Journal of Universal Computer Science, vol. 29, no. 3 (2023), 242-264
submitted: 21/7/2022, accepted: 12/12/2022, appeared: 28/3/2023

A Neuro Symbolic Approach for Contradiction Detection in Persian Text

Zeinab Rahimi
(NLP lab, Faculty of Computer Science and Engineering,
Shahid Beheshti University, Tehran, Iran
https://orcid.org/0000-0001-7553-251X, z_rahimi@sbu.ac.ir)

Mehrnoush Shamsfard
(NLP lab, Faculty of Computer Science and Engineering,
Shahid Beheshti University, Tehran, Iran
https://orcid.org/0000-0002-7027-7529, m-shams@sbu.ac.ir)

Abstract: Detection of semantic contradictory sentences is a challenging and fundamental issue for some NLP applications, such as textual entailments recognition. In this study, contradiction means different types of semantic confrontation, such as negation, antonymy, and numerical. Due to the lack of sufficient data to apply precise machine learning and, specifically, deep learning methods to Persian and other low-resource languages, rule-based approaches are of great interest. Also, recently, the emergence of new methods such as transfer learning has opened up the possibility of deep learning for low-resource languages. This paper introduces a hybrid contradiction detection approach for detecting seven categories of contradictions in Persian texts: Antonymy, negation, numerical, factive, structural, lexical and world knowledge. The proposed method consists of 1) a novel data mining method and 2) a transformer-based deep neural method for contradiction detection . Also, a simple baseline is presented for comparison. The data mining method uses frequent rule mining to extract appropriate contradiction detection rules employing a development set. Extracted rules are tested for different categories of contradictory sentences. In the first step, a classifier checks whether the rules work for an input sentence pair. Then, according to the result, rules are used for three categories of negation, numerical, and antonym. In this part, the highest F-measure is obtained for detecting the negation category (90%), the average F-measure for these three categories is 86%, and for the other four categories, in which the rules have a lower F-measure of 62%, the transformer-based method achieved 76%. The proposed hybrid approach has an overall f-measure of higher than 80%.

Keywords: Contradiction detection, data mining, BERT-based method, hybrid method, Persian text processing
Categories: I.7, I.2, I.5
DOI: 10.3897/jucs.90646

1 Introduction

Contradiction detection is a fundamental task in text understanding and has many possible applications, especially in textual inference and sentiment analysis. The contradiction relation represents the semantic confrontation between two entities and, in general, includes all antonymy, negation, and other semantic confrontations such as world knowledge, lexical, numeric, and structural. The words “contradiction,” “contrast,” “confrontation,” and “conflict” in this domain are sometimes used interchangeably, despite their differences. The concept of contradiction has different definitions and is categorized from different points of view, which is stated in details in section 1.1.

1.1 Types of contradiction

Contradiction usually occurs on two levels: word and structure. Table 1 shows different semantic contradiction types in word level [Safavi, 00]:

#	TYPE	DEFINITION	EXAMPLE
1	Negation	Refers to the cases where the negation of one is proof of the other. This type of contradiction is called a complementary contradiction	“man” and “woman,” “on” and “off,” or “open” and “closed”
2	Gradable adjectives	Refers to the words in a value spectrum of a specific attribute. Negating one is not a proof of the other.	“Cold-cool-warm-hot”
3	scalar contrasts	Statement of one does not necessarily mean the opposition of the other	“Cold and hot,” “old and young,” “big and small,” “short and long,” or “ugly and beautiful” (The statement “she is not ugly” does not necessarily mean that she is beautiful.)
4	Double-confrontation	Refers to the cases where a two-way relationship exists between two words. This means that if person A bought something from B, it is likely that B has sold it to A. Or, if A is the husband of B, then certainly B is A’s wife.	a couple (wife and husband) or buyer and seller
5	Lexical opposition	Where a negative affix is used. The proof of one is usually the negation of the other, and vice versa;	knowingly and unknowingly, polite and impolite, or safe and unsafe (“he is not a polite man,” means, “he is impolite.”)
6	Directional opposition	Where a reference point in time or space is considered, and a pair of words are measured based on that point	“Coming and going,” “up and down,” “bringing and taking,” “backward and forward,” “left and right,” or “north and south” (“Going” is essentially moving away from a point, and “coming” is getting closer to it.)

Table 1: Types of semantic contradiction at the word level [Safavi, 00]

These types are commonly attributed to words without considering their contexts. But in most cases, the focus is on the detection of contrast between sentences which occur in the form of different types of contradictions.

As mentioned before, the opposition (or antonymy) predominantly stands between two words or phrases where the negation of one is a further proof for the other, such as "on" and "off," or words or phrases created with negation affixes such as "accurate" and "inaccurate," while contradiction is mainly between two sentences or pieces of text, where two sentences are extremely unlikely to be true at the same time, such as "Mary was killed yesterday" and "Mary is having lunch at the restaurant now."

Table 2 shows different types of confrontation in textual entailments (a categorization for structure level contradiction) which is created in Marneffe research [De Marneffe, 08]:

#	TYPE	DEFINITION	EXAMPLE
1	Antonymy	The cases where antonym words appear in similar semantic roles in two sentences.	"Capital punishment is a catalyst for more crime" and "Capital punishment is a deterrent to crime"
2	Negation	Two sentences are contradictory with the negation of a position, using the negation sign	"a closely-divided Supreme Court said that judges and not judges must impose a death sentence" and "the Supreme Court ruled that only judges can impose the death sentence."
3	Numerical	Contradiction is due to a numerical contrast in two sentences	"More than 100 people were killed in the war" and "50 people were killed in the war"
4	Factive	A contradiction occurs when the manner of expressing an event or action in the first sentence creates an assumption in mind whose opposite is being stated in the second sentence	"The thieves did not intend to enter the bank" and "the thieves entered the bank,"
5	Structural	A conflict exists due to changes in the components of a relationship. That is, the semantic structure of the relation of the first sentence has changed in the second sentence, and the meaning has changed, but syntactically and without considering the first sentence, the second sentence may be correct.	"The Channel Tunnel stretches from England to France. It is the second-longest rail tunnel in the world, the longest being a tunnel in Japan" and "the Channel Tunnel connects France and Japan."
6	Lexical	A word or phrase is the opposite of the word or phrase in the second sentence. While these phrases are not necessarily opposite, they are contradictory in the context of these two sentences.	"The Canadian Parliament's Ethics Commission said the former immigration minister, Judy Sgro, did nothing wrong, and her staff had put her in a conflict of interests" and "the Canadian Parliament's Ethics Commission will accuse Judy Sgro."
7	World knowledge	two sentences contradict the background knowledge that they convey.	"One of the first Microsoft branches outside the USA was founded in 1989" and "Microsoft was established in 1990."

Table 2: Types of confrontation in textual entailments [De Marneffe, 08]

The categories 1, 2 and 3 in Table 2, have been studied in most papers. The last four categories, to the authors' knowledge, are less considered.

