# Subjective Bayes Method for Word Semantic Similarity Measurement

Junhua Wang, Xianglin Zuo, Wanli Zuo, Tao Peng
College of Computer Science and Technology
Key Laboratory of Symbolic Computation and Knowledge Engineering attached to the Ministry of Education
Jilin University
Changchun, China
e-mail: wangjh10@jlu.edu.cn, 295228473@qq.com, (corresponding)wanli@jlu.edu.cn, tpeng@jlu.edu.cn

*Abstract*—**Measuring semantic similarity between words is a classical problem in nature language processing, the result of which can promote many applications such as machine translation, word sense disambiguation, ontology mapping, computational linguistics, etc. This paper combines knowledge-based methods with statistical methods in measuring words similarity, the novel aspect of which is that subjective Bayes method is employed. Firstly, extract evidences based on WordNet; secondly, analyze reasonableness of candidate evidence using scatter plot; thirdly, generate sufficiency measure by statistics and piecewise linear interpolation technique; fourthly, obtain comprehensive posteriori by integrating uncertainty reasoning with conclusion uncertainty synthetic strategy; finally, we quantify word semantic similarity. On data set R&G (65), we conducted experiment through 5-fold cross validation, and the correlation of our experimental results with human judgment is 0.912, with 0.4% improvements over existing best practice, which show that using subjective Bayes method to measure word semantic similarity is reasonable and effective.**

***Keywords- Word Semantic Similarity; Scatter Plot; Piecewise linear interpolation; Subjective Bayes; WordNet***

## I. INTRODUCTION

Measuring semantic similarity between words has long been a research problem in nature language processing, the result of which can promote many applications such as machine translation [1], word sense disambiguation [2], information extraction [3], opinion mining [4], etc. So far, many approaches have been proposed for word semantic similarity measurement which can be grouped into two categories: knowledge-based and corpus-based method.

The foundation of knowledge-based method is semantic dictionary such as WordNet, MindNet and FrameNet. In the early work, R. Rada [5] computed word semantic similarity by calculating the semantic distance between corresponding concepts. Based on Rada's algorithm, P. Resnik [6] took into account the common ancestor of pending concepts. Wu and Palmer [7] measured semantic similarity by utilizing the depths of pending concepts and their nearest common ancestor based on concept hyponymy. In addition to the path length between concepts, Rigau and Agirre [8] considered not only the depth of concept, but also the local density they introduced. Jiang and Conrath [9] calculated concept semantic similarity by measuring the information of concepts and their nearest common ancestor. Lin [10] computed concept semantic similarity by measuring the information of concepts and their nearest common ancestor too, but in a different way. Leacock and Chodorow [2] introduced maximum depth of concept system on the basis of shortest distance of concepts for measuring concept semantic similarity. Hirst and St-Onge [11] suggested that if the path between concepts in question is short and the path's direction does not change frequently, then the concepts' semantics is similar. Yuhua Li et al. [12] combined path length, the depth of concept and information for measuring concept semantic similarity. Yang and Powers [13] used distance between concepts with seven parameters to measure concept semantic similarity based on three relationships (hyper/hyponym, hol/meronym, syn/antonym). Alexander B. et al. [14] analyzed earlier five typical method proposed by Jiang and Conrath, Hirst and St-Onge, Leacock and Chodorow, Lin, Resnik respectively. In recent years, some researchers improved previous methods to a certain extent. Marco A et al. [15] utilized synsets, synset relations and synset definitions to construct a weighted graph; and measured word semantic similarity by defining word distance based on the weighted graph. Peng et al. 16] combined WordNet with directed acyclic graph theory for word semantic similarity measurement. Giuseppe et al. [17] proposed a novel feature-based method by employing information theory. Cai et al. [18] combined improved distance-based method, information-based method with WordNet to measure word semantic similarity. David et al. [19-20] improved information-based method and feature-based method respectively. Liu et al. [21] measured concept semantic similarity with concept vector cosine similarity.

Corpus-based method quantified context similarity with large-scale corpus and statistical techniques to measure word semantic similarity. Dagan [22] used probabilistic model to calculate word distance. P. F. Brown [23] employed average mutual information technique to measure word semantic similarity. L. Lillian [24] adopted related entropy model to compute word semantic similarity. Lei Liu et al. [25] measured word semantic similarity by using pattern vector space model. Tao Xu et al. [26] built meaning vector by calculating word semantic similarity. Radinsky et al. [27] found the association between words by establishing concept temporal dynamics.

Corpus-based method is subject to the adopted corpus and cannot avoid data sparseness problem. Knowledge-based method is simple, effective and more intuitive; does not need

978-0-7695-5109-8/13 $31.00 © 2013 IEEE
DOI 10.1109/ICDMW.2013.47

corpus for training; but is more impacted by person's subjective consciousness. Based on WordNet, we utilize scatter plots, statistical techniques, piecewise linear interpolation and subjective Bayes method to model concept semantic similarity and obtain semantic similarity of related words. As shown in Fig. 1, the key steps of concept semantic similarity modeling are identifying evidence by scatter plot, generating sufficiency measure by statistics and piecewise linear interpolation technique, updating priori probability through uncertainty reasoning, and receiving comprehensive posteriori by conclusion uncertainty synthetic strategy. The novel respect of our approach is that Subjective Bayes method is used in word semantic similarity measurement for the first time. And experiment result shows that our method improved the effect of word semantic similarity measurement.

