Adapted Weighted Graph for Word Sense Disambiguation

Bagus Setya Rintyarna
Informatics Department
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
Muhammadiyah University of Jember
Jember, Indonesia
Email: bagus15@mhs.if.its.ac.id

Riyanarto Sarno
Informatics Department
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
Email: riyanarto@if.its.ac.id

***Abstract*—In Natural Language Processing, Word Sense Disambiguation is defined as the task to assign a suitable sense of words in a certain context. Word Sense Disambiguation takes an important role and considered as the core research problem in computational linguistics. In this research, we conduct an experiment with Adapted Lesk Algorithm compared to original Lesk Algorithm to improve the performance of weighted graph-based word sense disambiguation. Both Algorithms base their measure to the gloss of the dictionary used, not like the other similarity measure that base their measure to the path or information content of the concept being compared. Thus, both Lesk and Adapted Lesk has the highest coverage of part-of speech since they can measure between different part-of-speech. Results of the experiment indicate that Adapted Lesk improves the performance of weighted graph-based Word Sense Disambiguation by 19 % of precision compared to Original Lesk in individual similarity measure experiment.**

Keywords : Natural Language Processing; Word Sense Disambiguation.

I. Introduction

In the field of computational linguistics, Word Sense Disambiguation (WSD) can be defined as the computational task to identify the meaning of words in particular context [1]. The task is important in Natural Language Processing since almost all of English word has more than one meaning. In such field, the word is called polysemy. As the core research problem in computational linguistics, WSD takes an important role to remove ambiguity from text [2] by assigning appropriate sense of a word in a particular context. Moreover, WSD is also essential as a tool for text processing (eg.: machine translation and information retrieval) which is needed to extract information from text data since text left by the people from online activities is an important source of information [20].

According to Aggirre and Edmonds [2], there are four basic approach to WSD : 1) Knowledge-based, 2) Unsupervised corpus-based, 3) Supervised corpus-based and 4) Combination approach. The principal difference between Knowledge-based and corpus-based relies on the lexical resource used in the algorithm. While knowledge based takes advantage of lexical resource which is rich and systematic, corpus-based undertake a knowledge-lean resource which is potential to evolve or adapt as circumstances warrant [3]. Unsupervised corpus-based uses un-annotated corpora while supervised use aligned one [2]. Like in the other fields of study (eg.: data mining, process mining, decision mining), supervised takes advantage from Machine Learning technique [4], [5].

A knowledge-based WSD system employs lexical dictionary [2]. In this research, we propose an Adapted Weighted Graph WSD which uses large english lexical dictionary of Wordnet [6]; therefore this method is a promising knowledge-based wsd which is free of training sample [7]. Wordnet is a lexical database that is considered beyond Machine Readable Dictionary since it arranges concept in a rich semantic network based on psycholinguistic principle [8]. The contribution of this research focuses on the improvement of weighted graph-based WSD method proposed by Sinha and Mihalcea [9] by integrating a better similarity metrics. The similarity metrics to be adapted to Sinha and Mihalceas graph based WSD is Adapted Lesk [10].

We conduct experiment by using Senseval-3 dataset [11] to search the effect of Adapted Lesk Algorithm to the performance of the weighted-graph on WSD. Lesk algorithm is potential since in such WSD approach it is independent of part-of-speech. In other word Lesk is similarity measure with the highest coverage on part-of-speech since it is capable to return similarity measure across different part-of-speech (eg., verb-noun, adjective-noun). The other similarity measure (ie : Leacock and Chodorow, Wu and Palmer, Resnik, Lin, and Jiang and Conrath) can only measure the similarity between the same part-of-speech. Although the use of combined similarity measure outperforms individual one, yet Lesk work better in the use of individual similarity measure. In this paper we will prove that improving Lesk by using Adapted Lesk [10] will improve the entire system of graph-based WSD.