1.2 Paper focus

Although the existing methods are useful in recognizing most word-level oppositions, the differences in structural patterns of opposition in the Persian language necessitate their change, adaptation, and consideration within the scope of current research. Examples of these differences are given below.

- When using quantifiers to negate sentences in English, there is a negative quantifier and a positive verb, while in the same sentence in Persian a negative verb and a negative quantifier are used.

Example:

English: "no one came"

Persian: "هیچ کس نیامد" (no one did not come)

- When using negatives adverbs, the same pattern exists; a negative adverb and a positive verb in English, and a negative adverb and a negative verb in Persian.

Example:

English: "never"

Persian: "هرگز"

- When using the terms "neither"[1] in some sentences in Persian, a negative verb is used in both sentences, while in English, the verbs are positive. In addition, in Persian, the adverb ("هم") is the same for both positive and negative states, while in English, "neither" and "either" are used for negative and positive sentences, respectively.

Example:

English: "Mary does not like this food. Neither do I"

Persian: "مریم این غذا را دوست ندارد. من هم دوست ندارم" (Mary does not like this n food. I do not like it too.)

The contributions of this paper are listed as following:

1- Studying and considering the language specific contradiction Patterns for the Persian language and proposing a data mining method to automatically extract these patterns and rules.
2- Proposing a rule-based method according to the extracted patterns to recognize some sorts of contradictions in Persian sentences.
3- Providing a dataset for training neural models on contradiction detection task.
4- Proposing a BERT[2]-base deep learning method to handle some contradictory cases which are not properly recognizable through rules.
5- Presenting a classifier to determine which method is appropriate for each sentence pair.

The proposed total approach is a neuro-symbolic approach as it is a combination of neural methods (deep learning) and symbolic methods (data mining and rule-based detection). None of the two approaches are enough and can handle every type of

[1] هم

[2] Bidirectional Encoder Representations from Transformers

contradiction on its own but the combination of these two methods provides the best results for identifying contradictions in Persian language.

Due to this regard, we implemented our approach in the form of a combined system that includes two subsystems: 1) based on data mining and 2) based on deep learning. This hybrid algorithm is applied on input sentence pairs to identify the possible contradiction, and if the rules (extracted by the DM subsystem) are not matched, the decision is made by the deep learning subsystem.

The rest of the paper is organized as follows. Section 2 reviews related research. In Section 3, the proposed methods are explained in detail. In section 4, evaluation is discussed. Finally, section 5 presents the conclusion and suggestions for further work.

2 Related Work

There are few studies on the raw subject of automatic contradiction detection. There are some contradiction detection methods embedded in various applications of natural language processing, such as sentiment analysis and textual entailment. The proposed methods in related research can be classified into three categories:

1- Rule-based methods, including [Harabagiu, 06], [De Marneffe, 08], [Blanco, 11], [Asmi, 12], and [Pham, 13]
2- Machine learning and deep learning methods, including [Rocktäschel, 15], [Wang, 15], [Khandelwal, 19], and [Sifa, 19].
3- Other approaches, including [Shih, 12], [Vargas, 17], and [Li, 17].

In this section, related research and its details are presented in the following subsections.

2.1 Rule-based methods

According to the work of De Marneffe [De Marneffe, 08]), Harabagiu et al. in 2006 provide the first empirical results for contradiction detection. They consider three categories of negation, antonyms, and semantic information associated with contrast. It has been argued that this information can detect incompatible information (such as two conflicting answers to a question in a Q&A system) or identify compatible information (semantic similarity, redundancy, and textual entailment). The goal is to find contradictory information in the text, and a framework for identifying contradictions is proposed that addresses 3 types of negation, antonymy, and contradiction. Two approaches are considered in detecting negation: (1) direct negation (not, negative quantifiers such as no, no one, nothing, and negative adverbs like never) and (2) indirect negation (verbs such as deny, fail, refuse, prepositions like without, weak quantifiers like few, any, some, and cases of traditional negative polarity like anymore).

As discussed, De Marneffe [De Marneffe, 08] classifies contradiction into antonyms, negation, numerical, factive, structural, lexical, and world knowledge and argue that all these can be, in turn, classified into two categories: (1) negation and antonyms, the mismatch between date or number, and (2) the contradictions derived from the use of modal words, factive, lexical, and knowledge-based contradictions. In their study, they consider the first category in the following steps.

- Language analysis: Representation of a language to display text content.

- Graph alignment: Using Stanford parser, the text and the hypothesis are converted to dependency graphs and aligned.
- Filter non-coreferencing events.
- Extraction of contradictory features (including polarity, numbers and date contradiction, and structure).

Blanco [Blanco, 11]) analysed several cases about identifying negations in texts and their scope. The main idea is that in the detection of negation, two issues are important: (1) scope and (2) focus. The scope is part of the meaning that is negated and is a territory for the affected area. Based on that limitation, the negated or non-negated parts of the sentence are determined. For example, the sentence "all vegetarians do not eat meat" goes back to the entire vegetarian community, but the sentence "all plants are not eaten by vegetarians" does not mean that vegetarians just eat plants. The second issue, the focus, is about the part of the sentence which is the point of attention and how it can be negated. For example, in the sentence "that land was not large, it was huge," the magnitude of the land is worth considering. In their study, Blanco and Moldovan proposed rules for different categories of negation and added them to the representation of a negated sentence.

SankaraSubramanian in 2009 [SankaraSubramanian, 09] proposed a fuzzy rule-based algorithm with the following steps.

- Preprocessing and removing additional characters and reformatting abbreviations, numbers, and characters;
- Specifying the keywords of the document and initializing zero to three classes of aligned, conflicting, and unrelated cases;
- Comparing keywords and increasing the value of each of the three above variables in case of occurrence;
- Calculating the final value of variables and taking an initial decision;
- Calculating the matching ratio (the number of conflicting words from the sum of words, aligned and non-correlated) and taking the final decision based on the thresholds and the obtained ratio.

Asmi [Asmi, 12] investigated the identification of negation in a sentiment analysis system. In this regard, sentences that are negated are identified using a dependency parser, and the polarity of the negated sentences is calculated by a set of rules extracted from the senti-wordnet.

As mentioned, some contradiction detection works are embedded in applications of natural language processing, such as textual entailment. In the textual entailment task, there are two text fragments called the 'Text' and 'Hypothesis,' and the goal is to determine whether the meaning of the hypothesis is entailed in (can be inferred from) the text or the pair is contradictory or neutral. [Pham, 13] extracted a combination of shallow semantic representations using semantic role labelling and binary relations from sentences by a rule-based method. Initially, after the syntax analysis using the coreNLP[3] library, the Senna tool is used to label semantic roles. Then, using the REVERB tool, binary relations are extracted from sentences. This tool takes a piece of text with POS tags as input and creates output triples in the form (argument 1, relationship, argument 2). In the process of detecting the contradiction, two steps are taken to detect the conflicts of the frames and the contrast of the relationships. In the first step, the verbs of the text and

[3] https://stanfordnlp.github.io/CoreNLP/

the hypothesis are compared using sources such as VerbNet and VerbOcean and are placed in one of the matching, conflict, or non-related categories. Also, semantic frames of the text and hypothesis are scored using a conflict function according to the inconsistency of their events. In the second step, some extracted relations from the text and the hypothesis are compared, and decisions are made.