## II. SUBJECTIVE BAYES METHOD

Subjective Bayes [28] is an uncertainty reasoning model proposed by Duda et al through revising Bayes formula, which was successfully applied in PROSPECTOR expert system. Subjective Bayes method has a better theoretical foundation, and avoids a lot of statistical work as knowledge inputting being converted to sufficiency measure assignment and necessity measure assignment. Knowledge is represented by production rules in subjective Bayes method. It's form is IF $E$ THEN ($LS$, $LN$) $H$ ($P(H)$). Where $LS$, $LN$ are called sufficiency measure and necessity measure respectively. And each assertion $H$ has a prior probability $P(H)$.

*Definition 1 Sufficiency measurement* ($LS$). $LS$ represents support degree of the emergence of evidence $E$ for assertion $H$. Its range is from 0 to ∞.

$$LS= P(E|H)/P(E|\neg H) \tag{1}$$

*Definition 2 Necessity measurement* ($LN$). $LN$ represents support degree of the absence of evidence $E$ for assertion $H$. Its range is from 0 to ∞.

$$LN= P(\neg E|H)/P(\neg E|\neg H) \tag{2}$$

Uncertainty reasoning of subjective Bayes method is the process of obtaining $P(H|E)$ or $P(H|\neg E)$ with $P(H)$, $P(E)$, $LS$ and $LN$. The basic idea is: $H$'s credibility should be changed with new information acquisition; that is, the process of updating priori probability $P(H)$ to posterior probability $P(H|E)$ or $P(H|\neg E)$ with $P(E)$ based on $LS$ or $LN$.

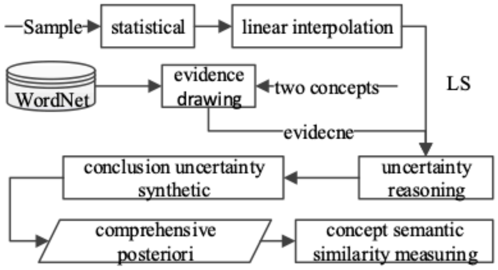


Figure 1. Concept Semantic Similarity Modeling

*Reasoning 1*. When $E$ is totally certain ($P(E)$=1); there is:

$$P(H|E)=LS\times P(H)/((LS-1)\times P(H)+1). \tag{3}$$

*Reasoning 2*. When $E$ is totally uncertain ($P(E)$=0); there is:

$$P(H|\neg E)=LN\times P(H)/((LN-1)\times P(H)+1). \tag{4}$$

*Reasoning 3*. When $E$ is uncertain ($0<P(E|S)<1$), $P(E|S)$ is the possibility of the existence of evidence $E$ with $S$ appearing; there is:

$$P(H|S)=P(H|E)\times P(E|S)+P(H|\neg E)\times P(\neg E|S). \tag{5}$$

Suppose there are $n$ knowledge supporting the same conclusion, whose antecedents are $n$ independent evidences $E_1,E_2,..,E_n$, then the conclusion uncertainty synthetic strategy is:

$$P(H \mid E_1E_2,...,E_n)=\frac{\prod_{i=1}^{n} LS_i \times P(H)}{1-P(H)+\prod_{i=1}^{n} LS_i \times P(H)}. \tag{6}$$

## III. EVIDENCE IDENTIFICATION BASED ON SCATTER PLOT

### A. Basic Concept

*Definition 3 Scatter plot*. Scatter plot is a graph using scatter distribution pattern reflecting variable statistical relationship with a variable as abscissa and another variable as ordinate. It can demonstrate the overall relationship trends and changing shape between factors and prediction object visually, provide decision support for selecting suitable mathematical expression to simulate variable relationship. It is the most intuitive graph for measuring the strength of relationships between variables.

*Definition 4 Hyponymy Graph* ($HG$). In WordNet, hyponymy relationship accounts for nearly 80 percent of all link types, and concepts are linked in a graph using hyper/hyponym (as shown in Fig. 2). The graph is named Hyponymy Graph, $HG$ = ($V$, $E$, $r$). Where, $V$ is a concept node set, $E$ is a binary relation on $V$, and $r$ is the root concept node of the Hyponymy Graph.

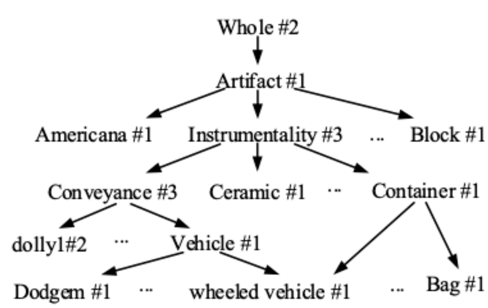


Figure 2. Concept Hyponymy Graph (part)

*Definition 5 Concept Path.* If the concept node sequence $P=(v_1,v_2,\ldots,v_n)$ satisfies condition $E(v_i,v_{i+1})$ $(0<i<n)$, then $P$ is called the concept path from $v_1$ to $v_n$.

*Definition 6 Concept Path Length.* If there is a concept path $P=(v_0,v_1,v_2,\ldots,v_n)$ from $v_0$ to $v_n$, then the length of $P$ is equal to $n$, denoted by $L_P=n$.

*Definition 7 Concept Depth.* The depth of $v$ is defined as the length of $P=(r,v_1,v_2,\ldots,v)$, which is the shortest concept path from $r$ to $v$, denoted by $D_v=L_P$.