II. Methods

In the field of computational linguistics, Word Sense Disambiguation (WSD) can be defined as the computational task to identify the meaning of words in particular context [1]. The task is important in Natural Language Processing since almost all of English word has more than one meaning. In such field,

ISBN: 978-1-4673-9879-4 (c) 2016 IEEE

the word is called polysemy. As the core research problem in computational linguistics, WSD take an important role to remove ambiguity from text [2] by assigning appropriate sense of a word in a particular context.

The contribution of this research concerns on the improvement of Graph Based WSD model proposed by Sinha and Mihalcea [9]. In such approach of WSD, the weight of the graph is extracted by using several similarity measure (ie : Leacock & Chodorow, Wu & Palmer, Resnik, Lin and Jiang & Conrath). This research proposes to improve the model by improving Lesk Algorithm with Adapted Lesk [10]. Lesk is potential since it is part-of-speech independent so by improving Lesk we argue that the WSD system will perform better. To evaluate the performance of the adapted weighted graph we will use Senseval-3 [11] all-words data set.

Given a sentence S to disambiguate containing n words (w_1, w_2,...,w_n). Each word has several word sense picked from WordNet Database Taxonomy (ws_1, ws_2,...,ws_o). Each word sense become candidate edge in the graph. The next step is extracting weight of the graph by counting similarity measure with several formulae between word sense. The new formula that will be evaluated in this research to be adapted in the graph is Adapted Lesk [10]. Step of the method proposed can be seen in Fig. 1.

To simplify the illustration of the methods, suppose we have a sentence : I like the red camera. Tagging using Stanford POS Tagger resulting part-of-speech of I(PRP)-like(VBP)-the(DT)-red(JJ)-camera(NN). Filtering to select verb, noun and adjective from the sentence resulting a sequence of words (w_1, w_2 and w_3) ie. : like-red-camera. The sense of the word like, red and camera is then picked from WordNet [6] database. Every sense of the word becomes edge of the graph. The similarity measure between word senses is then assigned by several similarity measure. Whenever the similarity between the senses is not zero, an edge is created to connect the node associated with the senses. The value of the similarity is then assigned as the weight of such vertice. The final step is to determine the most important sense by applying connectivity measure of the graph.

The contribution of this research is to evaluate the effect of the improvement of Lesk algorithm to entire WSD method since Lesk outperforms the other similarity measure in individual experiment and is part-of-speech independent [9].

A simple example to illustrate the result of every step of the method can be seen in this description (for the reason to simplify the description of the method we only pick 3 sense out of 5 sense of the word like from WordNet Version 3.0).

Step 1 : I like the red camera

Step 2 :
I(prp) like(vbp) - the(dt) red(jj) - camera(nn)
like-redcamera
like - camera red

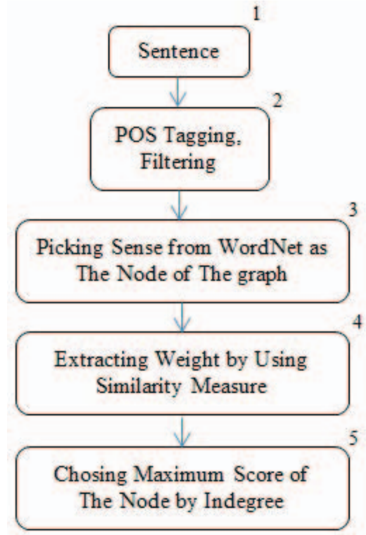

Fig. 1. Step of The Method Proposed

TABLE I
DESCRIPTION OF POS USED IN THE EXAMPLE

Tag	Description
PRP	Personal Pronoun
VBP	Verb in Present Tense
DT	Determiner
JJ	Adjective
NN	Noun, Singular or Mass

Step 3 : Picking Sense from WordNet

TABLE II
SENSE OF THE WORD "LIKE"