2.2 Machine learning based methods

Some researchers consider the contradiction detection task as a part of textual entailment, mainly after publishing of datasets such as SNLI (refer to section 2.2), which has contradiction tags as well as entailment ones. So the RTE[4] methods, which use classification algorithms and deep learning methods [Rocktäschel, 15] [Wang, 15], can be considered as related work in this section.

In 2015, Rocktäschel [Rocktäschel, 15] proposed a neural model that receives two sentences and detects entailment using LSTM (long short-term memory) units. The researchers extended the model with a word-by-word neural attention mechanism that encourages reasoning over entailments of pairs of words and phrases. The paper did not directly report the system performance for the contradiction class, but because of the novel idea and good overall performance of the system, the proposed network is a basis for many later studies.

Wang in 2015 [Wang, 15] proposed a special long short-term memory (LSTM) architecture for NLI[5]. Their model is built on a neural attention model for NLI but is based on a different idea. Instead of deriving sentence embedding for the text and the hypothesis for classification, they use a match-LSTM to perform word-by-word matching of the hypothesis with the text. The authors claimed this LSTM places more emphasis on important word-level matching results and reported promising outcomes in both entailment and contradiction classification tasks.

Lingam in 2018 [Lingam, 18] proposed an approach for detecting three different types of contradiction: negation, antonyms, and numeric mismatch. They derived several linguistic features from text and used them in a classification framework for detecting contradictions with artificial neural networks and deep learning techniques such as long short-term memory (LSTM) and Global Vectors for Word Representation (GloVe).

Tawfik and Speruit [Tawfik, 18] introduced an automated two-phase contradiction detection model that integrates semantic properties as input features to a Learning-to-Rank framework to identify key findings of a research article. Inconsistencies in the text were identified by a two-phase algorithm, claim retrieval and claim assertion. In the first phase, potential sentences relevant to the query were identified, and in the claim assertion phase, they evaluated whether sentences infer text entailment or contradiction. It also relies on negation, antonyms, and similarity measures to detect contradictions between findings.

Khandelwal [Khandelwal, 19] explored the decision choices for negation detection and the involved scope resolution using the popular transfer learning model BERT on three corpora: the BioScope Corpus, the Sherlock dataset, and the SFU Review Corpus. They reported state-of-the-art results for the scope resolution across all three datasets.

---

[4] Recognizing Textual Entailment

[5] Natural Language Inference

Sifa [Sifa, 19] examined a set of baseline methods for contradiction detection in German texts. For this purpose, the well-known Stanford Natural Language Inference (SNLI) data set (110,000 sentence pairs) was machine-translated from English to German. They trained and evaluated four classifiers on both the original and the translated data, using state-of-the-art textual data representations. Their chief contribution was the first large-scale assessment of this problem in German and the validation of machine translation as a data generation method.

Gao in 2021 [Gao, 21] proposed a novel model adaptation paradigm called adapting by pruning, which prunes neural connections in the pre-trained model to optimize the performance on the target task, with all remaining connections having their weights intact. They formulated adapting-by-pruning as an optimization problem with a differentiable loss, proposed an efficient algorithm to prune the model, proved that the algorithm is near-optimal under standard assumptions, and applied the algorithm to adapt BERT to tasks such as natural language inference (including contradiction class).

Tarunesh [Tarunesh, 21] created a test set (184K examples) for the Natural Language Inference task and benchmark state-of-the-art NLI systems on this set, which revealed fine-grained insights into the reasoning abilities of BERT and RoBERTa.

2.3 Other approaches

Shih in 2012[Shih, 12] presented knowledge scarcity as a challenge for contradiction extraction. In this regard, the authors used a web query to measure the frequency of phrases with a non-matching relationship and analysed the adaptation degree of these non-matches to the existence or non-existence of a contradiction.

Vargas [Vargas, 17] proposed a sentiment-based contradiction detection system that assumes oppositions as antonymy or opposite sentiments according to a unique aspect or attribute. In their approach, opposite sentiments according to a unique aspect or attribute. In their approach, the topics or attributes to which a sentence refers to are extracted. Then for each input sentence pair, the sentence polarity is measured for each extracted topic. If the polarity for a single topic differs in two sentences and their calculated similarity is lower than a threshold, the sentence pair is tagged as contradictory.

Li in 2017 [Li, 17] constructed a contradiction-specific word embedding (CWE). In this method, antonym and negation-based contrasts are used to create artificial contradictory sentences and use them as a training set to form a contradiction word embedding space. Afterward, the embedding is used to detect contradictions.

2.4 Persian related work

The work on the automatic recognition of contradiction in the Persian language is very limited. To the author knowledge, most works have been done in theoretical linguistics frameworks. In computational side, there is a research devoted to the negation problem in the context of opinion mining [Noferesti, 16] that uses weighted rule mining for finding patterns representing sentiment shifters from a domain-specific corpus.

In another work, Khodadadi and colleagues [Khodadadi, 15], use discourse signs in Persian textual corpus to recognize the contradiction, and a machine learning system is trained. This system uses discourse signs such as “but” to detect contradiction relations in one sentence.

Recently, Amirkhani and coleauges [Amirkhani, 20] presented a new dataset for NLI in the Persian language named FarsTail, which contains data on contradiction as well. They also provided the results of traditional and the state-of-the-art methods on FarsTail, including different embedding methods such as word2vec, fastText, ELMo, BERT, and LASER, as well as modeling approaches such as DecompAtt, ESIM, HBMP, ULMFiT, and cross-lingual transfer approach to provide a solid baseline for future research.

2.5 Related datasets

There are a few manually tagged datasets for contradiction detection, which include:

- The RTE[6] competition datasets

These datasets have been prepared annually for RTE competitions from 2005 to 2010 and include training, testing, and development subsets. The competitions have continued under TAC[7] and SemEval since then. RTE1 to RTE6 datasets have in total approximately 35000 sentence pairs manually labelled with three categories: "YES (entailment), NO (contradiction), and unknown."

- SICK[8] corpus

This collection consists of 10,000 English sentence pairs from two sources: ImageFlickr and SemEval2015 video descriptions, manually labelled with three categories: "Entailment, Contradiction, and Neutral."

- Stanford University SNLI[9] corpus

This corpus is a collection of 570k human-written English sentence pairs, manually labelled with three categories: "Entailment, Contradiction, and Neutral."

- The Multi-Genre Natural Language Inference (MultiNLI[10]) corpus is a crowd-sourced collection of 433k sentence pairs annotated with textual entailment information. The corpus is modelled on the SNLI corpus but differs in some ways; it covers a range of genres of spoken and written text and supports a distinctive cross-genre generalization evaluation.