*Definition 8 Concept Distance.* The distance between $v_1$ and $v_2$ is defined as the length of $P_1$ plus the length of $P_2$, where $P_1$ and $P_2$ is the shortest path from $v$ to $v_1$ and $v_2$, and $v$ is the nearest common ancestor of $v_1$ and $v_2$. We stipulate that the distance from concept to itself is zero.

### B. Evidence Identification

We identify $L(v_1,v_2)$ and $D_v$ as the candidate evidences for measuring the semantic similarity between $v_1$ and $v_2$,where $v$ is the nearest common ancestor of $v_1$ and $v_2$, $L(v_1,v_2)$ is the concept distance between $v_1$ and $v_2$; and analyze the overall relationship trends and changing shape between candidate evidences and concept semantic similarity with scatter plot to determine the candidate evidences' reasonableness based on manually annotated sample data R&G. Taking into account most of works were tested on R&G data set we also select it, which can facilitate the comparison with our work and current works. As shown in table I, $L(v_1,v_2)$ and $D_v$ are computed with WordNet1.6 owing to woodland missing from high versions, and the semantic similarity between $w_1$ and $w_2$ is converted to the semantic similarity between concept $v_1$ and $v_2$, where $v_1$ is the synset of $w_1$, $v_2$ is the synset of $w_2$, and the distance between $v_1$ and $v_2$ is the least distance for $w_1$ and $w_2$.

Based on table 1, we draw two scatter plots to analyze the overall relationship trends and changing shape between $L(v_1,v_2)$, $D_v$ and concept semantic similarity. In the left figure of Fig. 3, the abscissa is $L(v_1,v_2)$ and the ordinate is *Sim*. In the right figure of Fig. 3, the abscissa is $D_v$ and the ordinate is *Sim*. As shown in Fig. 3, the overall relationship trends between $L(v_1,v_2)$ and concept semantic similarity reveals negative correlation, whereas the overall relationship trends between $D_v$ and concept semantic similarity reveals positive correlation. The impact trend of $L(v_1,v_2)$ on concept semantic similarity is different from the impact trend of $Dv$ on concept semantic similarity, but $L(v_1, v_2)$ and $D_v$ both have significant correlation with concept semantic similarity. And $L(v_1,v_2)$ and $D_v$ are mutually independent according to evidence independence requirement for concluding uncertainty synthetic strategy. So choosing $L(v_1,v_2)$ and $D_v$ as the evidence for measuring concept semantic similarity is reasonable and effective.

Given the limitations of human recognition ability, Yang [12] suggests the bound of measuring concept semantic similarity with concept distance is 12; and the maximum concept depth of WordNet is 16. Thus, we set the range of $L(v_1,v_2)$ as {0,1,2,3,4,5,6,7,8,9,10,11,12}, the range of $D_v$ as {0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16}, and stipulate that the concept distance which is greater than 12 is 12.

TABLE I. DATA SET R&G(65)

| WORD | *S* | *LW* | *DW* | WORD | *S* | *LW* | *DW* |
|---|---|---|---|---|---|---|---|
| cord-smile | 0.02 | 12 | 0 | journey-car | 1.55 | 30 | 0 |
| noon-string | 0.04 | 30 | 0 | cemetery-mound | 1.69 | 8 | 1 |
| rooster-voyage | 0.04 | 30 | 0 | glass-jewel | 1.78 | 7 | 2 |
| fruit-furnace | 0.05 | 6 | 2 | magician-oracle | 1.82 | 2 | 4 |
| autograph-shore | 0.06 | 30 | 0 | crane-implement | 2.37 | 4 | 3 |
| automobile-wizard | 0.11 | 11 | 0 | lad-brother | 2.41 | 4 | 2 |
| mound-stove | 0.14 | 6 | 2 | sage-wizard | 2.46 | 5 | 2 |
| grin-implement | 0.18 | 30 | 0 | oracle-sage | 2.61 | 7 | 2 |
| asylum-fruit | 0.19 | 6 | 2 | bird-cock | 2.63 | 1 | 5 |
| asylum-monk | 0.39 | 10 | 0 | bird-crane | 2.63 | 3 | 5 |
| graveyard-madhouse | 0.42 | 12 | 1 | food-fruit | 2.69 | 4 | 3 |
| glass-magician | 0.44 | 8 | 0 | brother-monk | 2.74 | 1 | 5 |
| boy-rooster | 0.44 | 11 | 1 | asylum-madhouse | 3.04 | 1 | 7 |
| cushion-jewel | 0.45 | 6 | 2 | furnace-stove | 3.11 | 2 | 2 |
| monk-slave | 0.57 | 4 | 2 | magician-wizard | 3.21 | 0 | 4 |
| asylum-cemetery | 0.79 | 9 | 1 | hill-mound | 3.29 | 0 | 7 |
| coast-forest | 0.85 | 6 | 1 | cord-string | 3.41 | 1 | 4 |
| grin-lad | 0.88 | 30 | 0 | glass-tumbler | 3.45 | 1 | 5 |
| shore-woodland | 0.9 | 5 | 1 | serf-slave | 3.46 | 3 | 3 |
| monk-oracle | 0.91 | 7 | 2 | grin-smile | 3.46 | 0 | 7 |
| boy-sage | 0.96 | 5 | 2 | journey-voyage | 3.58 | 1 | 5 |
| automobile-cushion | 0.97 | 7 | 3 | autograph-signature | 3.59 | 1 | 5 |
| mound-shore | 0.97 | 4 | 3 | coast-shore | 3.6 | 1 | 4 |
| lad-wizard | 0.99 | 4 | 2 | forest-woodland | 3.65 | 0 | 3 |
| forest-graveyard | 1 | 7 | 1 | tool-implement | 3.66 | 1 | 4 |
| food-rooster | 1.09 | 12 | 0 | cock-rooster | 3.68 | 0 | 9 |
| cemetery-woodland | 1.18 | 7 | 1 | boy-lad | 3.82 | 1 | 4 |
| shore-voyage | 1.22 | 30 | 0 | cushion-pillow | 3.84 | 1 | 4 |
| bird-woodland | 1.24 | 7 | 1 | cemetery-graveyard | 3.88 | 0 | 6 |
| coast-hill | 1.26 | 4 | 3 | car-automobile | 3.92 | 0 | 7 |
| furnace-implement | 1.37 | 5 | 2 | gem-jewel | 3.94 | 0 | 6 |
| crane-rooster | 1.41 | 7 | 5 | midday-noon | 3.94 | 0 | 7 |
| hill-woodland | 1.48 | 5 | 1 | | | | |