Index	Senses	Notation
like#v#1	prefer or wish to do something	ws_1^1
like#v#2	find enjoyable or agreeable	ws_1^2
like#v#3	be fond of	ws_1^3

TABLE III
SENSE OF THE WORD "CAMERA"

Index	Senses	Notation
camera#n#1	equipment for taking photograph	ws_2^1
camera#n#2	television equipment consisting of lens system	ws_2^2

TABLE IV
SENSE OF THE WORD "RED"

Index	Senses	Notation
red#adj#1	of a color of the color spectrum next to orange	ws_3^1
red#adj#2	characterized by violence or bloodshed	ws_3^2
red#adj#3	especially of the face	ws_3^3

Step 4 and 5 : Choosing Maximum Score by Indegree

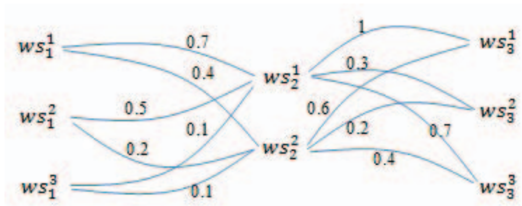

Fig. 2. Graph Extracted from The Example

As an example of the method proposed, the weight of the Graph In Fig. 2 is extracted by using adapted Lesk Algorithm. The complete measure between sense of like, camera and red is presented int Table 4. The result of the measure is then assigned as the weight of the graph as can be seen in Fig. 2. We use Wordnet::Similarity Tools from T. Pedersen and Jason Michellizi to generate most similarity measure used in this research.

TABLE V
RESULT OF MEASURE OF THE SENSE OF THE EXAMPLE

Index	Adapted Lesk Measure	After Normalization
ws_1^1 to ws_2^1	23	0.7
ws_1^1 to ws_2^2	12	0.4
ws_1^2 to ws_2^1	16	0.5
ws_1^2 to ws_2^2	8	0.2
ws_1^3 to ws_2^1	3	0.1
ws_1^3 to ws_2^2	2	0.1
ws_2^1 to ws_3^1	33	1
ws_2^1 to ws_3^2	9	0.3
ws_2^1 to ws_3^3	24	0.7
ws_2^2 to ws_3^1	20	0.6
ws_2^2 to ws_3^2	8	0.2
ws_2^2 to ws_3^3	12	0.4

The final step is to determine the edge with the maximum score by using Indegree Algorithm. Indegree Algorithm counts the value of indegree that is equivalent to the degree of the node. Like was used in [9], for an undirected weighted graph G with node N and edge E, G = (N, E), where w_{ab} is weight between Node Na and Nb, the indegree of node Na is defined as in Equation (1).

$$Indegree(N_a) = \sum_{(N_a,N_b)\epsilon E} w_{ab} \tag{1}$$

For graph in Fig. 2, Indegree Algorithm score is presented in Table 5. The node with the maximum score corresponds to the sense chosen for the word in that context of sentence. From Table 6, we can see that the sense chosen is like#v#1, camera#n#1 and red#adj#1. The bold number indicates the maximum score of Indegree of the sense.

TABLE VI
SCORE OF INDEGREE ALGORITHM

Word	Sense(Edge)	Indegree Score	Sense Chosen
like	ws_1^1	**1.1**	like#v#1
like	ws_1^2	0.7	
like	ws_1^3	0.2	
camera	ws_2^1	**3.3**	camera#n#1
camera	ws_2^2	1.9	
red	ws_3^1	**1.6**	red#adj#1
red	ws_3^2	0.5	
red	ws_3^3	1.1	

A. Similarity Measure

In this chapter we will briefly overview similarity measure used in the experiment. If S is similarity between two concept, Leacock and Chodorow [15] define similarity as in Equation (2) where l is length of shortest path of two concept being measured and D is maximum depth of the taxonomy.

$$S_{lch} = -\log \frac{l}{2*D} \tag{2}$$

Wu and Palmer [16] defines a term of Least Common Subsumer (LCS) to measure similarity S between concept c_1 and c_2 as can be seen in Equation (3).