- ES-Cn is an annotated contradiction dataset in Spanish within the news domain, where sentences are classified as compatible, contradictory, or unrelated information. The dataset consists of 7403 news items, of which 2431 contain compatible headline–body news items, 2473 contain contradictory headline–body news items, and 2499 are unrelated headline-body news items. Presently, four different types of contradictions are covered in the contradiction examples: negation, antonyms, numerical, and structural [Sepúlveda-Torres, 21].

- FarsTail[11] contains 10,367 samples provided in the Persian language and the indexed format to be useful for non-Persian researchers. The samples are generated from 3,539 multiple-choice questions with the least amount of annotator interventions, similar to the SciTail dataset. A multi-step process is adopted to ensure the quality of the dataset. The tag set consists of three common labels: contradiction, entailment, and neutral [Amirkhani, 20].

---

[6] Recognizing Textual Entailment
[7] Text Analysis Conference
[8] Sentences Involving Compositional Knowledge
[9] The Stanford Natural Language Inference
[10] https://www.nyu.edu/projects/bowman/multinli/
[11] https://github.com/dml-qom/FarsTail

Using ontologies and knowledge bases	**Opposite words ratio (from what adopted)**	**Using conflict feature set (TMP, LOC, sentiment ...)**	**Semantic role labelling**	**Sentiment analysis**	**Neural network, deep learning and word embedding**	**Rule-based**	**Algorithm /Contradiction type**
	*			*	*	*	antonymy
	*			*	*	*	negation
		*			*	*	numeric
*					*		Factive
			*		*		structural
*		*		*	*		lexical
*				*	*		World knowledge

Table 3: Algorithms presented for seven categories of contradiction [De Marneffe, 08]

After reviewing the contradiction types in section 1 and reviewing existing related work in Section 2, we will introduce our dataset and discuss and compare the three proposed methods (baseline, data mining, and Bert-based) for recognizing different types of contradiction in the next section.

3 The Proposed methods

To the authors' knowledge, the best results for contradiction detection systems are now obtained through machine learning and deep learning systems (such as work of Gao in 2021 [Gao, 21]. However, in Persian and possibly other low-resource languages [Shamsfard, 19], due to the lack of appropriate data with sufficient volume for system training, these methods were not applicable until recent research. Therefore, this study first focuses on introducing a practical approach based on general and specific rules in the Persian language. Secondly, for better results in the detection of some contradiction types that could not be achieved with rules, a medium-size contradiction dataset was created through the translation of the existing English corpora (both manual and machine-translation methods). Details of this dataset is stated in section 4.1. Then, a BERT-based model was trained using this generated corpus.

The proposed hybrid system consists of a rule-based contradiction subsystem and a BERT-based deep learning subsystem. We created two rule-based methods: basic and advanced. From now on the first method is called Basic-R and the second is called DM-R method. This Basic-R is based on a series of general features to identify semantic contradiction. The DM-R applies a data mining method (frequent rule mining) to the development set and automatically discovers the distinctive features of contradiction for the predefined contradiction categories. Details of these three systems are described in the following subsections.

3.1 The basic rule-based method (Basic-R)

In this section, the proposed rule-based baseline is introduced. So far, several rules and attributes, such as those in Marneffe research [De Marneffe, 08], have been used to find contradictions. In the Basic-R, a wide range of syntactic and semantic features are considered and used to solve the problem of contradiction detection, no new feature is introduced, and only some feature values are integrated and used as a comprehensive ruleset. Before feature extraction, a pre-processing procedure is applied to the sentence pairs.

In the pre-processing phase, the following tasks are performed.

- POS tagging with a 100-tag tag set performed with 92% precision by training a tagger over the Peykareh[12] corpus. [Bijankhan, 11]
- Name entity recognition (NER) with a seven-tag tag set performed with an accuracy of 85% by training a CRF model over the NER corpus of UT[13][Shahshahani, 19].
- Dependency parsing, performed with an approximate precision of 85% by training a model on the Dadegan[14] corpus [Rasooli, 13] using Malt-parser.
- Semantic role labelling, performed with an approximate precision of 75% by training a labeller on the Dadegan-SRL data [Mirzaei, 14].

The implemented features, which are eventually used as the rule set in the Basic-R, are as follows:

1- Sentiment agreement; Experience has proven that many contradictory sentence pairs have different polarities, so the feature of the sentiment dis/agreement of two sentences is considered a distinguishing feature. For example, “John slept very well last night” and “John had a nightmare last night” have opposite sentiments. For sentiment analysis, a simple lexicon-based system based on sentistrenght[15] lexicons are implemented and used. The lexicon consists of approximately 900 polar words. For calculating the sentiment score, the occurrence of polar words or their stems along with the negation word list is considered.
2- Named entities comparison; sometimes, the existing named entities in contradictory sentences are different, particularly the most famous ones such as location, date, or time. Therefore, this feature is also considered. For example, “Mary went to Paris yesterday” and “Mary went to London yesterday.” For this feature, after the NE extraction, the type and value of NEs are compared for possible inconsistency.

[12] http://dbrg.ut.ac.ir/Bijankhan/
[13] http://www.parsigan.ir/datasources/NER/8
[14] http://www.peykaregan.ir/dataset/پیکره-وابستگی-نحوی-زبان-فارسی
[15] http://sentistrength.wlv.ac.uk/

3- Different sizes of the two sentences; sometimes, the length of the sentences in the text and the hypothesis varies greatly. This feature can sometimes be decisive for identifying contradictions.

4- The adjective similarity of two sentences; when two sentences have the same syntactic structure but different adjectives, such as colours, it can be used to distinguish the contradiction. For example, “the woman has a black shirt” and “the woman is wearing a blue shirt.” For calculating this value, the similarity or antonymy of two corresponding adjectives is investigated by observing the ADJ part of speech tags in both sentences.

5- Verb similarity of two sentences; the verb is one of the most important elements of a sentence. Therefore, in many cases, the difference between verbs can clearly reflect the difference in the content of two contradictory sentences.

6- Negation; the negation in verbs with the same stem or negative adverbs can indicate the opposition in a pair of sentences. For example, “Ali went to school” and “Ali did not go to school.”

7- Common words; the number of common words in two sentences is usually an important factor in determining the opposition or similarity of sentences, particularly few other discriminative information exists.

8- Cosine similarity of two sentences; the cosine similarity by removing the stop-words and normalization by the ratio of the length of two sentences can be an appropriate criterion for determining the similarity or contradiction of a sentence pair. In this paper, BOW vectors are used. The similarity value in entailments or similar sentences is higher than in contradictory pairs.

9- SRL argument similarity; in many contradictory cases, e.g., structural contradictions, semantic arguments are used in different positions, so semantic tags can sometimes be a suitable attribute for identifying contradictions. For example, “water floats on oil” and “oil is floating on water.” In this system, “TMP” and “LOC” labels are investigated.

