Semantic Relation Extraction: A Review of Approaches, Datasets, and Evaluation Methods With Looking at the Methods and Datasets in the Persian Language

HAMID GHARAGOZLOU, JAVAD MOHAMMADZADEH, and AZAM BASTANFARD, SAEED SHIRY GHIDARY,

A large volume of unstructured data, especially text data, is generated and exchanged daily. Consequently, the importance of extracting patterns and discovering knowledge from textual data is significantly increasing. As the task of automatically recognizing the relations between two or more entities, semantic relation extraction has a prominent role in the exploitation of raw text. This article surveys different approaches and types of relation extraction in English and the most prominent proposed methods in Persian. We also introduce, analyze, and compare the most important datasets available for relation extraction in Persian and English. Furthermore, traditional and emerging evaluation metrics for supervised, semi-supervised, and unsupervised methods are described, along with pointers to commonly used performance evaluation datasets. Finally, we briefly describe challenges in extracting relationships in Persian and English and dataset creation challenges.

Additional Key Words and Phrases: Semantic relations, relation extraction, Natural Language Processing (NLP), automatic extraction, Persian text processing, linguistics, dataset, evaluation methods, information extraction

1 INTRODUCTION

Unstructured text sources account for most of the data produced and shared on the Internet [Aydar et al. 2020], making it difficult for humans to effectively manage, analyze, and extract relevant knowledge from this data [Abualigah and Altalhi 2022; Alhaj et al. 2022]. As the volume of text information on the Internet and in contemporary applications has expanded, so has an interest in text analysis to assist in processing vast amounts of unstructured text information [Abualigah et al. 2018; Mohammad and Abualigah 2018]. As a result, it is essential to create strategies for automatically extracting information from these documents, as they contain a wealth of critical data. The extracted data can be utilized to enhance access to and administration of knowledge buried inside massive text corpora [Abualigah et al. 2022; Pawar et al. 2017]. We leverage relational facts between the text's topics (entities) to make greater sense of the data and form more accurate conclusions. To gain comprehensive insights into specific domains like biology, banking, and social networks, utilizing computers is imperative for analyzing the relevant data [Aydar et al. 2020]. It is required to turn unstructured text into structured text. Annotating semantic information is a well-liked concept. However, the vast number and diversity of the data make human annotation impractical. Instead, we may have the machine organize the data by adding tags in our chosen format. Relationships between entities such as persons, organizations, and locations are often of interest [Bach and Badaskar 2011].

Named entities, relations, and events are the three types of data users often want to extract from documents [Pawar et al. 2017]. A named entity typically refers to a specific person, place, organization, or other things that a name can identify. For example, the sentence "In January 2015, Barack Obama flew to India" contains a named entity, which is a well-known individual in this case. Entity recognition is a method for finding each occurrence of a certain kind of anonymous entity in documents. A relation is often used to describe a strong relationship between two or more well-known entities. Consider the "having" relationship between a "product" and "feature" or "author" relationship between a "person" and a "book title" [Pawar et al. 2017].

Relation extraction is the process of anticipating an entity's properties and relations within a text. The sentence "Barack Obama was born in Honolulu, Hawaii" is used as an example by a relation classifier to try to predict the relationship "born in the city." Relation extraction is a vital component of **Natural Language Processing (NLP)**, used for downstream applications such as question answering systems, structured and semantic search, text summarization, and sentiment analysis. It is instrumental in creating relational knowledge graphs [Huang and Wang 2017].

Since the expression of semantic relations is language dependent, the relationship extraction process is also language dependent. Since most research is conducted in English, porting these approaches into other languages may be challenging. Although this article primarily focuses on English, we will also devote significant attention to exploring various aspects of relation extraction in Persian.

This article presents a comprehensive review of relation extraction in English and an overview of relation extraction methods, datasets, and unique challenges in Persian. In Section 3, we classify different types of relation extraction and their associated challenges. Section 4 classifies different approaches to relation extraction and introduces their advantages and limitations. Section 5 reviews prominent relation extraction datasets, including their features and statistical characteristics. In Section 6, we discuss evaluation methods for relation extraction. Continuing the survey, we examine relation extraction tasks in Persian and related topics. In Section 7, we review previous work on relation extraction from Persian texts, and in Section 8, we analyze existing Persian relation extraction datasets. In Section 9, we introduce the obstacles encountered in creating these datasets. Finally, Section 10 summarizes our study and offers suggestions for future work.