$$S_{wup} = \frac{2*Depth(LCS)}{Depth(c_1)+Depth(c_2)} \tag{3}$$

While Resnik [17] simply defines similarity S between two concept is equal with information content (IC) of the LCS between both concept.

$$S_{res} = IC(LCS) \tag{4}$$

with

$$IC(c) = -\log P(c) \tag{5}$$

If t_f is term frequeny of a concept, i_f is inherited frequency of a concept, and N is number of term in taxonomy, then

$$P(c) = \frac{t_f+i_f}{N} \tag{6}$$

Lin [18] introduces a similar formula of similarity measure with Wu and Palmer but using information content as in Equation (7).

$$S_{lin} = \frac{2*IC(LCS)}{IC(c_1)+IC(c_2)} \tag{7}$$

While Jiang & Conrath [19] uses both IC and LCS and but with different formula.

$$S_{jcn} = \frac{1}{IC(c_1)+IC(c_2)-2*IC(LCS)} \tag{8}$$

Adapted Lesk Algorithm is the improvement of Lesk Algorithm [12]. In this research, we use Adapted Lesk to improve weighted-graph-based WSD [9]. Original Lesk algorithm picks the gloss from traditional dictionaries such as Oxford Advanced Learners Dictionary of Current English and Websters 7th Collegiate while Adapted Lesk is based on Wordnet. Yet, for the purpose of fair comparation of the effect of Lesk and Adapted Lesk on the WSD method, in this experiment, we will pick the gloss for both algorithms from Wordnet Version 3.0.

The principal difference between Lesk and Adapted Lesk is that Adapted Lesk extents its comparison not just to the sense of words being disambiguate [10]. Adapted Lesk extents its comparison to the sense of words that are connected to the words to be disambiguated in certain relationship defined in WordNet as can be seen in Table 7.

TABLE VII
SCORE OF INDEGREE ALGORITHM

Noun	Verb	Adjective
Hypernym	Hypernym	Attribute
Hyponym	Troponym	Also See
Holonym	Also See	Similar to
Meronym		Pertainym of
Attribute		

B. WordNet

WordNet is an online lexical dictionary [7], [13] that is inspired by psycholinguistics theories of human lexical memory [13]. Thus it arranges its lexical information in the terms of word sense, not simply word form that is organized alphabetically. Rational representation of word sense in WordNet is called synonym sets (henceforth synsets). WordNet can be viewed as a graph and synsets is the node while the semantic and lexical relation between synsets is the edge [8].

As a large lexical dictionary, Wordnet has a wide range of application, not only in the field of Computational Lingusistics and Natural Language Processing, but also in another field of application. In [14] for example, Wordnet is used to support a business process application of a workflow management system.

III. RESULTS

To validate the proposed method, we use Precision, Recall and F-Measure. While precision is percentage of correctly tagged words out of the words addressed by the system, recall is out of all words in the test set. By using Senseval-3 Dataset which contains approximately 5000 words collected from two Wall Street Journal articles and one excerpt from Brown Corpus [11], we compare the proposed method with Sinha and Mihalceas baseline [7]. We conduct several experiment to evaluate the performance of Adapted Lesk on weighted graph WSD method compared to Original Lesk. In this experiment, we extract the weight of the graph by individual similarity measure including Adapted Lesk. This experiment is aimed to evaluate the performance of Adapted Lesk in individual similarity measure compared to the other similarity measure especially Lesk Algorithm in weighted graph WSD scheme. We evaluate the results on every part-of-speech (Noun, Verb and Adjective). The result of experiment is presented in Table 8 to Table 14.