10- Antonym; obviously, the occurrence of conflicting words and phrases (antonyms) in similar positions in sentence pairs can indicate contradiction. Therefore, this feature is one of the important features used. For example, “my clothes are still wet” and “my clothes are dry and warm.”

In the Basic-R, feature scores are normalized between 0 and 1, and then the weighted sum of these features is calculated for each sentence pair and converted to a contradiction score using thresholds. These thresholds and weights are determined using a part of the dataset that is separated as a development set with a simulated annealing algorithm for optimization. Also, some machine learning models are trained with different classifiers using the exact feature set for comparison. The details are stated in the evaluation section.

3.2 Using data mining to extract rules for contradiction detection (DM-R)

As stated before, semantic contradiction has different types, and for rule-based identification of contradictions the rules for each type should be found and applied. The main challenge of rule-based systems is to write the rules manually. To solve this problem, our second method is introduced in this section.

This system has two main parts. In the first part, using the associate rule mining method, the set of rules for identifying the contradictions is automatically derived from the frequency of the rules in the development set. In the second part, these rules are selected and applied according to the predefined categories. So, there are different algorithms for different categories of contradiction in sentences.

In the independent test of this system, three modes may occur: first, the type of input test data is ideally specified; second, a rule-based classifier is used to determine the input category; and third, all the algorithms are applied to the input, and the result is obtained by voting. Further details are given below.

A. For extracting the rules, the following steps are taken:

1-Extraction of dependency relations, semantic role labels, sentiment analysis labels, and part of speech tags (POS) for both sentences.

2-Keeping dependency relations that meet one of the following conditions and deleting the rest:

a. Relations in which the common words of two sentences appear;
b. Relations in which one side of an antonymy relationship appears;
c. Relations that have a certain tag like a "num."

At this step, these relationships need to be more general for frequent rule mining, so POS labels of words are placed inside the tuples instead of putting the exact words. For example, "amode (antonym-1, N-SING-COM)" or "num (NUM, N-SING-COM)."

3-Calculating the similarity of each two semantic arguments and putting them in the form of (arg1, arg2, similarity interval). The cosine similarity of BOW vectors was used, and similarity intervals were defined by 0.3 steps (0-03, 0.3-0.6, 0.6-1).

4-Determining the sentiment analysis label for both sentences and putting them in the form of (sentiment1, sentiment2). For example, (Positive, Negative) or (Positive, Positive). For sentiment analysis, a simple lexicon-based system based on sentistrenght lexicons is implemented and used. The lexicon consists of almost 900 polar words. For calculating the sentiment score, the occurrence of polar words or their stems along with the negation word list is considered.

5-Checking whether verbs are positive or negative (based on POS tags) and placing them in tuples for each pair of verbs and quantifiers or negative conjunctions in the two sentences. For example, (Verb1-NEG, Verb2-POS) or (Verb1-NEG, Qunt1-NEG).

6- Applying the associative rules mining using the WEKA toolkit (now that the gold label and some selected tuples for each sentence pair exist).

The following flowchart indicates the procedure of frequent rules extraction.

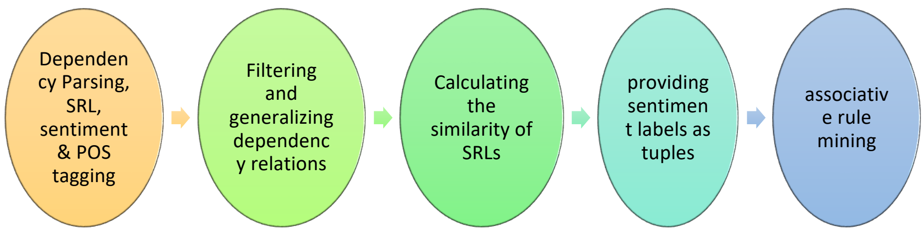

Figure 1: The procedure of frequent rules extraction

B. In the next step, the extracted frequent rules are manually assigned to different predefined contradiction categories. As discussed in the introduction, seven classes are

considered for contradiction types, including Antonym, Negation, Numerical, Structural, Factive, Lexical, and World Knowledge. Due to the complexity of the last three types, these three categories were merged under an “others” category. So, there are five predefined categories of numeric, negation, structural, anonym, and others. In the rule-based method, extracted rules are implemented for the four first categories, and the Basic-R is used for the “others” category. Table 4 shows some of the rules extracted for the predefined categories.
As stated before, to distinguish this method from the baseline rule-based method (Basic-R), it is called the DM-R.

3.1 BERT-based deep learning method using generated data

As stated at the beginning of Section 3, due to the emergence of BERT models [Devlin, 18] and their positive performance in similar tasks led the authors to fine-tune the Persian pre-trained BERT-based language models (ParsBert [Farahani, 21] and Persian Albert) on the translated data. As training data, part of SNLI, the whole MultiNLI datasets, and a nearly small part of manual translated data and FarsTail were machine translated. The results and implementation details of this deep learning subsystem are presented in Section 4.3.

For fine-tuning, the implementation[16] by Gao et al. (2021) was used, which is a PyTorch solution of the natural language inference (NLI) model based on Transformers.

Type	Rules	Example
Negation	(Verb1-NEG, Verb2-POS) Negation of verbs with the same stem (Verb1-NEG, Qunt1-NEG) (Verb1-NEG, Verb2-POS) Negative verb and quantifier in the first sentence and positive verb in the second	علی به مدرسه رفت. (Ali went to school.) علی به مدرسه نرفت. (Ali did not go to school) هیچ کس به مدرسه نرفت. (no one went to school) علی به مدرسه رفت. (Ali went to school)
	(Verb1-NEG, Qunt1-NEG) (Verb1-NEG, Verb2-POS) (Verb1-NEG, Quant 2-POS) Negative verb and quantifier in the first sentence and positive quantifier in the second	هیچ کس به مهمانی نیامد . (no one came to the party) همه به مهمانی آمدند. (everybody came to the party)
	(Verb1-NEG, ADV1-NEG) (Verb1-NEG, Verb2-POS) Negative verb and adverb in the first sentence and positive verb in the second one	هرگز به تهران برنمی گردم. (I never come back to Tehran) به تهران برگشتم. (I came back to Tehran)

[16] https://github.com/yg211/bert_nli

Numeric	Num (NUM1, N-SING-COM2) Num (NUM2, N-SING-COM2) [and basically comparison of the args]	سه دختر در خیابان نشسته بودند. (3 girls were sitting in the street) 5 دختر در خیابان بودند. (5 girls were in the street)
	Num (NUM1, antonym1) Num (NUM2, antonym2) number and antonyms in a sentence pair	سه دختر در مدرسه نشسته بودند. (3 girls were sitting at school) سه پسر در کلاس درس می خواندند. (3 boys were studying in the classroom)
Structural	(A0, A1, E2) Different agents and patients in SRL tags with the same words or stems	آب روی روغن شناور می ماند (Water floats on oil). روغن روی آب شناور است. (Oil is floating on water)

Table 4: Sample of extracted rules

4 Evaluation

4.1 Datasets

In this study, three datasets were used: (1) A thousand sentence pairs manually translated from the SNLI corpus of Stanford University, of which 324 are contradictory, and 676 are neutral and entailment; (2) A set of 250 sentence pairs (130 contradictory and 120 neutral sentence pairs), containing different categories of contradictions that were manually generated for testing the method; and (3) translated part of SNLI and the whole MultiNLI dataset machine-translated using Google Translate for training the deep learning subsystem.