2 REVIEW METHODOLOGY

This article reviews proposed approaches, datasets, and evaluation methods in semantic relation extraction. In addition, we investigate the methods and datasets for Persian relation extraction. This section introduces the review methodology, including the research questions, sources of information, search criteria, and key words.

Table 1. Questions Raised in the Research and Mapping Sections

Q#	Research Question	Mapping Section
RQ1.	What types of relationship extraction exist, and what approaches are used for each? What are the major challenges in each category?	Section 3
RQ2.	What are the most important datasets for extracting relationships? What are the characteristics and statistics of each of these datasets?	Section 4
RQ3.	What are the most important datasets for relation extraction? What are the characteristics and statistics of each of these datasets?	Section 5
RQ4.	How to evaluate different types of relationship extraction systems? What are the essential evaluation criteria?	Section 6
RQ5.	What are the different suggested approaches for relation extraction in Persian, and how do they work?	Section 7
RQ6.	What are the specific datasets for extracting information and relationships in the Persian language, and how are they made?	Section 8
RQ7.	What are the challenges in the relationship extraction task and dataset creation?	Section 9

2.1 Research Questions

To better understand relation extraction and its related concepts in English and Persian, we planned to carry out an exhaustive overview. Planning the review process requires research questions. Table 1 depicts the research questions and their mapping sections in this article.

2.2 Search Criteria and Key Words

To obtain a comprehensive and reliable set of sources, we searched various combinations of the key words mentioned in the following in widely used online scientific databases, including Springer, Wiley, ACM, IEEE Xplore, ScienceDirect, Semantic Scholar, and Google Scholar. The articles we used were extracted from conferences, journals, magazines, and transactions. Although most of the articles were journal papers, we tried to select articles published in more prestigious journals as our final resources.

Since semantic relation extraction is a field with a relatively long history, we initially set the time range of our search to "last 20 years." We then reviewed the resulting sources in an exploratory way. After gaining initial insights, we narrowed our search to articles published from the beginning of 2008 to the present to filter more recent research. The findings confirm that the speed of changes in relationship extraction methods and approaches has been much faster than the development of datasets. Finally, we focused on the methods proposed in recent years, specifically articles published since 2018.

We collected the most relevant articles to answer our research questions using the search term "relation extraction" combined with key words like "Persian," "semantic," "approaches," "dataset," "evaluating," and "evaluation metrics." Additionally, we searched for phrases such as "textual dataset creation methods," "textual dataset creation challenges," "Persian linguistic features," and "Persian text processing challenges."

2.3 Source of Information

We utilized the most influential online scientific databases, including IEEE Xplore, ACM Digital Library, Springer, Wiley Online Library, ScienceDirect, and Google Scholar. To identify and categorize datasets, we initially used Google Dataset Search to prepare an initial list. We also utilized other databases, such as the **Linguistic Data Consortium (LDC)**, available through UPenn (https://www.ldc.upenn.edu), and LINDAT, available through Charles University (https://www.lindat.cz). Additionally, we utilized Papers with Code (https://paperswithcode.com) as a valuable database to identify the specifications of datasets and explore the challenges, leaderboards, and benchmarks related to relation extraction.

3 CLASSIFICATION OF RELATION EXTRACTION TYPES

Based on the procedure, relation extraction may be broken down into several categories. Traditional relation extraction tasks primarily focus on the binary relationship between two entities, aiming to detect the relationship between them accurately in text. The limitations of these conventional activities hinder the thorough or accurate extraction of relationships. **Complex Relation Extraction (CoRE)** tasks have been devised to address these issues. This section introduces different relation extraction types.

3.1 Global Relation Extraction vs Mention-Level Relation Extraction

Relation extraction may be categorized into two categories in general: global relation extraction and mention-level relation extraction. A global system often produces a list as its output from a significant amount of text as an input. This system should generate a list of entity pairings with a particular semantic relationship. However, in determining the mention level, the algorithm evaluates the sentence containing two entities and identifies the presence or absence of a specific relationship between them Pawar et al. [2017].

3.2 Simple Relation Extraction vs Complex-Level Relation Extraction

Relation extraction may be broadly categorized into two types: simple and complex. Contrary to simple relation extraction, CoRE aims to extract more intricate relationships based on certain limitations and requirements. We will go through instances of CoRE, such as n-ary relation extraction, conditional relation extraction, few-shot relation extraction, **Continual Relation Extraction (CRE)**, multi-dimensional relation extraction, **Nested Relation Extraction (NRE)**, and overlapping relation extraction.