TABLE VIII
RESULT OF THE EXPERIMENT USING LEACOCK & CHODOROW

	Noun	Verb	Adjective	All
Precision	0.25	0.45	0.55	0.39
Recall	0.03	0.08	0.0005	0.001
F Measure	0.05	0.14	0.001	0.002

TABLE IX
RESULT OF THE EXPERIMENT USING WU & PALMER

	Noun	Verb	Adjective	All
Precision	0.29	0.43	0.52	0.45
Recall	0.03	0.08	0.001	0.001
F Measure	0.05	0.13	0.002	0.002

TABLE X
RESULT OF THE EXPERIMENT USING RESNIK

	Noun	Verb	Adjective	All
Precision	0.27	0.5	0.55	0.41
Recall	0.03	0.09	0.0005	0.001
F Measure	0.054	0.15	0.001	0.002

TABLE XI
RESULT OF THE EXPERIMENT USING LIN

	Noun	Verb	Adjective	All
Precision	0.27	0.5	0.55	0.41
Recall	0.03	0.09	0.0005	0.001
F Measure	0.054	0.15	0.001	0.002

TABLE XII
RESULT OF THE EXPERIMENT USING JIANG & CONRATH

	Noun	Verb	Adjective	All
Precision	0.21	0.45	0.55	0.375
Recall	0.02	0.08	0.0005	0.001
F Measure	0.04	0.14	0.001	0.002

TABLE XIII
RESULT OF THE EXPERIMENT USING LESK

	Noun	Verb	Adjective	All
Precision	0.09	0.4	0.38	0.25
Recall	0.01	0.07	0.0004	0.0007
F Measure	0.018	0.12	0.0008	0.001

From the result of the experiment presented in Table 8 to Table 14, we evaluate the performance of WSD that the weight

TABLE XIV
RESULT OF THE EXPERIMENT USING ADAPTED LESK

	Noun	Verb	Adjective	All
Precision	0.27	0.45	0.48	0.44
Recall	0.03	0.09	0.0005	0.001
F Measure	0.054	0.15	0.001	0.002

is extracted by using individual similarity measure (Leacock & Chodorow, Wu & Palmer, Resnik, Lin, and Jiang & Conrath) in every part of speech (Noun, Verb and Adjective). The goal of the experiment is to find the best performance of similarity measure in every part of speech. The result indicates that every similarity measure works best on Adjective. For Noun, Wu & Palmer yield the best performance, while for Verb, Resnik and Lin work best. As Adapted Weighted Graph for Word Sense Disambiguation is an improvement of weighted graph wsd proposed by Sinha and Mihalcea [7] by using Adapted Lesk [10], we try to compare the performance of Adapted Lesk (this research) and Lesk as the baseline method. From Table 11 and Table 12 we notice that Adapted Lesk outperform Lesk by 0.18 on Noun, 0.05 on Verb, and 0.1 on Adjective. Overall, Adapted Lesk outperforms Lesk by 0.19 on precision.

In the future work, we plan to combine similarity measure that achieved the best performance on certain part of speech to get the best performance of the method proposed. We also plan to integrate a better state of the art similarity measure in Sinha and Mihalceaś random walk WSD model to increase the recall as well as another better centrality algorithm instead of Indegree Algorithm.

IV. CONCLUSION

In order to improve the performance of weighted graph WSD, we improve Lesk similarity measure by using Adapted Lesk in random walk wsd framework. Several similarity measures are compared and examined. Experiments indicate that Wu & Palmer work best for noun while Resnik and Lin work best for verb. Almost all similarity measures achieve best performance on adjective. Results of the experiment indicate that Precision of Adapted Lesk outperform Lesk in weighted graph-based WSD method in individual use similarity measure by 19 %.

REFERENCES

[1] R. Navigli., Word Sense Disambiguation : A Survey, *ACM Computing Surveys*, Vol. 41, No. 2, pp. 1-69, doi : 10.1145/1459352.1459355, (2009).

[2] E. Agirre., P. Edmonds., Word Sense Disambiguation : Algorithm and Application, *Springer*, pp. 1-28, doi : 10.1007/978-1-4020-4809-8, (2007).

