Artificial Intelligence Applications And Innovations Vol. 1 No. 3 (2024)

Article history:
Received 5 March 2024
Revised 20 April 2024
Accepted 03 June 2024
Published online 01 July 2024

A Deductive Word Sense Disambiguation Approach Based on Data Mining and Knowledge Extraction in Expert Systems

Zahra. Pourbahman[1], Niloofar. Rastin[2], Mostafa. Fakhrahmad[3*]

[1] Department of Computer Science, Shahed University, Tehran, Iran
[2] Faculty of Computer Engineering, Iranian E-University, Tehran, Iran
[3] Electronic and Computer Engineering, Shiraz University, Shiraz, Iran

*** Corresponding author email address**: Iranfakhrahmad@shirazu.ac.ir

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Pourbahman, Z., Rastin, N., & Fakhrahmad, M. (2024). A Deductive Word Sense Disambiguation Approach Based on Data Mining and Knowledge Extraction in Expert Systems. *Artificial Intelligence Applications and Innovations, 1*(3), 20-30.
https://doi.org/10.61838/jaiai.1.3.3

A B S T R A C T

Word Sense Disambiguation (WSD) involves assigning the appropriate sense to ambiguous words. WSD is one of the most challenging problems in several Natural Language Processing (NLP) tasks, such as machine translation. This paper proposes a novel approach consisting of four main components. In the first part, a mining process is used to construct a tree structure that represents helpful knowledge about the conceptual relationships between each ambiguous word and its relevant context. In the second part, a Knowledge Base (KB) is constructed based on the chains derived from the tree structure. The third part involves designing an expert system for lexical ambiguity resolution using the forward chaining strategy. In the final part, the KB is upgraded to improve its effectiveness in determining the correct senses of ambiguous words. The performance of the proposed approach is evaluated on the TWA corpus. The results demonstrate the effectiveness of the proposed expert system.

Keywords: *Natural language processing, Expert system, Word sense disambiguation, Lexical ambiguity resolution, Forward chaining.*

1. Introduction

Machine translation is the task of mapping a natural language to another while no human is involved. One major issue in machine translation systems is resolving ambiguity in translation or interpreting the appropriate meaning using linguistic statistics [1].

Each natural language has a considerable set of vocabulary. Since many words have multiple senses, the translation process gets more confusing. The human mind selects the word sense by interpreting the context or considering the available clues [2]. Two different approaches have been proposed to simulate such behaviour in digital translation: rule-based and corpus-based methods.

Rule-based machine translation uses grammatical and linguistic restrictions to translate polysemous words and has been applied in conventional machine translation systems. The main drawback of rule-based approaches is that they demand a significant amount of data and human experts with linguistic knowledge to specify a set of rules, which is expensive in real-world applications.

On the other hand, corpus-based machine translation uses the knowledge automatically extracted from a parallel corpus to translate a polysemous word in the source language to its sense in the target language. This approach may transfer the message inaccurately in the translation process. Corpus-based methods fall into two main categories: example-based and statistics-based methods.

Example-based machine translation uses a bilingual sentence-aligned parallel corpus including sentence pairs (i.e., pairs of source sentences and their translations). Also, it applies the best algorithm in which an input sentence is matched against the examples stored in the corpus by analogy [3].

Statistical-based machine translation uses a mathematical model with parameters estimated from a bilingual corpus. In some statistical translation models, words are treated as independent entities rather than considering their structural connectivity. In such cases, a set of probabilities is evaluated based on the frequency of words to solve the ambiguity resolution problem.

Although example-based and statistics-based techniques often outperform rule-based methods [4], they require the collection, storage, and processing of a large bilingual corpus [5]. Another challenge is the lack of efficient algorithms for extracting knowledge from this large corpus for ambiguity resolution [6].

In this paper, we propose an efficient approach to WSD that does not require a large or bilingual corpus for translation. The proposed approach consists of the following components:

1. Preprocessing: All paragraphs with ambiguous words in their first sense are selected for integration. Similarly, this integration is performed for all paragraphs including the ambiguous words in the second sense. As a result, for every target word, a text file of two long paragraphs is created.
2. Data mining process: A tree structure is created based on the relationship between each ambiguous word and its relevant context.
3. Constructing the KB: The KB includes the chains derived from the previously generated tree structure. Moreover, the degree of importance for each chain is computed and stored in the KB.
4. Lexical ambiguity resolution: An expert system is proposed to determine the meaning of ambiguous words. The reasoning strategy used in the inference engine of the expert system is forward chaining.
5. Enriching KB of the proposed expert system: The KB is updated based on the results achieved in the previous phase.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 introduces the proposed approach. In Section 4, the experiments are reported and explained. Section 5 concludes the paper and describes future work.

2. Related Work

In the method proposed by [7], WSD was done based on the concept, the structure, and the meaning of the words. In this method, WordNet was used as the KB resource to discover the relationships between the target word and its surrounding words in the context. Then, the Naive Bayes classifier was applied to determine the correct sense of the target word.

The method proposed by [8] extracted syntactic and lexical features based on WordNet. They also extracted semantic features using the K-Nearest-Neighbour method to estimate the similarity of represented vectors of the ambiguous words and their contexts based on different similarity measures.

The authors of [9] used a monolingual corpus and built a sub-graph for each ambiguous word based on its context. Then, a similarity function was employed to estimate the most similar sense based on built sub-graphs.

In [10], a WSD method was proposed that could take additional advantage of the sense-labelled examples by exploiting the information obtained from a multilingual representation. They showed that using the features in a multilingual space could improve the performance of a WSD system compared to the traditional systems that use only monolingual features.