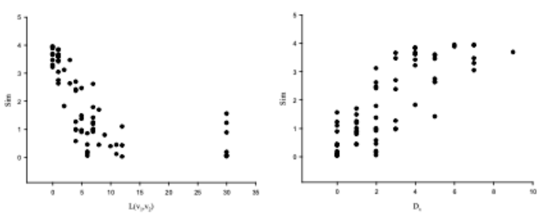

Figure 3. Evidence Analyze Scatter Plots

## IV. WORD SEMANTIC SIMILARITY MODELING

Subjective Bayes uncertainty reasoning is based on knowledge base, the knowledge of which is in the form of rules. Rule is comprised of antecedent, consequent, sufficiency measure, necessity measure, and priori probability. For concept semantic similarity measurement, the antecedent of rule is a single evidence($L(v_1,v_2)$= *LValue*, *LValue* = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}; or $D_v$=*DValue*, *DValue*= {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16}), the consequent of rule is $C=c_1$ (consistency) or $C=c_2$ (unconsistency), the value of $P(C=c_1)$ and $P(C= c_2)$ are 0.5 and 0.5. And the evidence of $L(v_1,v_2)$= *LValue* or $D_v$=*DValue* is certain, so the uncertainty reasoning for receiving Posteriori follows reasoning 1. As reasoning1 does not involve necessity measure, we only consider the value of sufficiency measure, and do not evaluate necessity measure.

## A. Sufficiency Measure Assignment

As shown in equation (1), the key to compute sufficiency measure is to obtain conditional probability $P(L(v_1,v_2)|C)$ and $P(D_v|C)$. In this paper, we quantify $P(L(v_1,v_2)|C)$ and $P(D_v|C)$ with discrete function $LRG(L(v_1,v_2))$ and $DRG(D_v)$.

$$P\left(L\left(v_1,v_2\right)=LValue_i \mid C=c_1\right)=\frac{LRG\left(LValue_i\right)}{\sum_{i=0}^{12} LRG\left(LValue_i\right)} \tag{7}$$

$$P\left(L\left(v_1,v_2\right)=LValue_i \mid C=c_2\right)=\frac{1-LRG\left(LValue_i\right)}{13-\sum_{i=0}^{12} LRG\left(LValue_i\right)} \tag{8}$$

$$P\left(D_v=DValue_j \mid C=c_1\right)=\frac{DRG\left(DValue_j\right)}{\sum_{j=0}^{16} DRG\left(DValue_j\right)} \tag{9}$$

$$P\left(D_v=DValue_j \mid C=c_2\right)=\frac{1-DRG\left(DValue_j\right)}{17-\sum_{j=0}^{16} DRG\left(DValue_j\right)} \tag{10}$$

Where, $LRG(L(v_1,v_2))$ and $DRG(D_v)$ are generated by statistical techniques and piecewise linear interpolation method based on human-annotated sample set R&G(65).

### 1) Statistical Policy

We use statistical techniques to calculate the function value ($LRG(LValue_i)$ or $DRG(DValue_j)$) for variable $L(v_1,v_2) = LValue_i$ or $Dv = DValue_j$ appeared in sample set, as shown in equation (11-12).

$$LRG(LValue_i)=LRGMean(LValue_i) \tag{11}$$

$$DRG(DValue_j))=DRGMean(DValue_j) \tag{12}$$

Where, $LRGMean(LValue_i)$ is equal to a quarter of mean value in R&G samples with $L(v_1,v_2)=LValue_i$, $DRGMean(DValue_j)$ is equal to a quarter of mean value in R&G samples with $D_v=DValue_j$.

### 2) Piecewise Linear Interpolation Strategy

*Definition 9 Piecewise Linear Interpolation.* In each interval $[x_i,x_{i+1}]$, use one order polynomial (Straight line) approaching $f(x)$. In other words, curve is replaced by segments of lines.

$$f(x)=(x-x_{i+1})/(x_i-x_{i+1})*y_i+(x-x_i)/(x_{i+1}-x_i)*y_{i+1} \tag{13}$$

We use piecewise linear interpolation strategy to calculate the function value $LRG(LValue_i)$ or $DRG(DValue_j)$ for variable $L(v_1,v_2)=LValue_i$ or $Dv=DValue_j$ not appearing in sample set, as shown in equation (14-15).

$$LRG(LValue_i)=LRGInter(LValue_i) \tag{14}$$

$$DRG(DValue_j))=DRGInter(DValue_j) \tag{15}$$

Where, the value of $LRGInter(LValue_i)$ is obtained by using one order polynomial approaching $LRGMean(LValue_i)$ in interval $[LValue_k,LValue_{k+1}]$, $LValue_k$ is the nearest sample from $LValue_i$ for the left, $LValue_{k+1}$ is the nearest sample from $LValue_i$ for the right; the value of $DRGInter(DValue_j)$ is obtained by using one order polynomial approaching $DRGMean(DValue_j)$ in interval $[DValue_t,DValue_{t+1}]$, $DValue_t$ is the nearest sample from $DValue_j$ for the left, $DValue_{t+1}$ is the nearest sample from $DValue_j$ for the right.