3.3 Sentence-Level Relation Extraction vs Document-Level Relation Extraction

Sentence-based annotated training data is employed for sentence-level relation extraction, where sentences in the training set are labeled with triples using sentence-triple alignment annotation. The trained model then predicts new relations for fresh entity pairings. However, a significant challenge in real-world scenarios is the need for labeled data [Aydar et al. 2020].

Sentence-level approaches limit the comprehension of entity-pair relations across a text. As a result, it overlooks relations that can only be understood by carefully understanding many phrases in a text [Aydar et al. 2020]. Real-world relation extraction often involves processing large amounts of document-based data, which requires a document-level extraction model. The specified relation, in this instance, may span numerous lines or pages. The issue is more overwhelming because of this characteristic than it would be to extract an intra-sentence relationship [Jiang et al. 2020].

3.4 Binary Relation Extraction vs n-ary Relation Extraction

The binary relationship between two entities is primarily the focus of classic relation extraction activities. **Binary relation extraction (BiRE)** is a learning-based technique that falls into three main categories: supervised, semi-supervised, and unsupervised. Simple BiRE has improved significantly and provides several effective and practical solutions. The straightforward BiRE technique, however, cannot satisfy the demands of the rapidly expanding field of intelligent applications. Researchers have addressed certain BiRE limitations by utilizing the CoRE technique. In contrast, n-ary relation extraction aims to extract relationships between multiple entities from one or more phrases. Due to its potential applications in identifying cause-and-effect relationships and predicting drug-gene-mutation occurrences, NRE has gained increased research interest [Jiang et al. 2020].

3.5 Conditional Relation Extraction

Conditional relations are relationships formed under one or more constraints, like temporal or geographical ones. For instance, we know that the ternary relationship between "Obama," "President," and "United States" only holds true from 2008 to 2017. Using this information to answer knowledge-based questions might result in significant errors. A conditional relation is often expressed as (f, and, m, sh), where "sh" is the requirement that maintains the relation's validity and "f, and, m" is the same as the primary subject-attribute-object trinity. Therefore, President Barack Hussein Obama's situation is a temporary separation (2008–2017) [Jiang et al. 2020].

Today, massive knowledge bases like DBpedia, Freebase, and YAGO contain millions of entity instances and relationships. Only some of them, however, take into account conditional relations or external factors; this substantially restricts the utility of current knowledge bases in sophisticated reasoning tasks and necessitates urgent conditional relation extraction research [Jiang et al. 2020; Liu et al. 2019].

3.6 Few-Shot Relation Extraction

Most of the time, a relation only has a small number of instances, causing typical relation extraction models to be useless. Few-shot learning and few-shot relation extraction are the novel paradigms that have shown promise for solving this issue [Jiang et al. 2020].

In real-world applications, labeled instances are often scarce due to linguistic diversity, distinct domains, and the cost of human annotation, which can challenge the development of effective models. However, improving the performance of such models is possible. Alternatives include techniques that need a small number of instances (few-shot) or none at all (zero-shot) [Sainz et al. 2021].

3.7 Continual Relation Extraction

As new information becomes available, it must be extracted from unstructured text, making CRE crucial. CRE is a more practical and beneficial setting since the updating process is continuous and iterative. However, providing the relation extractor access to all training cases in previously performed tasks would be unfeasible due to storage and computing resource limits. The standard relation extraction setup, where the extractor is typically taught from the beginning with full access to the training corpus, contrasts with our constant learning approach [Wu et al. 2021].

Traditional relation extraction algorithms can only effectively manage the expanding kinds of relations in real life if they typically presuppose a predefined set of standardized relations and train on a fixed dataset. The CRE method addresses this problem. CRE seeks to aid the model in learning new relations while preserving accurate categorization of existing ones compared to classical relation extraction [Zhao et al. 2022].