The method proposed by [11] extracted lexical features using the English-Chinese bilingual dictionary and employed Vector Space Model to translate ambiguous words in text snippets based on extracted features.

The method proposed by [12] generated clusters using a clustering method on web-scale n-grams. After that, for each context word, the closest cluster was used for WSD. This method captured some semantic relations that were not in WordNet.

The authors of [11] used two windows for context detection: one representing words that occur to the left of the ambiguous word and another for those to the right. They evaluated the frequency of context words to classify each ambiguous word using the Naive Bayes classifier.

The method proposed by [13] used the Collins English Dictionary to connect words to their senses based on their textual definitions. Then, the sense of each ambiguous word was detected based on these relations. The authors of [14] extended this approach by considering semantic relations between ambiguous words and their contexts based on WordNet.

The authors of [15] proposed an approach to model WSD as a simple distributed constraint optimization problem. This approach applied information from various knowledge resources such as Part-Of-Speech tagging to make a model in a multi-agent setting.

The method proposed by [16] presented a WSD method to overcome the feature sparseness problem. To do this, an automatically-created thesaurus was used to append related words to a specific context in order to improve the effectiveness of candidate selection for an ambiguous word.

The method proposed by [17] integrated active learning strategies with a WSD method. In this approach, three different uncertainty sampling-based active learning algorithms were implemented and used with the Support Vector Machine classifier. The proposed approach reduced the number of annotated samples while improving the quality of disambiguation models.

The authors of [18] applied the semantic diffusion kernel to WSD. Semantic diffusion kernel modelled semantic similarity using a diffusion process on a graph defined by lexicon and co-occurrence information.

Recent research in WSD has explored various approaches in the field of deep learning to improve accuracy and address challenges in different languages. Attention-based models, such as stacked bidirectional Long Short-Term Memory networks with self-attention mechanisms, have shown promising results in identifying appropriate word senses [19]. Some studies have focused on refining target word embeddings by modelling the correlation between target context and gloss using multi-head attention mechanisms [20]. Additionally, combining gated-dilated convolutional neural networks with multi-head self-attention has been proposed to extract discriminative features and improve WSD accuracy [21]. Ensemble methods have also been utilized, as in [22], where multiple BERT models are combined to improve disambiguation through pre-trained language models.

While these approaches have shown improvements in accuracy, they also have notable limitations. Most deep learning-based models require extensive labeled datasets, which can be difficult to obtain, especially for low-resource languages. Additionally, many of these models are computationally expensive, requiring significant resources to train and deploy. Furthermore, while ensemble methods like the one in [23] improve performance, they often involve the use of multiple classifiers, adding complexity and reducing interpretability.

In contrast, our approach addresses these limitations by eliminating the need for large-scale labeled training data and the computational burden of multiple classifiers. Instead, we construct a structured knowledge base that iteratively improves through enrichment. This method not only reduces dependency on vast labeled datasets but also provides a more interpretable, efficient, and adaptable solution compared to both traditional ensemble methods and other deep learning-based techniques. By focusing on knowledge base enrichment, our approach fills an existing gap in the literature by offering a scalable and resource-efficient alternative for WSD that balances accuracy with interpretability.

3. Proposed Expert System

This paper introduces an ambiguity resolution system with a new approach to WSD. The details are described in the following subsections.

3.1. *Preprocessing*

The data used in this paper is TWA. It consists of six parts, each representing a bi-sense ambiguous word and including many paragraphs containing the target word. Some paragraphs point to the first sense of the ambiguous word, and the rest refer to the second sense. All paragraphs that correspond to each sense are integrated and constitute a long paragraph. After removing stop words, a bag of words is created for each sense of each ambiguous word. The degree of importance of each word in each bag of words is evaluated as Equation 1.

$$W_t = \frac{TF_{t,BW}}{|BW|} \qquad (1)$$

Where t and BW denote a word and a bag of words, respectively. Moreover, $TF_{t,BW}$ is the term frequency of t in BW.

3.2. Data mining process

As a result of the preprocessing step, there is a bag of words for each sense of each ambiguous word. The mining process aims to create a tree structure for each ambiguous word. Each tree has two subtrees, each corresponding to one of the senses. According to Figure 1, the target ambiguous word and its corresponding senses constitute Level 0 and Level 1 of the tree, respectively. The bag of words related to each sense of the target ambiguous word constitutes Level 2 of the tree. In this step, n words with the highest relevance to the ambiguous target word are selected, based on probabilities in the preprocessing evaluated using Equation 1. For each of these words, a paragraph is extracted from Wikipedia that contains the word. Each paragraph is separately pre-processed and analysed similarly to Level 2.