As sample limited, the priori probability of $Dv$=16 cannot be calculated by statistical techniques and piecewise linear interpolation method. We stipulate that the value of $DRG(16)$ is equal to 1, as two concepts with $Dv$=16 must be the same concept.

## B. Concept Semantic Similarity Measurement

When it comes to measuring semantic similarity between concepts by human beings, we can find that human tends to decide similarity between concepts through comparing their common features and different features. In this paper, $P(C=c_1|L(v_1,v_2)=LValue_i, D_v=DValue_j)$ reflecting the probability of concepts' semantic being same is considered as the common information of concepts, $P(C=c_2|L(v_1,v_2)=LValue_i, D_v=DValue_j)$ reflecting the probability of concepts' semantic being irrelevant is considered as the difference of concepts, and on this basis, the concept semantic similarity model can be defined as follows.

$$Sim=\frac{\alpha V_1}{\alpha V_1+\beta V_2} \tag{16}$$

Where, $V_1=P(C=c_1|L(v_1,v_2)=LValue_i,D_v=DValue_j)$, $V_2=P(C=c_2|L(v_1,v_2)=LValue_i,D_v=DValue_j)$, $\alpha$ and $\beta$ are adjustment factors, $\alpha=\sum_{i\in[0,12]} LRG(i)\sum_{j\in[0,16]} DRG(j)$, $\beta=\sum_{i\in[0,12]}(1-LRG(i))\sum_{j\in[0,16]}(1-DRG(j))$.

As shown in equation (16), concept semantic similarity measuring is converted to computation of $P(C=c_1|L(v_1,v_2)=LValue_i,D_v=DValue_j)$ and $P(C=c_2|L(v_1,v_2)=LValue_i,D_v=DValue_j)$. While evidences $L(v_1,v_2)=LValue_i$ and $D_v=DValue_j$ are independent of each other, we can obtain $P(C=c_1|L(v_1,v_2)=LValue_i,D_v=DValue_j)$ and $P(C=c_2|L(v_1,v_2)=LValue_i,D_v=DValue_j)$ by utilizing conclusion uncertainty synthetic strategy.

$$V_1=\frac{LS_1\times LS_2\times P\left(C=c_1\right)}{1-P\left(C=c_1\right)+LS_1\times LS_2\times P\left(C=c_1\right)} \tag{17}$$

$$V_2=\frac{LS_3\times LS_4\times P\left(C=c_2\right)}{1-P\left(C=c_2\right)+LS_3\times LS_4\times P\left(C=c_2\right)} \tag{18}$$

$LS_1$ is the sufficiency measure of the rule IF $L(v_1,v_2)=LValue_i$ THEN $C=c_1$. It represents the support degree of the emergence of evidence $L(v_1,v_2)=LValue_i$ for assertion $C=c_1$.

$$LS_1 = \frac{P\left(L\left(v_1, v_2\right) = LValue_i \mid C = c_1\right)}{P\left(L\left(v_1, v_2\right) = LValue_i \mid C = c_2\right)} \quad (19)$$

$LS_2$ is the sufficiency measure of the rule IF $D_v$=$DValue_j$ THEN $C$=$c_1$. It represents the support degree of the emergence of evidence $D_v$=$DValue_j$ for assertion $C$= $c_1$.

$$LS_2 = \frac{P\left(D_v = DValue_j \mid C = c_1\right)}{P\left(D_v = DValue_j \mid C = c_2\right)} \quad (20)$$

$LS_3$ is the sufficiency measure of the rule IF $L(v_1,v_2)$=$LValue_i$ THEN $C$=$c_2$. It represents the support degree of the emergence of evidence $L(v_1,v_2)$=$LValue_i$ for assertion $C$= $c_2$.

$$LS_3 = \frac{P\left(L\left(v_1, v_2\right) = LValue_i \mid C = c_2\right)}{P\left(L\left(v_1, v_2\right) = LValue_i \mid C = c_1\right)} \quad (21)$$

$LS_4$ is the sufficiency measure of the rule IF $D_v$=$DValue_j$ THEN $C$= $c_2$. It represents the support degree of the emergence of evidence $D_v$=$DValue_j$ for assertion $C$= $c_2$.

The way of obtaining $P(L(v_1,v_2)$=$LValue_i$|C=$c_1$), $P(D_v$=$DValue_j$|C=$c_1$), $P(L(v_1,v_2)$=$LValue_i$|C=$c_2$) and $P(D_v$=$DValue_j$|C=$c_2$) has been introduced above. Thus, the concept semantic similarity modeling has been concluded.

The semantic similarity between $w_1$ and $w_2$ is converted to the semantic similarity between concept $v_1$ and $v_2$, where $v_1$ is the synset of $w_1$, $v_2$ is the synset of $w_2$, and the distance between $v_1$ and $v_2$ is the least for $w_1$ and $w_2$. We give the word semantic similarity algorithm $Sim_{SB}$ in Fig. 4.