3.8 Multi-Dimensional Relation Extraction

Videos and images have also developed into valuable resources due to the rapid expansion of online information. Multi-dimensional connection extraction is used to extract relationships from this massive database. As a live form of information transmission, images and videos may convey a wealth of knowledge. On the one hand, people prefer to use visuals to convey some of their knowledge of broad perception rather than express it directly. On the other hand, combining a

Table 2. The Most Significant Challenges in Different Relation Extraction Types

Relation Extraction Type	Most Important Limitations and Challenges
Sentence level	• There is not adequate training data. • Poor cross-sentence reasoning performance.
Document level	• Various document formats, long dependencies.
N-ary	• Absence of end-to-end models: most preparatory work calls for all references to entities, but getting them might be time consuming. An end-to-end model may be quite beneficial in practical applications.
Conditional	• Dependency complexity. • Conditions may be expressed in a variety of ways in free writing. The conditional dimension needs to be formalized using a broad framework. • Data insufficiency: not enough specific data is currently available for conditional relation extraction.
Few-shot	• Severe performance degradation, especially when using minimal training settings.
Continual	• Namely catastrophic forgetting: when a neural network is used to learn a series of tasks, the performance of the learned model for the earlier tasks may be negatively impacted by the learning of the later tasks. • Order-sensitivity: this term describes the situation where task performance changes depending on the order in which they are presented to the user.
Multi-dimensional	• General knowledge: the link drawn from already completed activities has a solid relationship for data input, such as "the ball in the image is red." It will be challenging to apply this little knowledge to additional research. One of the main goals of extracting a multi-dimensional relationship is to figure out how to extract a conceptual and informative relationship, such as commonsense knowledge. • Even though some research has started to create multi-dimensional knowledge databases, more research is needed before they compete with current knowledge bases.
Nested	• The complexity of structure: sentences include numerous nested entities and relationships or a high number of clauses. Direct analysis of nested structures is challenging due to the complex sentence structure. • In a sentence, the subject may only be mentioned once, and it usually takes the form of a reference, such as when pronouns are used. Therefore, it is necessary to locate its actual existence.
Overlapping	• The complexity of relations: a sentence may have none or several relationships between two entities. • Unknown entities and relationships: it is challenging to locate entities and relationships accurately since their locations are unknown.

complex set has often produced positive outcomes. These events show how crucial it is to capture relationships in pictures or videos instead of merely using plain words [Jiang et al. 2020].

3.9 Nested Relation Extraction

The notation for the more common BiRE is (arg1, relation, arg2), whereas the notation for the more complex NRE is either ((arg1, relation1, arg2), relation2, arg3), or (arg1, relation1, (arg2, relation2, arg3)). Although standard BiRE loses some details, which causes our triads to become uninformed and incomplete, NRE aids in more properly expressing the core sentence's meaning. NRE is beneficial for downstream tasks, such as question answering, that rely heavily on the precision and comprehensiveness of triads. Current studies have emphasized NRE. For instance, NESTIE learns syntactic patterns for relationships written in nested forms, whereas StuffIE leverages the Stanford lexical and dependency parsing database to extract nested relations [Jiang et al. 2020].

3.10 Overlapping Relation Extraction

Almost every method for finding relationships starts with the idea that each sentence has a single relational truth. Even so, relational facts in sentences are often complicated, and many relational triplets may overlap in a statement. As a result, the extraction of overlapping relations has garnered much interest. The primary purpose of this task is to extract all potential relationship information from a phrase while considering the intricate overlap between triplets. Sequence-to-sequence approaches and graph-based methods are two existing techniques for uncovering overlapping relations. Sequence-to-sequenced approaches use unorganized text as an input and produce relational triples as a sequential output after decoding them instantly. Graph-based approaches build a graph neural network for the combined extraction of entities and overlapping relations [Hang et al. 2021].

In Table 2, the most noteworthy challenges of different relation extraction types are summarized.

4 CLASSIFING RELATION EXTRACTION APPROACHES

The presented approaches for relation extraction can be categorized from different perspectives. In this section, we classify these approaches in terms of learning methods after a general classification. Next, because most of the techniques presented in recent years are based on machine learning and deep neural networks, we describe some of the most influential works in this category. The advantages and limitations of each category are also discussed.

4.1 A General Classification

Relation extraction methods can be broadly categorized into four groups: pattern based, structure based, language based, and statistical. Pattern-based approaches extract taxonomic and non-taxonomic relationships between entities by matching textual patterns. Structure-based approaches have been proposed for semi-structured texts like Wikipedia articles. These methods attempt to identify and utilize pre-existing text structures in the relation extraction process. In statistical models, the relation extraction process is based on the distribution of textual semantic relations. In language-based extraction methods, linguistic activities such as morphological, syntactic, and semantic analysis are utilized [Fadaei and Shamsfard 2010]. The advantages and limitations of these approaches are summarized in Table 3.