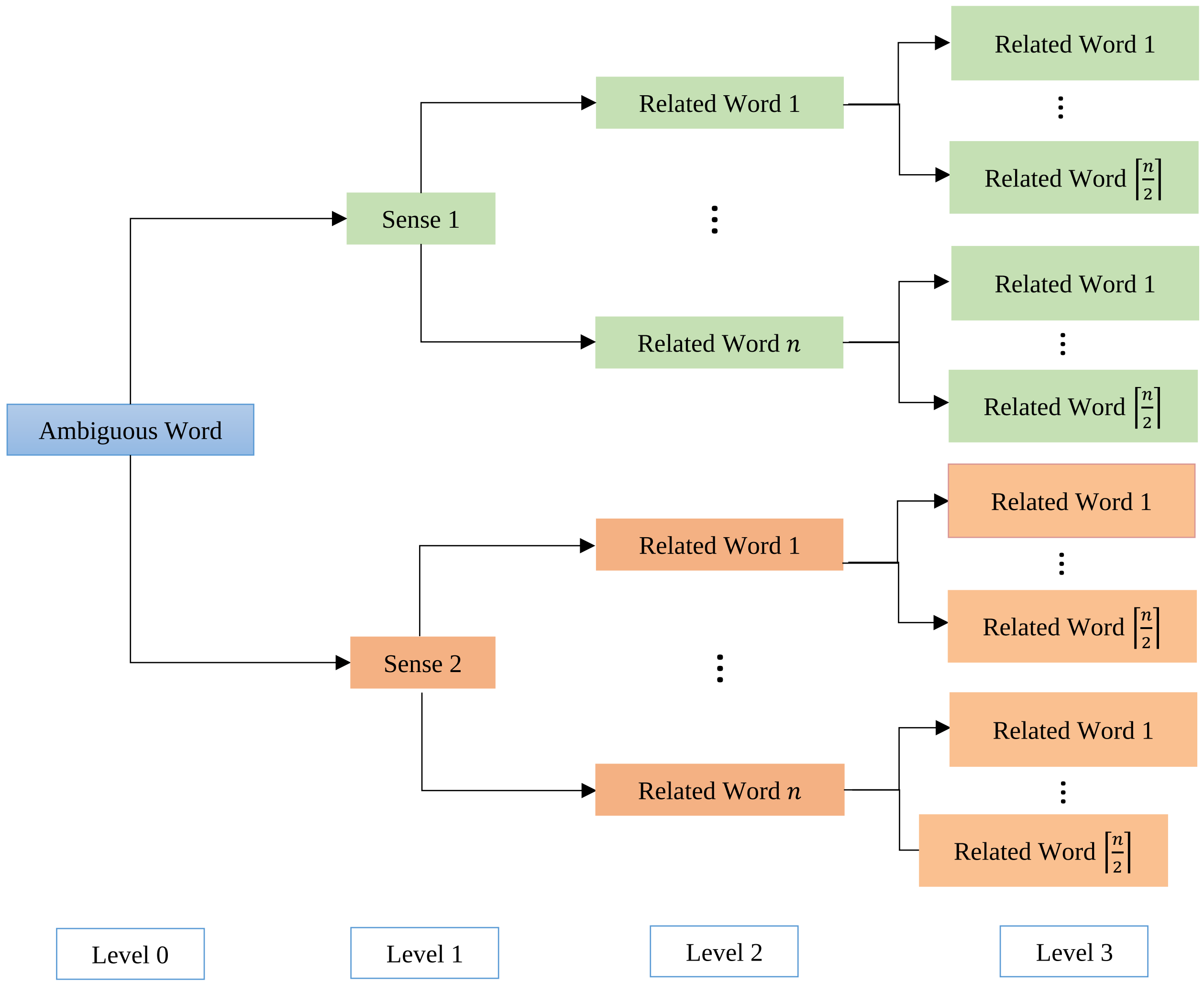

Figure 1. A schematic example of hierarchical structure corresponding to an ambiguous word

However, at this Level, $\lceil \frac{n}{2} \rceil$ words with the highest relevance to the target word are selected. These $\lceil \frac{n}{2} \rceil$ words constitute Level 3 of the tree structure. As a result, a tree structure with two subtrees is generated for each ambiguous word. Each tree is used to construct a KB for an ambiguous word.

3.3. Constructing the KB of the expert system

In this section, each generated tree is decomposed into two sets of chains with length 1, each related to one sense of the target ambiguous word. Each set of chains is inserted into a part of the KB denoted as KB-part 1 and KB-part 2. A degree of importance is assigned to each chain using Equation 1. At this stage, each pair of chains in the KB is compared. According to the transitive relation, if the last element of one chain is equal to the first element of the other, they are connected. The degree of importance of the generated chain is computed as the multiplication of the degree of importance of two primary chains. For instance, let A is the target ambiguous word, A → B (w_1) and B → C(w_2) are two chains with length 1. These two chains are connected to create a chain of length 2, A → B → C ($w_3 = w_1 \times w_2$). This new chain is inserted into the KB. This process repeats until no new chains are generated. So, different chains with different lengths are inserted into the KB.

The first and each subsequent element of each generated chain constitute some new chains with length 1. The degree of importance of the base chain is assigned to these new chains. These new chains, along with their weights, are inserted into the KB. For instance, let A → B → C → D (w_5) is one of the generated chains for the ambiguous word A in the KB. Then, A → B (w_5), A → C (w_5), and A → D (w_5) are inserted into the corresponding part of the KB.

KB-Part 1:

Ambiguous word → Related word 1 (w_1)
Related word 1 → Related word 2 (w_2)
Related word 2 → Related word 3 (w_3)

Generated chains:

Ambiguous word → Related word 1 → Related word 2 ($w_4 = w_1 * w_2$)
Ambiguous word → Related word 1 → Related word 2 → Related word 3 ($w_5 = w_3 * w_4$)

New chains after decomposition:

Ambiguous word → Related word 1 (w_4)
Ambiguous word → Related word 2 (w_4)
Ambiguous word → Related word 1 (w_5)
Ambiguous word → Related word 2 (w_5)
Ambiguous word → Related word 3 (w_5)

Final KB-Part 2:

Ambiguous word → Related word 1 (w_1)
Related word 1 → Related word 2 (w_2)
Related word 2 → Related word 3 (w_3)
Ambiguous word → Related word 1 → Related word 2 (w_4)
Ambiguous word → Related word 1 → Related word 2 → Related word 3 (w_5)
Ambiguous word → Related word 1 (w_4)
Ambiguous word → Related word 2 (w_4)
Ambiguous word → Related word 1 (w_5)
Ambiguous word → Related word 2 (w_5)
Ambiguous word → Related word 3 (w_5)

Figure 2. An example of the content of KB-Part 1 and steps of deduction

In this way, valuable words for WSD are determined, and their degree of importance is automatically increased. An example of this process in KB-Part 1 is presented in Figure 2.