The gold datasets statistics and sample distribution are presented in Table 5 (Dataset1 and Dataset2), and the details of the automatically generated dataset are indicated in Table 6.

Dataset	**type**	**Sentence pairs**	**Contradictory pairs**
Dataset1	Manual Translated SNLI	1000	325
Dataset2	Negation	42	26
	Numeric	47	28
	Antonym	52	26
	Structural	31	15
	WK+Factive+Lexical	78	35
	SUM	250	130

Table 5: Gold datasets statistics and sample distribution

Data source	**# Of sentence pairs**	**Contradictory pairs**
SNLI-test	10000	3333
SNLI-dev	10000	3333
MNLI(Dataset3)	400000	125000
All automatically translated data	420000	130000

Table 6: Description of the machine-translated data

As mentioned before, parts of the SNLI corpus and the whole MultiNLI dataset were machine-translated to be used in deep learning. After translation, only the MultiNLI dataset was chosen for deep learning experiments due to the following reasons:

- SNLI sentence pairs have similar and simple structures and also insufficient coverage for general topics. For example, many of these sentences are in the form of [a man/a woman/two boys/two girls/...] [verb] [a place such as on the beach/at home/…]. The official webpage of MultiNLI confirmed this claim thus "the corpus is modeled on the SNLI corpus, but differs in that covers a range of genres of spoken and written text, and supports a distinctive cross-genre generalization evaluation."

- By only using MultiNLI, the results can be compared to other systems with higher precision and fairness.

Therefore, in the experiments section, the machine-translated MultiNLI dataset is called Dataset3.

4.2 Metrics

Evaluation measurements are similar to most NLP tasks and consist of precision, recall, and F-measure, defined as follows:

$$\text{Precision(P)} = \frac{\text{correct system decisions}}{\text{all system decisions}} \tag{1}$$

$$\text{Recall(R)} = \frac{\text{correct system decisions}}{\text{what system should have decided}} \tag{2}$$

$$F - measure = \frac{2 \cdot P \cdot R}{(P + R)} \tag{3}$$

4.3 Results and Discussion

The results of the presented system evaluations are as follows. First, the evaluation of the Basic-R is reported in table 7. Also, as mentioned, some machine learning models is trained using extracted features in Section 3.1 to compare with the Basic-R. In the machine learning models, dataset1 is used for training (668 samples out of 1000 to balance the classes) the model, and different classifiers are tested in the Weka[17] toolkit.

[17] https://www.cs.waikato.ac.nz/ml/weka/s

Among the tested classifiers, the best results were obtained through naïve Bayes, RBF network, and Meta.Multi.class.Classifiers.

System name	**Train/Development set**	**Test set**	**Precision**	**Recall**	**F-measure**
Basic-R	Dataset1	Dataset2	62	68.2	65.1
Basic-R	Dataset2	Dataset1	53.7	86.4	66.2
ML-naïve Bayes	Dataset1	Dataset2	55.5	72	62.9
ML-RBF networks	Dataset1	Dataset2	56	65.5	60.4
ML-Meta.Multi class.Classifier	Dataset1	Dataset2	60.3	64	62.1

Table 7: Baseline evaluation of the two gold datasets

Considering the fact that the test samples in dataset2 are from different contradiction categories, it is clear that more complex categories such as world knowledge or factive samples cannot be identified well with rules. And the style of samples in dataset1 and dataset2 are different, so it is expected that the results of the Basic-R method would not be impressive. Also, due to the small volume of dataset1 and the fact that the dataset1 is part of SNLI, which is not multi genre, the average performance of non-deep machine learning approaches can also be justified.

Secondly, the evaluation of the DM-R is presented. As mentioned, test samples are provided for each contradiction category. In this section, an algorithm (rule matching) is applied to its associated samples. The results are shown in Table 8.

Also, the evaluations are performed for the dataset without considering indicated labels (such as negation and numeric), and all algorithms were applied to each sentence pair. The results which are indicated in the last row of Table 8 show that the rules are sufficiently discriminative for unlabelled data (having only contradiction or not contradiction tags).

As can be seen in Table 8, the DM-R for the first categories (Negation, Numeric, and Antonym) has an excellent performance. This is because, in the development dataset, there are proper examples of these types, so their rules are well extracted. On the other hand, sentence pairs belonging to these categories are usually not too complicated.

For the structural category, the performance is not promising due to the lack of a semantic role labelling tool with proper accuracy in the Persian language. For the "others" category, proper, specific rules could not be extracted, and only the samples were tested using the Basic-R. The results are not satisfactory due to the complexity of the sentences and lack of knowledge resources. Still, the overall performance of the proposed system is promising.

Category	Precision	Recall	F-measure
Negation	92.8	85.7	89.1
Numeric	89.3	82.14	85.6
Antonym	86.5	84	85.2
Structural	67	46	54.5
WK+Factive+Lexical	62	68.2	65.1
Average	79.52	73.21	75.9
Without considering indicated labels	73.2	72.5	72.9

Table 8: Evaluation of separate algorithms for each category (DM-R)

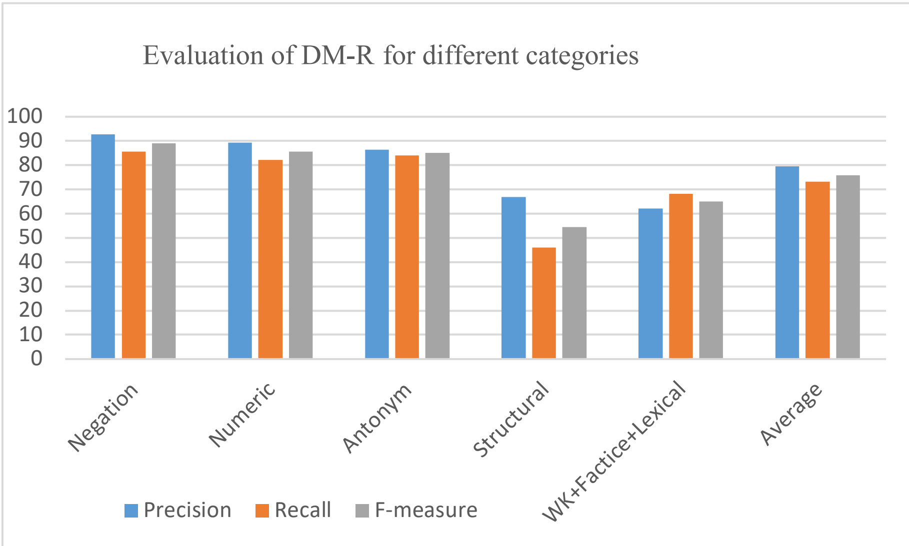

Figure 2: The evaluation of separate algorithms (DM-R) for each category

We can say that this is somehow because these categories are not well modelled by rules. To be more clear, it can be said that without involving knowledge or using accurate semantic resources and tools, it is very difficult to find accurate rules for detecting more complex contradiction categories. Therefore, as stated in Section 3, a BERT-based deep learning module was introduced to improve the overall performance.