## V. Experiment and Result Analysis

We conducted experiment through 5-fold cross validation on data set R&G (65), which is divided into five sub-samples. One sub-sample is retained for validating the model; the other sub-samples are used for training. Cross-validation is repeated five times, once for each sub-sample.

### A. Experimental Result

We give the experimental results in table II-VI for each sub-sample.

```
input: words w1 and w2; sample data R&G
output: Sim
1. (LRG, DRG)←generateFunction(R&G)
2. PD←setProbabilityDistribution(LRG, DRG)
3. (c1, c2)←getConcept(w1, w2)
4. (L, D)←getEvidence(c1, c2)
   //LValuei=L
   //DValuej=D
5. (LS1,LS2,LS3,LS4)←getLS(L,D,PD)
6. (V1,V2)←getPosteriori(LS1,LS2,LS3,LS4)
7. (α, β)←setParameter(LRG, DRG)
8. Sim←(α*V1)/( α*V1+β*V2)
9. Return Sim
```

Figure 4. $Sim_{SB}$ Algorithm

TABLE II. Sub-sample Data Set 1

| Sample | R&G | $Sim_{D-S}$ | Sample | R&G | $Sim_{D-S}$ |
|---|---|---|---|---|---|
| glass-magician | 0.44 | 0.0363 | cemetery-mound | 1.69 | 0.0941 |
| asylum-cemetery | 0.79 | 0.0613 | crane-implement | 2.37 | 0.4355 |
| coast-forest | 0.85 | 0.0166 | oracle-sage | 2.61 | 0.1782 |
| monk-oracle | 0.91 | 0.1782 | autograph-signature | 3.59 | 0.9218 |
| lad-wizard | 0.99 | 0.2240 | car-automobile | 3.92 | 0.9747 |
| forest-graveyard | 1 | 0.1314 | midday-noon | 3.94 | 0.9747 |
| journey-car | 1.55 | 0.0136 | | | |

TABLE III. Sub-sample Data Set 2

| Sample | R&G | $Sim_{D-S}$ | Sample | R&G | $Sim_{D-S}$ |
|---|---|---|---|---|---|
| automobile-wizard | 0.11 | 0.0196 | glass-jewel | 1.78 | 0.1739 |
| grin-implement | 0.18 | 0.0264 | sage-wizard | 2.46 | 0.1572 |
| asylum-monk | 0.39 | 0.0286 | food-fruit | 2.69 | 0.3835 |
| graveyard-madhouse | 0.42 | 0.0562 | journey-voyage | 3.58 | 0.9134 |
| grin-lad | 0.88 | 0.0264 | coast-shore | 3.6 | 0.9523 |
| boy-sage | 0.96 | 0.1572 | tool-implement | 3.66 | 0.9523 |
| cemetery-woodland | 1.18 | 0.1543 | | | |

TABLE IV. Sub-sample Data Set 3

| Sample | R&G | $Sim_{D-S}$ | Sample | R&G | $Sim_{D-S}$ |
|---|---|---|---|---|---|
| rooster-voyage | 0.04 | 0.0341 | bird-crane | 2.63 | 0.9366 |
| autograph-shore | 0.06 | 0.0341 | asylum-madhouse | 3.04 | 0.9844 |
| asylum-fruit | 0.19 | 0.0463 | furnace-stove | 3.11 | 0.7460 |
| monk-slave | 0.57 | 0.2968 | glass-tumbler | 3.45 | 0.9328 |
| bird-woodland | 1.24 | 0.1486 | cock-rooster | 3.68 | 0.9927 |
| coast-hill | 1.26 | 0.5600 | cemetery-graveyard | 3.88 | 0.9985 |
| magician-oracle | 1.82 | 0.9819 | | | |

TABLE V. Sub-sample Data Set 4

| Sample | R&G | $Sim_{D-S}$ | Sample | R&G | $Sim_{D-S}$ |
|---|---|---|---|---|---|
| cord-smile | 0.02 | 0.0204 | lad-brother | 2.41 | 0.2517 |
| mound-stove | 0.14 | 0.0491 | bird-cock | 2.63 | 0.9416 |
| boy-rooster | 0.44 | 0.0094 | cord-string | 3.41 | 0.9622 |
| cushion-jewel | 0.45 | 0.0491 | grin-smile | 3.46 | 0.9896 |
| mound-shore | 0.97 | 0.4304 | forest-woodland | 3.65 | 0.9335 |
| shore-voyage | 1.22 | 0.0204 | boy-lad | 3.82 | 0.9622 |
| hill-woodland | 1.48 | 0.1569 | | | |

TABLE VI. Sub-sample Data Set 5

| Sample | R&G | $Sim_{D-S}$ | Sample | R&G | $Sim_{D-S}$ |
|---|---|---|---|---|---|
| noon-string | 0.04 | 0.0216 | brother-monk | 2.74 | 0.9579 |
| fruit-furnace | 0.05 | 0.0565 | magician-wizard | 3.21 | 0.9855 |
| shore-woodland | 0.9 | 0.1891 | hill-mound | 3.29 | 0.9926 |
| automobile-cushion | 0.97 | 0.4080 | serf-slave | 3.46 | 0.6986 |
| food-rooster | 1.09 | 0.0216 | cushion-pillow | 3.84 | 0.9630 |
| furnace-implement | 1.37 | 0.2670 | cemetery-graveyard | 3.88 | 0.9980 |
| crane-rooster | 1.41 | 0.6875 | | | |