Table 3. Advantages and Limitations of Pattern-Based, Structure-Based, Language-Based, and Statistical Methods

Approach	Advantages	Limitations/Disadvantages
Pattern-based techniques	• No need for strong processing power • Ability to extract taxonomic and non-taxonomic relations	• Inability to recognize all patterns • Limited functionality • The need for human resources to define the rules and patterns • Ignoring omnidirectional relationships between pairs
Structure-based techniques	• Ability to extract taxonomic and non-taxonomic relations	• Unsuitable for unstructured text
Statistical methods	• Suitable for choosing the most appropriate inter-constituent relationships	• Common challenges in clustering problems (because these methods usually rely on clustering techniques)
Language-based techniques	• Considering syntactic and semantic features simultaneously	• Availability of powerful tools like morphological analyzers, chunkers, and parsers

4.2 Classification Based on Learning Approach

From the point of view of a learning approach, the three types of methods used in semantic relation extraction are supervised, semi-supervised, and unsupervised. In addition, **Reinforcement Learning (RL)** and distant-supervised learning have recently attracted the attention of researchers as two powerful approaches for relation extraction. In this section, we explain these five approaches.

In the supervised approach, the relationships between the entities of a text are manually labeled, and then a classifier is designed and trained. The training process is based on the features of the sentences in which those entities appeared. Supervised relation extraction and its reliance on labeled data has several advantages, including higher reliability than other approaches. However, this approach also has several limitations, such as the need for a large volume of labeled data, its expensive and time-consuming nature, its strong dependence on expert staff, the high computing time required in the training phase, and decreased efficiency due to the need for preprocessing tasks. Despite its limitations, supervised learning remains a popular and effective approach, providing accurate results and allowing for a straightforward interpretation of the data [Nasser et al. 2019].

In semi-supervised methods, a small number of manually labeled samples are used. Next, these samples (known as seeds) are searched within the text. Then, patterns are extracted from the sentences that contain these seeds. These patterns are used to extract additional samples. Similarly, semi-supervised methods have flaws. Due to the specificity of the patterns, there is a low recall. Because a pattern does not necessarily represent a specific relation, relation instances may be incorrectly identified, resulting in additional false instances [Nasser et al. 2019; Pawar et al. 2017]. Semi-supervised relation extraction does not require extensive human intervention. Nonetheless, this method has limitations, such as its inability to extract cross-sentence and document-level relationships. Although semi-supervised learning can be helpful in situations with limited labeled data, its limitations must be considered to ensure accurate and exhaustive results.

Unsupervised methods do not require labeled data and therefore do not have a low recall. The disadvantage of these approaches is that they require populating a knowledge base using the extracted relationships. To do this, we must link the extracted relations to the knowledge base relations [Nasser et al. 2019]. Unsupervised learning has the advantage of adequate performance in extracting inter-constituent relationships without requiring extensive human intervention to label relationships. However, it also has limitations and disadvantages. One major limitation is the lack of prior knowledge about relations, which can result in lower accuracy of models. Additionally, the models' low accuracy is due to the lack of human intervention.

RL has helped remove noisy data and improve features in classification problems for text processing. This learning approach significantly improves classification performance. When using RL to extract relations, the extractor module is considered an RL agent [Zeng et al. 2018]. The classification task is a sequential decision-making process in which the agent receives and classifies a sample relation at each stage. After that, the environment gives immediate and subsequent rewards to the agent. When the environment correctly classifies the sample, a positive reward is assigned to the agent; otherwise, it receives a negative reward. Finally, the agent learns optimal behavior by maximizing the total rewards and then can classify the samples as accurately as possible. Finally, the agent finds an optimal value for the policy weights [Gharagozlou et al. 2022].

The advantages and limitations of these five learning approaches are demonstrated in Table 4.

Table 4. Advantages and Limitations of Supervised, Semi-supervised, Unsupervised, Reinforcement Learning, and Distant-supervised Approaches

Approach	Advantages	Limitations/Disadvantages
Supervised relation extraction	• *Higher* reliability compared to other approaches (the decision-making basis for these methods is provided by humans). • No relation restriction.	• The need for a large volume of labeled data • Expensive and time consuming • Strong dependence on expert staff • High computing time in the training phase • Decreased efficiency due to the need for preprocessing tasks
Semi-supervised relation extraction	• No need for extensive human intervention.	• Inability to extract cross-sentence relations • Inability to extract document-level relations
Unsupervised relation extraction	• Adequate performance in extracting the inter-constituent relationship. • No need for extensive human intervention to label relationships.	• Lack of prior knowledge about relations • Low accuracy of models due to lack of human intervention
Distant-supervised relation extraction	• No need for human experts. • Low consumption, affordable.	• Dealing with large amounts of noisy data • Low generalizability
Reinforcement learning	• RL-based models are quite similar to human learning, hence it is possible to achieve perfection by using it. • RL-based models are compelling in classifying unbalanced relationships.	• Need for a lot of data and computation • The possibility of a large number of states, a phenomenon that can affect the efficiency of the model