3.4. *Lexical ambiguity resolution*

In this section, the expert system uses the generated KBs for lexical ambiguity resolution. The task of the expert system is to detect the sense of the ambiguous word in a new paragraph based on the existing chains in both parts of the KB (KB-Part 1 and KB-Part 2) corresponding to the ambiguous word. The schema of the proposed expert system is presented in Figure 3.

The inference algorithm of the proposed expert system is presented in Figure 4. According to this algorithm, two scores are evaluated for the test paragraph across two parts of the KB related to the ambiguous word. In line 2, each word in the bag of words of the test paragraph is compared with the last elements of the chains present in each part of the KB. The degree of importance of those chains matched with the word is accumulated with the corresponding score of the KB part (lines 3-8). At last, the sense of the ambiguous word is determined by comparing these scores (lines 9-11).

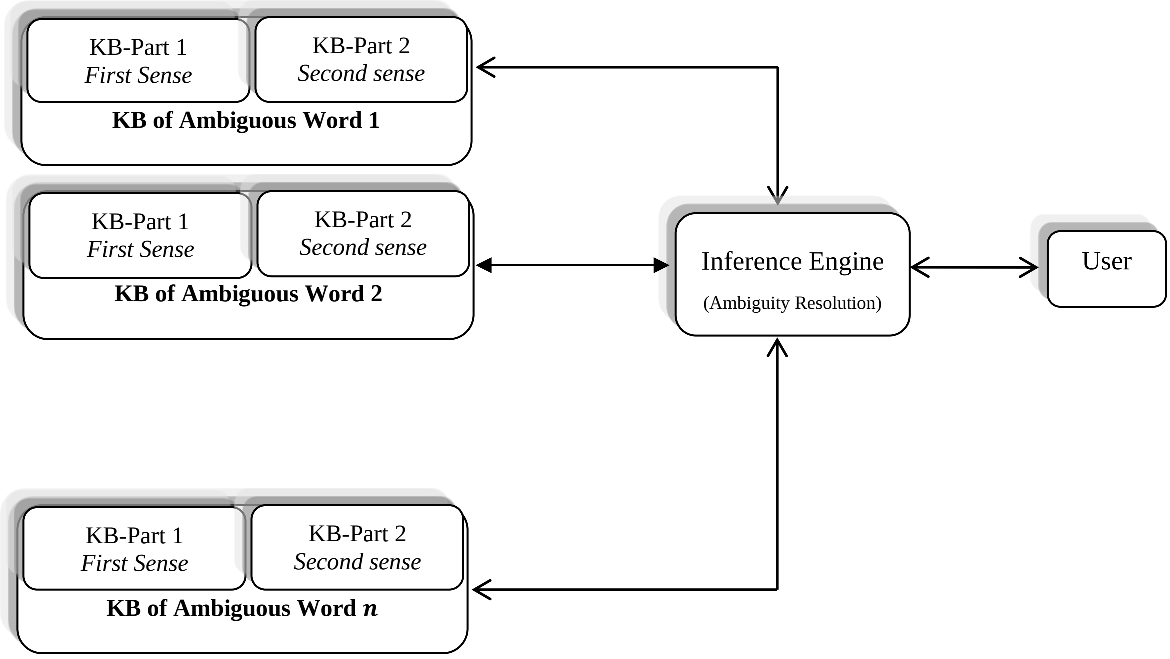

Figure 3. The structure of the proposed expert system used for WSD

variables: $score_1$ *(probability of ambiguous word to be in the first sense)*

$score_2$ *(probability of ambiguous word to be in the second sense)*

w_j *(the weight of* j^{th} *chain)*

t_i *(the* i^{th} *term)*

c_j *(the* i^{th} *chain)*

1: *initialize* $score_1$ *and* $score_2$ *to zero*

2: *for each* t_i *of the test paragraph, p:*

3: *for each* c_j *in KB-Part 1:*

4: *if* t_i *is matched with the last element of* c_j*:*

5: $score_1 = w_j + score_1$

6: *for each* c_j *in KB-Part 2:*

7: *if* t_i *is matched with the last element of* c_j*:*

8: $score_2 = w_j + score_2$

9: *if* $score_1 > score_2$*:*

print "the first sense"

10: *else:*

11: *print "the second sense"*

Figure 4. The proposed algorithm for WSD

3.5. Enriching the KB of the proposed expert system

In this section, a knowledge-learning process is applied to enrich the KB of the expert system. A paragraph containing an ambiguous word is used as the input of the inference engine of the expert system. Following the inference process, the expert system provides the sense of the ambiguous word as its output. If the expert system accurately identifies the meaning of the ambiguous word, new chains of length 1 are extracted based on the words in the input paragraph. The first element of these chains is the ambiguous word and the last element is one word from the bag of words created for the input paragraph. The degree of importance of each new chain is computed based on Equation 1. These chains are inserted into the part of the KB related to the correct sense of the ambiguous word. As a result, the future inferences will be made with more knowledge in the KB. New paragraphs, along with the previous ones, are introduced to the enriched expert system to demonstrate the effect of enriching the KB. The results indicate that enriching the KB improves the memorization and generalization capabilities of the proposed expert system.

4. Experimental Results

The first section of the experiments is dedicated to evaluating the proposed expert system using the TWA corpus. In the second part of the experiments, various test paragraphs from Wikipedia are introduced to the proposed expert system.

