● -

*

(430065)

:

: ; ; ; ;
: TP391.1 : A

Chinese textual entailment recognition using statistical and lexical semantic features

LIU Maof-fu, LI Yan, GU Jin-guang

(*College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China*)

Abstract: In order to further analyze and understand the text, textual entailment recognition should be paid more attention to, especially the ones in Chinese text pair. The statistical and lexical semantic features, associated with Chinese text pair, are extracted after the Chinese text preprocessing, such as Chinese word segmentation and stop words removal. The textual entailment recognition is actually one classification task and the classification model based on support vector machine can be designed and implemented using the extracted statistical and lexical semantic features. The final experiment results demonstrate the effective and feasibility of the classification model using the textual statistical and semantic features.

Key words: textual entailment; statistical feature; lexical semantic feature; support vector machine; contradiction

0

ˇ

β Â h ч Ц Ц

[1,2] ч Ц Ш Ц Ц

Ь н ˇ ч Ц Ц Â

Ц ˇ ˇ Ш ˇ

ˇ à

à à Ц [3-5] Â

Ь ˇ Ц

RITE^ Recognizing

Inference in TExt~ ї н Ш˘ З Ш

[7] Â RITE Ни BC

^ Binary-Class~ MC^ Multi-Class~ Н Â

BC Н ^ *T, H*˜ ї ˇ ч

T *H* Ш ї Ни ' MC Н

^ *T, H*˜ ї F^ Forward

entailment~ à R^ Reverse entailment~ à ї

B Bidirectional entailment~ à C Contradiction~

I^ Independence~ Ш ˇ ч ї

Â ˇ ˇ

З Â 1 ₂ RITE

ˇ ї *T1 H1* ˇ *T2 H2* ї

ˇ *T3* Ц *H3* н ↕ Â

1'

T1 Ь Ь Ч ^

ˇ Ч ~~ ⁼ ч

ê ⁼ ч ЇÂ

H1 Ь Ч Ц Ш ч Â

T2 ˇ

Н Â

H2 Ий Ш ˇ

Н Â

T3 ⁼ ч ї . ˇ

' '

' Ни ð Ц ˇ Natural Language Generation Oriented Event Semantic Computation˜

61100133; Ь ˇ Resource Construction and Statistical Analysis of Chinese Textual Inferences˜

61173062 . Ни ð Ц ˇ Semantic Web Model Ba2011 j yte126

: ↔ (1977) ˇ ˇ ˇ ˇ CCF (E200013837S)

ˇ Н ч à ' (1987-

˙) ˇ . ˇ ˇ Н ч ' (1974) ˇ . ˇ ˇ ˇ Н ч

Ц Â E-mail: liumaofu@wust.edu.cn

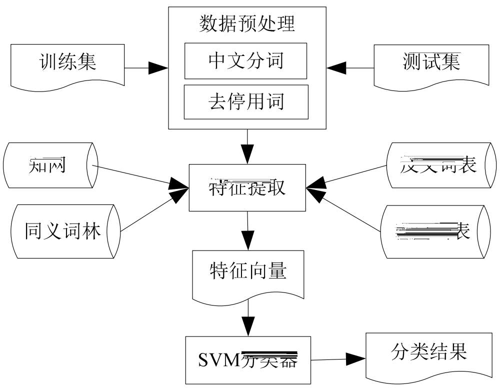
10

LIBSVM[2]

11 6 5

1

SVM 1



1

ICTCLAS 1 *T2* *H2*

T2

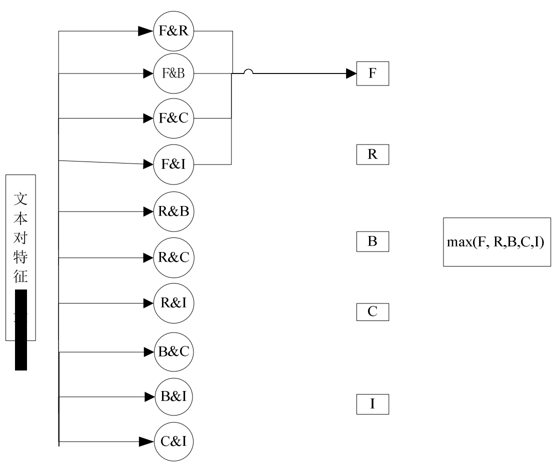
H2

[8.9]

SVM

RITE BC RITE MC

2



ì **t h** *T H* ˘ *n* ˘

TF*IDF Â

ˆ 6˜

˘ Ң

Â

ˆ 6˜ Â

$$\mathrm{Sim}_{LCS}(T,H)=\frac{\mathrm{len}(\mathrm{LCS}(T,H))}{\min(length(T),length(H))}$$ ˆ 6˜

ì LCS(T,H) *T H* Â

2.2

1 ↓ ˜ ˘

RITE BC MC ⊢ 33

Â 1 ˘ WUST BC

MC ⊢ 11 ß ÂUIOWA

' liumaofu@wust.edu.cn

' 430065