### B. Result Analysis

We measure the accuracy of algorithm $Sim_{SB}$ by calculating correlation coefficient of experimental results with manual annotation. Correlation coefficient is an index for measuring linear correlation between two random variables. It has been widely used in various fields of science. Correlation coefficient ($r$) is defined in equation (22) with range of [-1,1], where $r$>0 indicates a positive correlation, $r$<0 indicates a negative correlation, |$r$| represents the degree of correlation between variables. Specially, $r$=1 is called perfect positive correlation, $r$=-1 is called perfect negative correlation, $r$=0 is called irrelevant. IF |$r$| is greater than 0.8 then the two variables have a strong linear correlation.

$$r_{XY} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \quad (22)$$

We calculate the correlation coefficient of experimental results with manual annotation data for each sub-sample. As shown in Fig. 5, there is a strong correlation between each experimental result with R&G data. The mean correlation for 5-fold cross validation experiment on R&G(65) is 0.912, which is higher than the best known method as shown in Fig. 6. All the experiments of the methods listed in Fig. 6 are conducted on R&G(65). And we only use the concept distance and the concept depth of the nearest common ancestor for measuring concept semantic similarity; this can reduce the time spending on knowledge base searching.

## VI. SUMMARY

This paper combines knowledge-based methods with statistical methods in measuring words similarity, the novel aspect of which is that Subjective Bayes method is employed. The correlation of our experimental results with human judgment is 0.912, with 0.4% improvements over existing best practice, which show that using subjective Bayes method to measure word semantic similarity is reasonable and effective. As we only use concept distance and the concept depth of the nearest common ancestor for measuring concept semantic similarity, this can reduce the time spending on knowledge base searching.

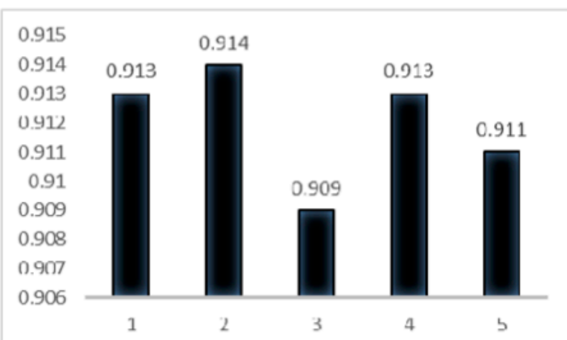


Figure 5. Cross-validation results comparison chart

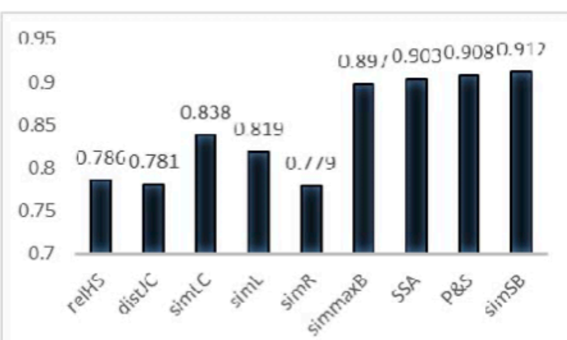


Figure 6. Algorithm results comparison chart for R&G(65)

We hope that our method will be useful in promoting intelligent machine translation, improving information retrieval accuracy, and enabling other tasks that normally require measuring semantic similarity between words. In the future, we will use this computational model of word semantic relatedness in word sense disambiguation.

## ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under grant No.60973040; the Key Scientific and Technological Break-through Program of Jilin Province under grant No.20130206051GX; the National Natural Science Foundation of China under grant No.61300148; and the National Natural Science Foundation of China under grant No.60903098.

## REFERENCES

[1] Zhou Ming, Ding Yuan, and Huang Changning, "Improving Translation Selection with a New Translation Model Trained by Independent Monolingual Corpora," Computational Linguistics and Chinese Language Processing, Vol. 6, No. 1, Feb. 2001, pp. 1-26, doi: 10.1.1.329.8202.

[2] Claudia Leacock, and Martin Chodorow, "Combining local context and WordNet similarity for word sense identification," MIT Press, 1998, pp. 265-283, doi: 10.1234/12345678.

[3] Hany Hassan, Ahmed Hassan, and Ossama Emam, "Unsupervised Information Extraction Approach Using Graph Mutual Reinforcement," Proc. the 2006 Conference on Empirical Methods in Natural Language Processing, ACM, July 2006, pp. 501-508, doi: 10.3115/1610075.1610144.

[4] Cambria, E., Schuller, B., Yunqing Xia, and Havasi C., "New Avenues in Opinion Mining and Sentiment Analysis," Intelligent Systems, Vol. 28, No. 2, March-April 2013, pp. 15-21, doi: 10.1109/MIS.2013.30.

[5] Roy. Rada, Hafedh. Mili, Ellen. Bicknell, and Maria Blettner, "Development and Application of a Metric on Semantic Nets," IEEE Transactions on Systems, Vol. 19, No. 1, Jan/Feb 1989, pp. 17-30, doi: 10.1109/21.24528.

[6] Philip Resnik, "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," Proc. International Joint Conference on Artificial Intelligence, IEEE Press, Aug. 1995, pp. 448-453, doi: 10.1109/CIS.2012.39.

[7] Wu Zhibiao, and Martha. Palmer, "Verbs Semantics and Lexical Selection," Proc. the 32nd Annual Meeting of the Association for Computational Linguistics, ACM, June 1994, pp. 133-138, doi: 10.1.1.14.1869.