Table 1. The accuracy of the different methods evaluated on TWA corpus

Ambiguous word	[7]	[9]	[8]	[10]	[24]	[11]	Proposed Method
Bass	85.49	87.85	88.2	90.65	92.5	92.73	93.85
Crane	87.00	85.26	84.4	75.79	84.2	88.00	84.43
Motion	74.32	84.57	86.2	81.09	78.6	79.21	85.93
Palm	84.21	89.55	87.1	73.13	82.6	84.16	90.21
Plant	67.00	69.14	71.9	79.14	76.6	72.63	65.87
Tank	71.03	70.14	76.1	77.61	79.1	71.29	75.16
Average	76.67	80.16	81.6	78.91	81.1	79.64	82.57

4.1. Experiments on TWA

The proposed expert system is evaluated using the TWA corpus*, which contains six ambiguous words with two senses. For the evaluation, 10-fold cross-validation is employed. The set of all related paragraphs for each ambiguous word is divided into ten equal folds. Nine folds are used to build the KB of the expert system, while the remaining fold is used as test data. The above procedure is repeated 10 times ensuring that each fold is used as test data once.

The average accuracy is reported in Table 1. In this table, the proposed approach is compared with several other methods discussed in the related work section. The highest accuracy values are presented in bold. As can be seen, the proposed approach outperforms all other methods for the "bass" and "palm" ambiguous words. The high accuracy in these cases suggests that the proposed expert system effectively captures and utilizes contextual cues, which aligns with the theory discussed in earlier sections regarding the system's ability to leverage a rich knowledge base and contextual relationships.

Moreover, while the proposed method does not achieve the highest accuracy in every individual case (e.g., for the ambiguous words "plant" and "crane"), it consistently performs well across all cases, ultimately achieving the best average accuracy overall. This indicates that the system is robust and generalizes well across different types of ambiguity. The ability of the proposed expert system to maintain strong performance, even in cases where other methods outperform it for specific ambiguous words, can be attributed to its comprehensive approach, which incorporates both local and global contextual information.

4.2. Experiments Using Wikipedia

Wikipedia is currently used in machine translation [23], text clustering [25], text categorization [26], semantic analysis [27, 28], and query expansion [29]. Therefore, to

*https://web.eecs.umich.edu/~mihalcea/downloads.html#tw have a comprehensive experiment in generalization of the proposed expert system, 4 test paragraphs are extracted from Wikipedia for each of the ambiguous words palm, crane, and bass; 2 related to the first sense and 2 related to the second sense.

To report the results, a column chart is used to compare weights evaluated by the proposed expert system. These charts are shown in Figures 5 to 7. The vertical axis of each chart represents the evaluated weights, while the horizontal axis represents the test numbers. Additionally, the lighter and darker bars indicate the evaluated weights corresponding to the first and second parts of the KB, respectively.

According to Figure 5, the first sense of the ambiguous word palm is hand (represented by lighter bars) and the second sense is tree (represented by darker bars). As can be seen, the proposed expert system correctly disambiguates the word "palm" in 3 out of 4 test cases, demonstrating the system’s capacity to distinguish between the two senses based on the context of the paragraph. This result shows that the system effectively uses contextual clues to disambiguate words.

Similarly, Figure 6 illustrates the system's performance on the word "crane". The first sense of "crane" refers to the bird (represented by lighter bars), and the second sense refers to the machine (represented by darker bars); both are clearly distinguishable. In this case, the expert system correctly disambiguates the word "crane" in 2 out of 4 test cases. While the performance is slightly lower than for "palm", the system still demonstrates its ability to effectively handle context-dependent word sense disambiguation. This aligns with the system's design, which utilizes both local and global contextual information to identify the correct sense.

Finally, Figure 7 presents the results for the word "bass," where the system successfully disambiguates the word in all 4 test cases. In this case, the lighter bars represent the first sense, "fish", while the darker bars correspond to the second sense, "music". The perfect disambiguation rate for this word indicates the system’s robustness and precision. The proposed expert system can accurately differentiate between senses with high consistency when context is sufficiently informative.

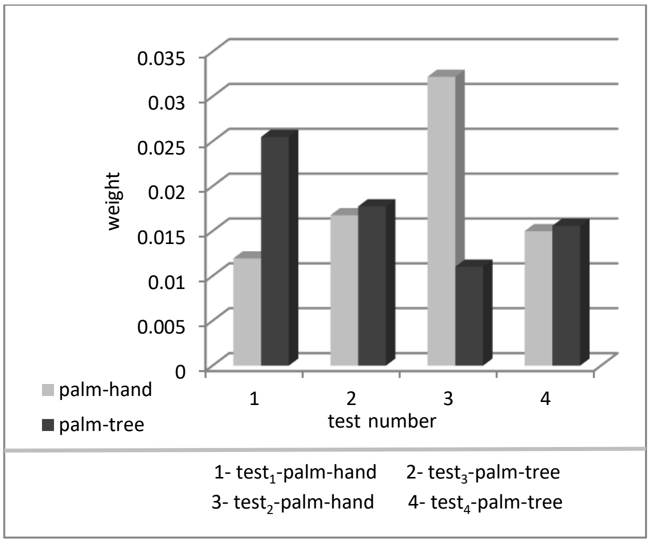

Figure 5. Column chart comparing total evaluated weights for palm:
test$_i$-palm-hand: the ith test paragraph in which palm means hand
test$_i$-palm-tree: the ith test paragraph in which palm means tree