For training and evaluating this module, various datasets were used. As stated in Section 3.3, the code from Gao et al. (2021) was used for fine-tuning the BERT models on datasets. At first, the English MultiNLI dataset was evaluated with different BERT models (BERT-based and ALBERT) to make sure of obtaining the same results as the authors. Secondly, the Persian deep NLI model was trained with the translated and shuffled MultiNLI dataset (Dataset3) and similar Persian BERT models (BERT-based[18] and ALBERT[19]). Also, these models were trained on the FarsTail and evaluated in two

[18] https://huggingface.co/HooshvareLab/bert-base-parsbert-uncased
[19] https://huggingface.co/HooshvareLab/albert-fa-zwnj-base-v2

ways: once each on its own test datasets and once on Dataset1 (for a fair comparison with rule-based methods). A description of the machine-translated data is shown in Table 6. As indicated in the related datasets in Section 2, FarsTail consists of almost 10000 sentence pairs (7000 for the training set and 1500 for each of the test and development sets). The evaluation results of the BERT-based deep learning method for the "contradiction" class are indicated in Tables 9 and 10.

Bert Model	Train dataset	Test dataset	P	R	F
Bert-base-v2	MultiNLI train	MultiNLI test	76	74	75
Bert-base-v3	MultiNLI train	MultiNLI test	73	70	72
Albert	MultiNLI train	MultiNLI test	77	70	73.8
Bert-base-v2	FarsTail train	FarsTail test	77	72	74
Bert-base-v3	FarsTail train	FarsTail test	66	70	68
Albert	FarsTail train	FarsTail test	67	64	66

Table 9: Evaluation of the BERT-based deep system on Dataset3 and FarsTail with their test datasets for the "contradiction" class

Train dataset	Test dataset	P	R	F
Dataset3	Dataset1	66	75	70.2
FarsTail train	Dataset1	51	62	56

Table 10: Evaluation of the BERT-based deep system on Dataset1 for the "contradiction" class

Table 9 shows the performance of different Persian Bert models finetuned with two datasets of translated MultiNLI and FarsTail. The results show that ParsBert version 2 is tuned better for both datasets in the task of contradiction detection. Apparently, the changes made in version3 was not better for this task. Also, Table 10 shows the performance of two Bert models trained on translated MultiNLI and FarsTail on the dataset1 (shared among all tests) and the translated set performed better. This issue is probably due to the fact that FarsTail dataset is more in the field of history, religion and literature and is a bit domain specific, but the MultiNLI is of different genres and is more similar to the style and domain of the dataset1 which is general.

It should be noted that according to the authors' studies (and the benchmark of Paperswithcode[20] website on the Natural Language Inference on MultiNLI so far), the best models for identifying textual implications for English on the MultiNLI dataset have a performance of approximately 91% (in F-measure), which is obtained through very large or new BERT models such as ALBERT-xxlarge or DoBerta that are not available in Persian. However, the best comparable model, which is a similar BERT-based model for Persian, has a performance of 87.7% in the F-measure for the contradiction class [Gao, 21]. Therefore, the Persian model based on BERT that uses translated MultiNLI is comparable to the English model.

Among similar Persian studies, Khodadadi and colleagues [Khodadadi, 15] have detected contradiction relations in one sentence. As explained in the previous paragraph, the deep learning algorithm of Gao [Gao, 21] code was tested, which reported the best results on the MultiNLI dataset with an implementable system for

[20] https://paperswithcode.com/sota/natural-language-inference-on-multinli

Persian with a Persian dataset. In Table 11, the performance of four algorithms of 1) Basic-R, 2) best machine learning model, 3) BERT-based deep learning method, and 4) DM-R are compared.

System name	Precision	Recall	F-measure
Basic-R	53.7	86.4	66.2
ML-naïve Bayes	55.5	72	62.9
DM-R	79.52	73.21	75.9
Persian Deep learning (ParsBert-V2)	76	74	75

Table 11: Best performance comparison of all implemented systems

It seems that the deep learning method have better recall on test data. The reason for this issue probably is that the deep network has discovered more complex features and patterns from data. By examining some test samples for which the corresponding rules was not matched truly in the DM-R, but is correctly identified by the deep learning method, the idea comes to mind that for those pairs of input sentences which do not precisely match the rules extracted for simpler contradiction categories (such as negation or numeric), the trained model of deep learning should be used to achieve better accuracy.

So as the results indicate, the best total performance for all types of contradiction is achieved through the BERT-based deep learning method. It is, however, lower than the DM system in three categories of Negation, Numerical, and Antonymy (F-measure=87). Therefore, the two methods of DM-R (to be used for the three mentioned categories) and BERT-based deep (for others) are combined. The final total performance is approximately 80 for all categories.

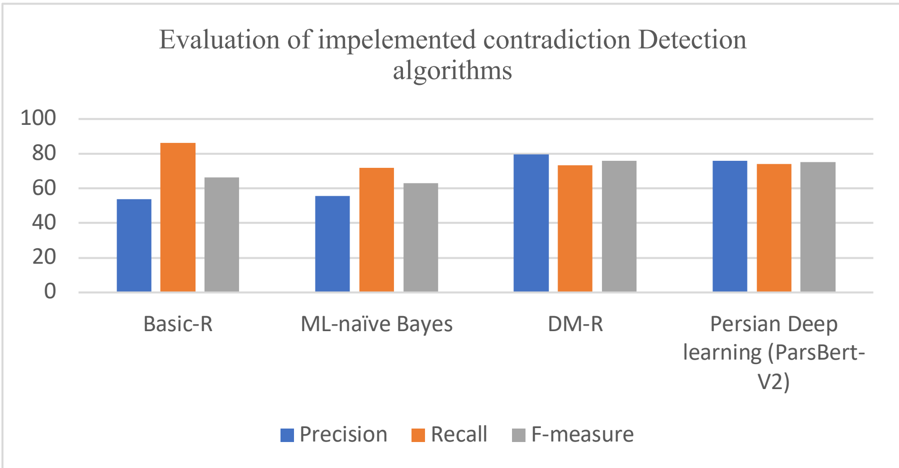

Figure 3: Evaluation of the implemented algorithms

On the whole, this paper presented a data mining-based contradiction detection method with excellent performance in specific categories (Negation, Antonym, and Numeric). Two points can be stated about this system: 1) In some contradiction types, it is more difficult to find rules or extra knowledge bases are necessary to cover the required semantic information, and 2) some sentences may be very long or complicated; so, the extracted rules may not cover such cases, and the DM system is not sufficient for contradiction detection in all sentences.

