_i$*, *dis_SEL$_j$* and *dis_SEN$_k$*. The third *dis_by$_i$* and *dis_with$_i$* are distances between a candidate and function words denoted by *fw* including 被 ('by') and 把 ('with') which are in the same sub-sentence.

The first and second distances are calculated with Equation 1. The third distances are calculated according to Equation 2. When *ctr* is before *ph* or *fw*, *sign* is 1, otherwise is -1. *n* represents the number of words between *ctr* and *fw*. n_0 is 1 when *ctr* and *ph* are in the same sub-sentence, otherwise it is 0. n_1 means the number of sub-sentences between *ctr* and *ph*. n_2 or n_3 means the number of full sentences or paragraphs between *ctr* and *ph*. Every one of distance values maps a distance feature.

$$dis(ctr, ph) = sign * \sum_{i=0}^{3} (n_i * 10^i) \quad (1)$$
$$dis(ctr, fw) = sign * n \quad (2)$$

It is an important feature identifying a trajectory whether 被('by') or 把('with') appears in the same sub-sentence with it. If 被('by') or 把('with') and *ctr* appear in the same sub-sentence, the value of the feature *hby* or *hwith* is 1, otherwise it is 0.

4 Experiments

4.1 Qualification of the Values of the Parameters

Before the experiments, we determined the parameters of Winnow: the promotion factors α and β, and the threshold θ, using the enumeration method. The values of these three parameters for one experiment were selected when the *F*-score was maximal at all runs in the numeric area and on the changing step of every parameter in Table 4 confirmed according to our experience.

4.2 Experiments with the Different Feature Spaces

The following experiments were done separately according to four different feature spaces defined in Table 5 according to features in Sect. 3.2. The values of the three parameters of following experiments are shown in Table 6.

Parameter	Min	Max	Step
α	1.0	2.0	0.1
β	0.0	0.9	0.1
θ	1	6	1

Table 4: The Numeric Areas and Steps on the Parameters

Features in the feature space	Name
Parts-of-speech of candidates	*Fea*00
POS of candidates and distances between a candidate and spatial expressions	*Fea*01
All features mentioned in Sect. 3.2	*Fea*10

Table 5: The Definitions of Feature Spaces

Feature Space	α	β	θ
*Fea*00	1.2	0.2	2
*Fea*01	1.6	0.3	2
*Fea*10	1.7	0.5	2

Table 6 Values of the Parameters

First, the performance of *Fea*01 being much better than *Fea*00 in Figure 5 shows that the feature of spatial expressions is effective, and a trajectory and a spatial expression are bounding.

Second, Figure 6 shows that the performance of *Fea*10 is much stronger than that of *Fea*01. The *F*-score on the testing database is much lower than on training dataset, but the precision is much higher for *Fea*01. In contrast, the performances of *Fea*10 between training database and testing database fall at the same degree. Therefore, the features on function words and verbs are effective.
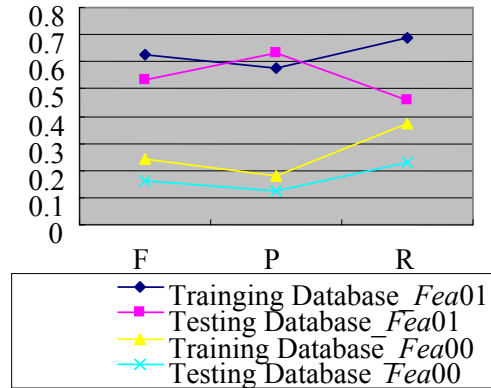

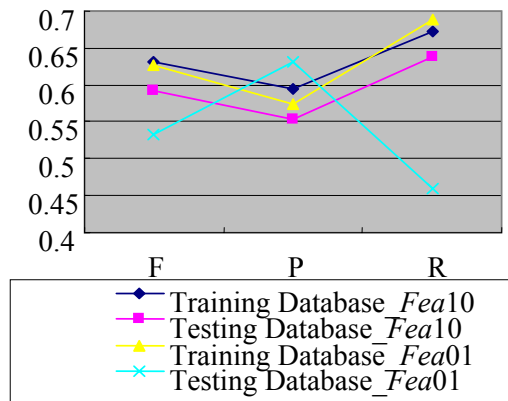Figure 5: The Results on the Feature Spaces *Fea*01 and *Fea*00