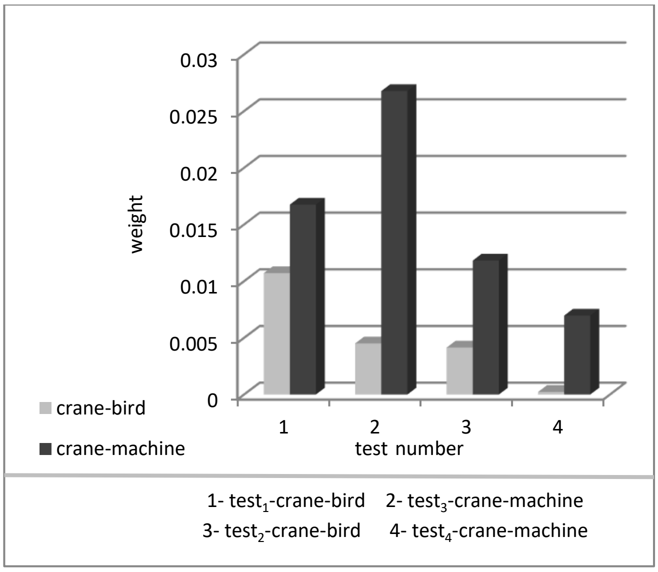

Figure 6. Column chart comparing total evaluated weights for crane:

test$_i$-crane-bird: the ith test paragraph in which crane means bird

test$_i$-crane-tree: the ith test paragraph in which crane means machine

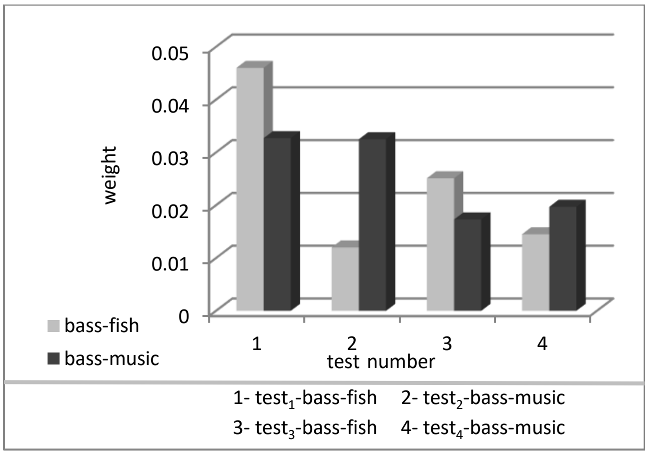

Figure 7. Column chart comparing total evaluated weights for bass:

$test_i$-bass-hand: the i[th] test paragraph in which bass means fish

$test_i$-bass-tree: the i[th] test paragraph in which bass means music

After updating the KB (as described in Section 3.5), the experiments are repeated using 12 recent evaluation tests extracted from Wikipedia. The results are presented in Figures 8 to 10.

In Figure 8, the ambiguous word "palm" is correctly disambiguated in all test cases after the update. Notably, the sense of "palm" in the test paragraph test1-palm-hand is now correctly identified, whereas before the update, the system had misclassified this instance. This demonstrates the efficacy of the KB update in improving the model's performance in handling context-specific word senses. The chart clearly indicates the enhanced accuracy achieved post-update.

Figure 9 shows the results for the word "crane", where the updated KB enables the system to correctly determine the sense of the word in 3 out of 4 test cases. Specifically, before updating the KB, the system failed to disambiguate the word "crane" in both test cases where the first sense, "bird", was intended. After the update, the system correctly identifies the bird sense in the first test instances. This improvement reflects the system's enhanced ability to distinguish between senses when the KB is enriched with more accurate and contextually relevant information, thereby leading to better disambiguation performance.

In Figure 10, the word "bass" shows consistent performance, with the system correctly identifying the word’s sense in all test cases both before and after the KB update. The results for "bass" remain unchanged, which suggests that the system’s ability to disambiguate this word was already robust and did not require significant adjustments after the KB update. This consistency further validates the overall effectiveness of the proposed expert system in identifying word senses, particularly when the word’s context is clear and informative.

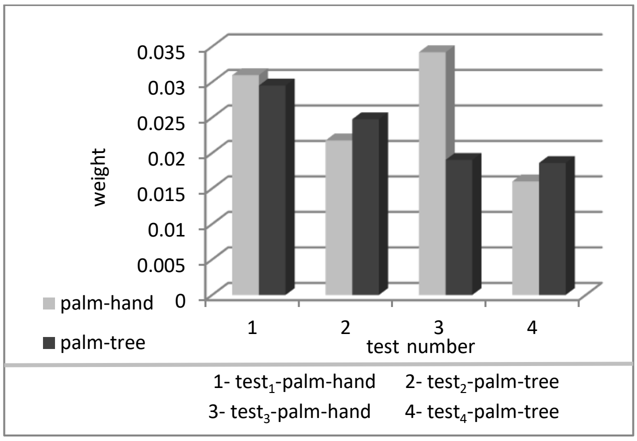

Figure 8. Column chart comparing total evaluated weights after updating KB for palm:

$test_i$-palm-hand: the i[th] test paragraph in which palm means hand

$test_i$-palm-tree: the i[th] test paragraph in which palm means tree

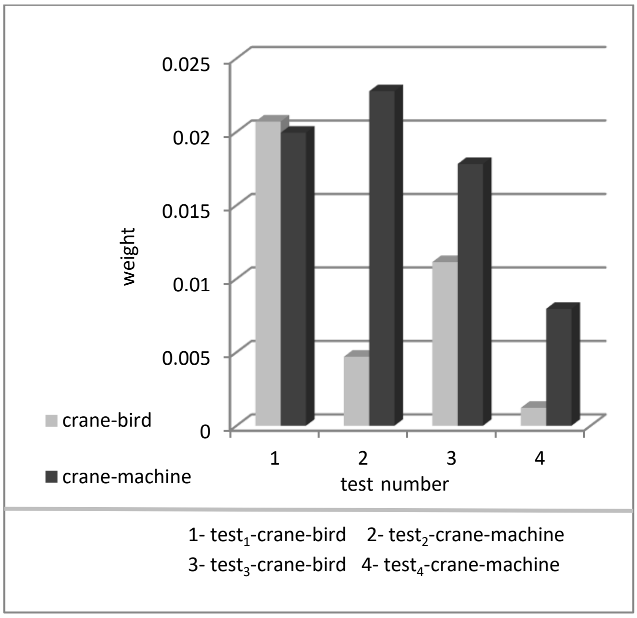

Figure 9. Column chart comparing total evaluated weights after updating KB for crane:

$test_i$-crane-bird: the i[th] test paragraph in which crane means bird

$test_i$-crane-tree: the i[th] test paragraph in which crane means machine

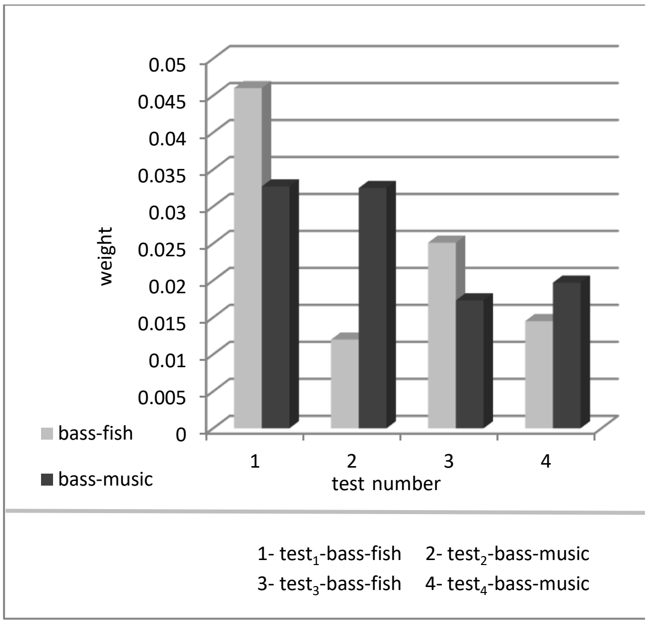

Figure 10. Column chart comparing total evaluated weights after updating KB for bass:
testᵢ-bass-hand: the ith test paragraph in which bass means fish
testᵢ-bass-tree: the ith test paragraph in which bass means music

5. Conclusion

In this paper, we propose a novel word sense disambiguation method that addresses key limitations in current approaches. While existing methods often rely on large bilingual corpora or complex linguistic rules, our approach eliminates these dependencies by constructing a structured knowledge base using a data mining process. This process builds a tree structure to capture the relationships between ambiguous words and their contexts, which is then used to construct the KB. An expert system for lexical ambiguity resolution is introduced, with the KB being iteratively updated to improve accuracy. Our results demonstrate that this approach effectively disambiguates words in a variety of contexts, offering a more resource-efficient and interpretable solution compared to other methods. By eliminating the need for large-scale labelled datasets and multiple classifiers, our method presents a scalable alternative to existing approaches. Future work could explore enriching the KB with new test cases and applying graph structures to represent word-context relationships, which may further improve the system's adaptability and performance.

Authors' Contributions

All authors equally contributed to this study.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors declare that they have no conflict of interest. The authors also declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

The study placed a high emphasis on ethical considerations. Informed consent obtained from all participants, ensuring they are fully aware of the nature of the study and their role in it. Confidentiality strictly maintained, with data anonymized to protect individual privacy. The study adhered to the ethical guidelines for research with human subjects as outlined in the Declaration of Helsinki.

References

[1] D. W. Oard and A. R. Diekema, "Cross-language information retrieval," *Annual Review of Information Science and Technology (ARIST)*, vol. 33, pp. 223-256, 1998. [Online]. Available: https://www.asist.org/publications/arist.

[2] S. M. Fakhrahmad, M. H. Sadreddini, and M. Zolghadri Jahromi, "A proposed expert system for word sense disambiguation: deductive ambiguity resolution based on data mining and forward chaining," *Expert Systems*, vol. 32, no. 2, pp. 178-191, 2015, doi: 10.1111/exsy.12075.

[3] S. D. Samantaray, "Example based machine translation approach for Indian languages," in *Proceedings of ICCS*, 2004, pp. 1-10. [Online]. Available: https://www.iccs-meeting.org/iccs2004.

[4] A. L. Lagarda, V. Alabau, F. Casacuberta, R. Silva, and E. Diaz-de-Liano, "Statistical

post-editing of a rule-based machine translation system," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, 2009, pp. 217-220, doi: 10.3115/1620853.1620913.

[5] F. Ahmed and A. Nürnberger, "Arabic/english word translation disambiguation using parallel corpora and matching schemes," in *Proceedings of the 12th Annual Conference of the European Association for Machine Translation*, 2008, pp. 6-11. [Online]. Available: https://aclanthology.org/2008.eamt-1.

[6] S. Abdul-Rauf and H. Schwenk, "Exploiting comparable corpora with TER and TERp," in *Proceedings of the 2nd Workshop on Building and Using Comparable Corpora: From Parallel to Non-Parallel Corpora (BUCC)*, 2009, pp. 46-54, doi: 10.3115/1690339.1690351.

[7] A. H. Rasekh, M. H. Sadreddini, and S. M. Fakhrahmad, "Word sense disambiguation based on lexical and semantic features using naive Bayes classifier," *Journal of Computing and Security*, vol. 1, no. 2, pp. 123-132, 2014. [Online]. Available: https://sid.ir/paper/341649/en.