For a better contradiction detection system, a new dataset for the Persian natural language inference was introduced, which is the MultiNLI translated by Google. Using this dataset, a BERT-based inference model was trained for Persian texts to identify contradictions. This model has an overall performance of 75 for contradiction. So, a hybrid system was created to achieve higher performance. As the results indicate, this hybrid system has the best performance among Persian contradiction detection systems.

5 Conclusions

In this study, a rule-based baseline, a data mining system, and a BERT-based deep learning system were introduced to explore the semantic contradiction in Persian sentences. These systems were compared to various machine learning and deep learning methods on Persian texts and performed better. The baseline method is based on a series of general features to identify semantic contradiction, but the DM system, using a development set, automatically discovers the distinctive features of contradiction (especially for Persian texts) in the seven categories. In this regard, the frequent rule mining method was used, and the implementation of frequent rules assigned to each category led to promising results for some categories. As for a low-resource language such as Persian, machine learning and deep learning methods may not be effective due to a lack of appropriate datasets, developing rule-based methods with proper functioning can help identify semantic contradictions. The proposed data mining methods with extracting rules has an acceptable performance comparable to the best systems in the world. In the combinational category, due to the complexity of sentences and lack of appropriate knowledge resources, performance was not satisfactory. Therefore, a BERT-based deep system was presented to cover the shortcomings in other categories. The presented hybrid system has a performance of 73 (F-measure) for contradiction detection in Persian texts.

For future work, training a contradiction embedding model is planned for a possible better performance of detecting contradictory sentence pairs. Also, it is possible to examine other Bert models (such as mBert) for training. It is assumed that adding data or filtering less precise translated part of data would improve deep learning results.

References

[Amirkhani, 20] Amirkhani, H., et al., Farstail: A persian natural language inference dataset. arXiv preprint arXiv:2009.08820, 2020.

[Asmi, 11] Asmi, A. and T. Ishaya. Negation identification and calculation in sentiment analysis. In: The second international conference on advances in information mining and management. 2012.

[Bijankhan, 11] Bijankhan M, Sheykhzadegan J, Bahrani M, Ghayoomi M. Lessons from building a Persian written corpus: Peykare. Language resources and evaluation. 2011.

[Blanco, 11] Blanco, E. and D. Moldovan. Some issues on detecting negation from text. In: Twenty-Fourth International FLAIRS Conference. 2011.45(2):143-64.

[De Marneffe, 08] De Marneffe, M.-C., A.N. Rafferty, and C.D. Manning. Finding contradictions in text. In: Proceedings of ACL-08: HLT. 2008.

[Devlin, 18] Devlin, J., et al., Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[Farahani, 21] Farahani, M., et al., Parsbert: Transformer-based model for persian language understanding. Neural Processing Letters, 2021. 53(6): p. 3831-3847.

[Gao, 21] Gao, Y., N. Colombo, and W. Wang, Adapting by Pruning: A Case Study on BERT. arXiv preprint arXiv:2105.03343, 2021.

[Harabagiu, 06] Harabagiu, S., A. Hickl, and F. Lacatusu. Negation, contrast and contradiction in text processing. In: AAAI. 2006.

[Khandelwal, 19] Khandelwal, A. and S. Sawant, Negbert: A transfer learning approach for negation detection and scope resolution. arXiv preprint arXiv:1911.04211, 2019.

[Khodadadi, 15] Khodadadi, H., S. Rahati Quchani, and A. Estaji, Contrast Relation Recognition in Persian discourse using supervised learning methods. Signal and Data Processing, 2015. 12(2): p. 13-22.

[Li, 17] Li, L., B. Qin, and T. Liu, Contradiction detection with contradiction-specific word embedding. Algorithms, 2017. 10(2): p. 59.

[Mirzaei, 14] Mirzaei A, Moloodi AS. The first semantic role corpus in Persian language. Language Science. 2014 Sep 23;2(3):48-29.

[Noferesti, 16] Noferesti, S. and M. Shamsfard. Using data mining techniques for sentiment shifter identification. in Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). 2016.

[Pham, 13] Pham, M.Q.N., M.L. Nguyen, and A. Shimazu, Using shallow semantic parsing and relation extraction for finding contradiction in text. 2013.

[Rasooli, 13] Rasooli MS, Kouhestani M, Moloodi A. Development of a Persian syntactic dependency treebank. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2013 Jun (pp. 306-314).

[Rocktäschel, 15] Rocktäschel, T., et al., Reasoning about entailment with neural attention. arXiv preprint arXiv:1509.06664, 2015.

[Safavi, 00] Safavi, K., A brief introduction logical semantics. Literary Text Research, 2000. 4(9.10): p. 143-181.

[SankaraSubramanian, 09] SankaraSubramanian, B. and R.V.K. Mehta, Contradiction Analysis in Text Mining: A Fuzzy Logic Approach. SCSVMV University, kanchipuram, INDIA, 2009.

[Sepúlveda-Torres, 21] Sepúlveda-Torres, R., A. Bonet-Jover, and E. Saquete, "Here Are the Rules: Ignore All Rules": Automatic Contradiction Detection in Spanish. Applied Sciences, 2021. 11(7): p. 3060.[Lingam, 18] Lingam, V., et al., Deep learning for conflicting statements detection in text. 2018, PeerJ Preprints.

[Sifa, 19] Sifa, R., et al. Towards contradiction detection in german: a translation-driven approach. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI). 2019. IEEE.

[Shahshahani, 19] Shahshahani M S, Mohseni M, Shakery A, Faili H. PAYMA: A Tagged Corpus of Persian Named Entities. JSDP 2019; 16 (1) :91-110.

[Shamsfard, 19] Shamsfard, M. Challenges and Opportunities in Processing Low Resource Languages: A study on Persian. International Conference Language Technologies for All (LT4All), 2019.

[Shih, 12] Shih, C., et al., Validating contradiction in texts using online co-mention pattern checking. ACM Transactions on Asian Language Information Processing (TALIP), 2012. 11(4): p. 1-21.

[Tarunesh, 21] Tarunesh, I., S. Aditya, and M. Choudhury, Trusting roberta over bert: Insights from checklisting the natural language inference task. arXiv preprint arXiv:2107.07229, 2021.

[Tawfik, 18] Tawfik, N.S. and M.R. Spruit. Automated contradiction detection in biomedical literature. In: International Conference on Machine Learning and Data Mining in Pattern Recognition. 2018. Springer.

[Vargas, 17] Vargas, D.S. and V. Moreira, Identifying sentiment-based contradictions. Journal of Information and Data Management, 2017. 8(3): p. 242-242.

[Wang, 15] Wang, S. and J. Jiang, Learning natural language inference with LSTM. arXiv preprint arXiv:1512.08849, 2015.