[3] T. Pedersen., Word Sense Disambiguation : Algorithm and Application, *Springer*, pp. 133-166, doi : 10.1007/978-1-4020-4809-8, (2007).

[4] R. Sarno., P. L. I. Sari., H. Ginardi., D. Sunaryono., I. Mukhlash., Decision Mining for Multi Choiche Workflow Patterns, *International Conference on Computer, Control, Informatics and Its Applications*, pp. 337-342, doi : 10.1109/IC3INA.2013.6819197, (2013).

[5] R. Sarno, R. D. Dewandono., T. Ahmad., M. Farid Naufal., F. Sinaga., Hybrid Association Rule Learning and Process Mining for Fraud Detection, *IAENG International Journal of Computer Science*, 42 : 2, pp. 59-72, (2015).

[6] G. A Miller., Wordnet : A Lexical Database for English, *Communication of the ACM*, (1995)

[7] R. Mihalcea., Knowledge-Based Approach for WSD, Word Sense Disambiguation Algorithm and Application, *Springer*, pp. 107-131, doi : 10.1007/978-1-4020-4809-8, (2007).

[8] R. Navigli., M. Lapata., An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 4, pp. 678-692, doi : 10.1109/TPAMI.2009.36, (2010).

[9] R. Sinha., R. Mihalcea., Unsupervised graph-based Word Sense Disambiguation using Measures of Word Semantic Similarity, *International Conference on Semantic Computing*, pp. 363-369, doi : 10.1109/ICSC.2007.87, (2007).

[10] S. Banerjee., T. Pedersen., An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet, *Computational Linguistics and Intelligent Text Processing, Springer Berlin Heidelberg*, pp. 136-145, doi : 10.1007/3-540-45715-1_11, (2002).

[11] B. Snyder., M. Palmer., The English All-words task : Senseval 3, *Third International Workshop on Evaluation of Systems for the Semantic Analysis of Text*, pp. 41-43, (2004).

[12] M. Lesk., Automatic Sense Disambiguation Using Machine Readable Dictionaries : How to tell a pine cone from an ice cream cone, *Proceeding of The SIGDOG Conference, Toronto*, pp. 24-26, ISBN 0-89791-224-1, (1986).

[13] C. Fellbaum., WordNet : An Electronic Lexical Database,*MIP Press*, 1998.

[14] R. Sarno., C. A. Djeni., I. Mukhlash., D. Sunaryono., Developing A Workflow Management System for Enterprise Resource Planning, *Journal of Theoretical and Applied Information Technology*, Vol 72. No. 3, 2015.

[15] C. Leacock., M. Chodorow., Combinging Local Context and WordNet Sense Similarity for Word Sense Identification, *The MIT Press*, pp. 265-283, ISBN : 9780262272551, (1998).

[16] Z. Wu., M. Palmer., Verb Semantics and Lexical Selection, *Proceeding of the 32nd Annual Meeting of The Association for Computational Linguistics*, pp. 133-138, (1994).

[17] P. Resnik., Using Information Content to Evaluate Semantic Similarity, *Proceeding of 14th International Joint Conference on Artificial Intelligence*, pp. 448-453, ISBN : 978-1558603639 (1995).

[18] D. Lin., An Information-Theoretic definition of Similarity, *Proceeding of the 15th International Conference on Machine Learning*,pp. 296-304, ISBN 1-55860-556-8, (1998).

[19] J. Jiang., D. Conrath., Semantic Similarity based on Corpus Statistics and Lexical Taxonomy, *Proceeding of The International Conference on Research on Computational Linguistics*, (1997).

[20] B. Y. Pratama., R. Sarno., Personality Classification Based on Twitter Text Using Naive Bayes, KNN and SVM, *International Conference on Data and Software Engineering*, pp. 170-174, doi : 10.1109/ICODSE.2015.7436992, 2015