[8] Eneko Agirre, and German. Rigau, "A Proposal for Word Sense Disambiguation Using Conceptual Distance," Proc. 1st International Conference on Recent Advances in Natural Language Processing, IEEE Press, Oct. 1995, pp. 35-43, doi: 10.1.1.56.1667.

[9] Jay J. Jiang, and David W, "Conrath. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy," Proc. International Conference on Research in Computational Linguistics, IEEE Press, Sep. 1997, pp. 1-15, doi: 10.1.1.269.3598.

[10] Dekang Lin, "An Information-theoretic Definition of Similarity," Proc. the 15th International Conference on Machine Learning, ACM, 1998, pp. 296-304, doi: 10.1.1.55.1832.

[11] Graeme Hirst, and David St-Onge, "Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms," MIT Press, 1998, pp. 305-332, doi: 10.1.1.50.8426.

[12] Li Yuhua, Zuhair A. Bandar, and David McLean, "An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources," IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No. 4, July-Aug. 2003, pp. 871-882, doi: 10.1109/TKDE.2003.1209005.

[13] Yang Dongqiang, and David M.W. Powers, "Measuring Semantic Similarity in the Taxonomy of WordNet," Proc. the 28th Australasian

Computer Science Conference, ACM, 2005, pp. 315-322, doi: 10.1.1.87.678.

[14] Alexander Budanitsky, and Graeme Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness," Computational Linguistics, Vol. 32, No. 1, March 2006, pp. 13-47, doi: 10.1.1.106.7617.

[15] Marco A. Alvarez, and SeungJin Lim, "A Graph Modeling of Semantic Similarity between Words," Proc. International Conference on Semantic Computing, IEEE Press, Sept. 2007, pp. 355-362, doi: 10.1109/ICSC.2007.23.

[16] Qin Peng, Lu Zhao, Yan Yu, and Wu Fang, "A New Measure of Word Semantic Similarity based on WordNet Hierarchy and DAG Theory," Proc. International Conference on Web Information Systems and Mining, IEEE Press, Nov. 2009, pp. 181-185, doi: 10.1109/WISM.2009.44.

[17] Giuseppe Pirrò, "A Semantic Similarity Metric Combining Features and Intrinsic Information Content," Data & Knowledge Engineering, Vol. 68, No. 11, Nov. 2009, pp. 1289-1308, doi: 10.1016/j.datak.2009.06.008.

[18] Cai Songmei, and Lu Zhao, "An Improved Semantic Similarity Measure for Word Pairs," Proc. 2010 International Conference on e-Education, e-Business, e-Management and e-Learning, IEEE Press, Jan. 2010, pp. 212-216, doi: 10.1109/IC4E.2010.20.

[19] David Sánchez, Montserrat Batet, and David Isern, "Ontology-based Information Content Computation," Knowledge-Based Systems, Vol. 24, No. 2, March 2011, pp. 297-303, doi: 10.1016/j.knosys.2010.10.001.

[20] David Sánchez, Montserrat Batet, David Isern, and Aïda Valls, "Ontology-based Semantic Similarity: A New Feature-based Approach," Expert Systems with Applications, Vol. 39, No. 9, July 2012, pp. 7718-7728, doi: 10.1016/j.eswa.2012.01.082.

[21] Liu Hongzhe, Bao Hong, and Xu De, "Concept Vector for Semantic Similarity and Relatedness based on WordNet Structure," Journal of Systems and Software, Vol. 85, No. 2, Feb. 2012, pp. 370-381, doi: 10.1016/j.jss.2011.08.029.

[22] Ido Dagan, Lillian Lee, and Fernando C. N. Pereira, "Similarity-based Models of Word Cooccurrence Probabilities," Machine Learning, Vol. 34, No. 1-3, Feb. 1999, pp. 135-142, doi: 10.1.1.110.892.

[23] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer, "Word Sense Disambiguation Using Statistical Methods," Proc. the 29th annual meeting on Association for Computational Linguistics, ACM, June 1991, pp. 264-270, doi: 10.1.1.14.1232.

[24] Lillian Lee, "Similarity-based Approaches to Natural Language Processing," PhD thesis, MA: Harvard University, 1997, doi: 10.1.1.136.3516.

[25] Liu Lei, Zhong Maoshang, and Lu Ruzhan, "Measuring Word Similarity Based on Pattern Vector Space Model," Proc. International Conference on Artificial Intelligence and Computational Intelligence, IEEE Press, Nov. 2009, pp. 72-76, doi: 10.1109/AICI.2009.249.

[26] Xu Tao, Qu Weiguang, Tang Xuri, Ding Dexin, Li Bin, and Li Hui, "Computing Word Similarity on Large-Scale Corpus," Proc. 2009 4th International Conference on Innovative Computing, Information and Control, IEEE Press, Dec. 2009, pp. 1076-1079, doi: 10.1109/ICICIC.2009.145.

[27] Kira Radinsky, Eugene Agichten, Evgeniy Gabrilovich, and Shaul Markovitch, "A Word at a Time: Computing Word Relatedness using Temporal Semantic Analysis," Proc. the 20th international conference on World Wide Web, ACM, Mar. 2011, pp. 337-346, doi: 10.1145/1963405.1963455.

[28] Richard O. Duda, Peter E. Hart, and Nils J. Nilsson, "Subjective Bayesian Methods for Rule-based Inference Systems," Proc. the national computer conference and exposition, ACM, June 1976, pp. 1075-1082, doi: 10.1145/1499799.1499948.