[8] A. Rezapour, S. M. Fakhrahmad, and M. H. Sadreddini, "Applying various distance functions and feature extraction schemes to ambiguity resolution," *Intelligent Data Analysis*, vol. 22, no. 3, pp. 617-638, 2018, doi: 10.3233/IDA-173385.

[9] S. Rahmani, S. M. Fakhrahmad, and M. H. Sadreddini, "Co-occurrence graph-based context adaptation: a new unsupervised approach to word sense disambiguation," *Digital Scholarship in the Humanities*, vol. 36, no. 2, pp. 449-471, 2021, doi: 10.1093/llc/fqz048.

[10] C. Banea and R. Mihalcea, "Word sense disambiguation with multilingual features," in *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*, 2011. [Online]. Available: https://aclanthology.org/W11-0100.

[11] H. Ji, "One sense per context cluster: Improving word sense disambiguation using web-scale phrase clustering," in *2010 4th International Universal Communication Symposium*, 2010, pp. 181-184, doi: 10.1109/IUCS.2010.5666225.

[12] W. A. Gale, K. W. Church, and D. Yarowsky, "A method for disambiguating word senses in a large corpus," *Computers and the Humanities*, vol. 26, pp. 415-439, 1992, doi: 10.1007/BF00136984.

[13] J. Veronis and N. Ide, "Word sense disambiguation with very large neural networks extracted from machine readable dictionaries," in *COLING 1990 Volume 2: Papers Presented to the 13th International Conference on Computational Linguistics*, 1990, doi: 10.3115/997939.998006.

[14] G. Tsatsaronis, M. Vazirgiannis, and I. Androutsopoulos, "Word sense disambiguation with spreading activation networks generated from thesauri," in *IJCAI*, 2007, vol. 27, pp. 223-252. [Online]. Available: https://ijcai.org/proceedings/2007.

[15] S. Reddy and A. Inumella, "WSD as a distributed constraint optimization problem," in *Proceedings of the ACL 2010 Student Research Workshop*, 2010, pp. 13-18. [Online]. Available: https://aclanthology.org/P10-3003.

[16] F. Tacoa, D. Bollegala, and M. Ishizuka, "A context expansion method for supervised word sense disambiguation," in *2012 IEEE Sixth International Conference on Semantic Computing*, 2012, pp. 339-341, doi: 10.1109/ICSC.2012.27.

[17] Y. Chen, H. Cao, Q. Mei, K. Zheng, and H. Xu, "Applying active learning to supervised word sense disambiguation in MEDLINE," *Journal of the American Medical Informatics Association*, vol. 20, no. 5, pp. 1001-1006, 2013, doi: 10.1136/amiajnl-2012-001244.

[18] T. Wang, J. Rao, and Q. Hu, "Supervised word sense disambiguation using semantic diffusion kernel," *Engineering Applications of Artificial Intelligence*, vol. 27, pp. 167-174, 2014, doi: 10.1016/j.engappai.2013.08.007.

[19] B. K. Mishra and S. Jain, "Word Sense Disambiguation for Indic Language using Bi-LSTM," *Multimedia Tools and Applications*, pp. 1-26, 2024, doi: 10.1007/s11042-024-19499-9.

[20] X. Zhang *et al.*, "Word Sense Disambiguation by Refining Target Word Embedding," in *Proceedings of the ACM Web Conference 2023*, 2023, pp. 1405-1414, doi: 10.1145/3543507.3583191.

[21] C. X. Zhang, Y. L. Zhang, and X. Y. Gao, "Multi-head self-attention gated-dilated convolutional neural network for word sense disambiguation," *IEEE Access*, vol. 11, pp. 14202-14210, 2023, doi: 10.1109/ACCESS.2023.3243574.

[22] S. Kaddoura and R. Nassar, "EnhancedBERT: A feature-rich ensemble model for Arabic word sense disambiguation with statistical analysis and optimized data collection," *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 1, p. 101911, 2024, doi: 10.1016/j.jksuci.2023.101911.

[23] D. Tufiș, R. Ion, Ș. D. Dumitrescu, and D. Ștefănescu, "Wikipedia as an SMT Training Corpus," in *Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP 2013)*, Hissar, Bulgaria, 2013. [Online]. Available: https://aclanthology.org/R13-1091.

[24] X. Wang and D. Martinez, "Word sense disambiguation using automatically translated sense examples," in *Proceedings of the Cross-Language Knowledge Induction Workshop*, 2006, doi: 10.3115/1608842.1608849.

[25] R. Nayak, R. Mills, C. De-Vries, and S. Geva, "Clustering and labeling a web scale document collection using Wikipedia clusters," in *Proceedings of the 5th International Workshop on Web-Scale Knowledge Representation Retrieval & Reasoning*, 2014, pp. 23-30, doi: 10.1145/2663792.2663803.

[26] N. Bloom, M. Theune, and F. De Jong, "Document categorization using multilingual associative networks based on Wikipedia," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 841-846, doi: 10.1145/2740908.2743003.

[27] A. Saif, N. Omar, M. J. Ab Aziz, U. Z. Zainodin, and N. Salim, "Semantic concept model using Wikipedia semantic features," *Journal of Information Science*, vol. 44, no. 4, pp. 526-551, 2018, doi: 10.1177/0165551517706231.

[28] Z. Wu and C. Giles, "Sense-aware semantic analysis: A multi-prototype word representation model using wikipedia," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2015, vol. 29, 1 ed., doi: 10.1609/aaai.v29i1.9496.

[29] H. Husni, Y. Kustiyahningsih, F. H. Rachman, E. M. S. Rochman, and H. Yulian, "Query expansion using pseudo relevance feedback based on the bahasa version of the wikipedia dataset," in *AIP Conference Proceedings*, 2023, vol. 2679, 1 ed., doi: 10.1063/5.0111273.