Figure 6: The Results on the Feature Spaces *Fea*10 and *Fea*01

We calculated the recalls on different mappings on *Fea*10 as shown in Figure 7. The method is most effective to 1:1 mapping, and can resolve the ellipsis of the object or subject in some degree.
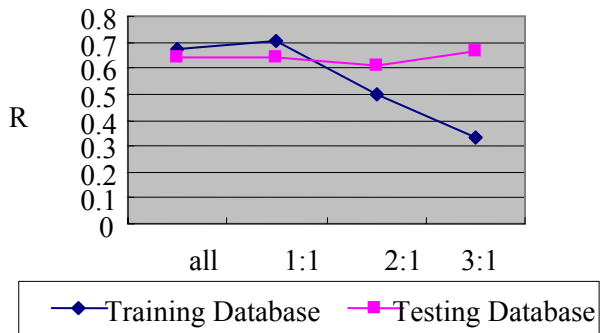

Figure 7: The Results on the Different Mappings

The recalls of syntactic locations are showed in Figure 8 for *Fea*10. The performance for the sub-sentence relation is over 85%. It is critical to resolve the data sparseness in the other two situations in the future.
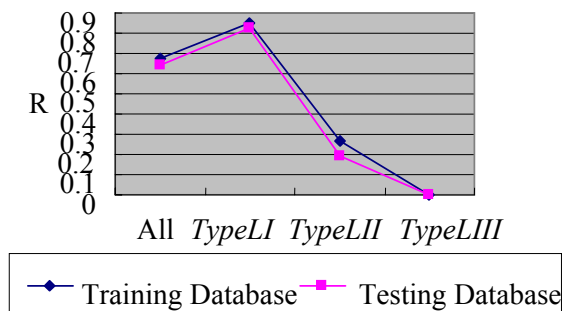
Figure 8: The Results on the Different Syntactic Location Relations

5   Conclusion

We showed the concepts of the spatial relation and trajectory and the descriptive strategy in Chinese texts. Based on these studies, we proposed that the extraction of the trajectory is a binary classification problem. We also proposed the method to extract the trajectories at the text level with a linear classifier. In the end, we presented experiments from different points of view to show that it is effective with a binary classifier with shallow syntactic features to extract a cognitive concept of a trajectory.

Further studies on the feature selecting in the semantics and at the text level are needed to improve the results and resolve the grammar and the semantic ellipsis of a trajectory, furthermore, to resolve the data sparseness. Moreover, we have found that about 94% words as trajectories are included in HowNet, which is a common sense knowledge base (Dong and Dong, 2002). Therefore, it is essential to take advantage of HowNet for better results in the next phases.

6   Acknowledgements

References

Bob Coyne, Richard Sproat. 2001. *WordsEye: An Automatic Text-to-Scene Conversion System*. Proceedings of the SIGGRAPH 2001 Annual Conference on Computer graphics, Los Angeles, CA USA.

Zhengdong Dong, Qiang Dong. 2007. *HowNet*. http://www.keenage.com/

Richard Johansson. 2006. *Natural Language Processing Methods for Automatic Illustration of Text*. ISSN 1652-4691, Licentiate Thesis, LU-CS-LIC. Lund, Sweden.

Ronald W. Langacker. 1987. Foundations of Cognitive Grammar (2 Vols). Stanford, Calif.: Standford University Press.

Nicholos Littlestone. 1988. *Learning Quickly when Irrelevant Attributes Abound: A New Linear-threshold Algorithm*. Machine Learning 2:285–318

Ruqian Lu, Songmiao Zhang. 2002. *From Story to Animation – Full Life Cycle Computer Aided Animation Generation* (in Chinese). Acta Automatica Sinica, 28(3): 322-348

Yajuan Lv. 2003. *Research on Bilingual Corpus Alignment and Automatic Translation Knowledge Acquisition*. Dissertation for the doctoral degree in engineering, Harbin Institute of Technology, Harbin, China.

Mark T. Maybury. 1990. *Using Discourse Focus, Temporal Focus, and Spatial Focus to Generate Multisentential Text.* Proceedings of the Fifth International Workshop on Natural Language Generation: 70-78, Dawson, PA.

Dan Roth. 1998. *Learning to Resolve Natural Language Ambiguities: A Unified Approach*. Proceedings of AAAI-98, 15th Conference of the American Association for Artificial Intelligence: 806-813.

Shaohua Zou. 1989. *The Grammar Function of Spatial Function Words in Chinese* (in Chinese). ZHONG-GUO WUWEN, 3:173-178