4.3 Machine Learning and Deep Learning Methods

Over the past decade, machine learning based methods have been widely used to solve text processing problems. A wide range of machine learning based methods have been used for the relation extraction task, from support vector machine and similar techniques to deep neural networks and pretrained language models. In recent years, deep learning and modern neural networks have caused fundamental developments in many fields. The widespread use of deep neural networks in text processing tasks, including semantic relation extraction, has led to the construction of improved and more advanced models. The advantages and limitations of different types of machine learning in relation extraction are exhibited in Table 5.

Some of the most important relation extraction methods based on machine learning and deep learning are introduced in the following.

Socher et al. [2012] proposed a deep **Recurrent Neural Network (RNN)** architecture that defines a combined vector representation of words and phrases in a parse tree format. A vector and a matrix represent each expression. The vector encodes the semantic information of an expression, and the matrix encodes the degree of its influence on the meaning of syntactically neighboring expressions. Xu et al. [2015] introduced a model based on **Long Short-Term Memory (LSTM)** that utilizes SDP (Shortest Dependency Path). They use a directed graph as a dependency tree to model dependencies between entities. Zhang et al. [2015] employ **Bidirectional Long Short-Term Memory (BiLSTM)** to improve sentence representation. As an extension, Zhou et al. [2016] added an attention mechanism to the BiLSTM architecture to obtain more meaningful parts of sentences.

Adel and Strötgen [2021] provide an enriched attention mechanism to strengthen neural relation extraction. The model has four essential modules: two stacked LSTMs, an input layer, an

Table 5. Most Important Challenges in Different Relation Extraction Types

Method	Advantages	Limitations/Disadvantages
Traditional methods (like information retrieval and statistical methods)	• A multitude of methods and a long history • Efficient extraction of syntactic features	• Ignoring semantic features and relations • Low accuracy in extracting correct relations • Low generalizability
Non-neural machine learning models	• Higher accuracy than traditional non-machine learning methods	• Difficulty in defining and extracting features • Quality and accuracy of the model depend a lot on the quality of the features extracted
Machine learning models: CNNs	• The ability to properly represent the hierarchical structures of sentences • Efficient evaluation of sentence matching	• Attention to one-sided features • Ignoring complex semantic features in sentences
Machine learning models: LSTMs	• The ability to maintain sequential dependency information • Ability to efficiently generate embedded sentence	• Creating an embedded representation of sentences using only a neural network scheme
Language models (based on BERT and transformers)	• The ability to understand and identify syntactic rules and semantic relations simultaneously • Linguistic prediction capability, including word and next sentence prediction and masked word prediction • Ability to generate new texts with correct syntax and semantic rules (this feature is quite useful for automatic generation tasks)	• The need for high processing power to retrain language models such as BERT • A very high number of model parameters • Different treatment (different accuracy) with positive and negative samples

enriched attention mechanism, and the output layer. Yang et al. [2018] propose an ensemble neural architecture using adaptive boosting LSTMs powered by an attention mechanism for relation extraction. LSTM's role is to embed sentences; like other systems, the attention mechanism plays its classical role. The final neural classifier is made using adaptive boosting. This approach leads to a robust joint ensemble network that can extract relationships accurately. Nguyen and Verspoor [2019] proposed an end-to-end neural network employing a BiLSTM-CRF and a biaffine attention for **Named Entity Recognition (NER)** and **Relation Classification (RC)** tasks, respectively. Nayak and Ng [2019] developed a reusable multi-factor attention model consisting of a dependency parser to extract syntactic features, a linear attention mechanism to calculate semantic similarity, and a formulation to measure the dependency distance of words.

Zeng et al. [2014] introduced a **Convolutional Neural Network (CNN)** based model that combines several local features to obtain a global feature representative of all local features.

Mintz et al. [2009] presented a new perspective entitled "Relation extraction without labeled data." It was suggested to use the distant supervision approach in relation extraction. Han et al. [2018a] proposed an instance-level adversarial training model to enhance the distant supervised relation extraction by reducing noisy samples. Liu et al. [2018] proposed another distant-supervised method for word-level relation extraction by building an STP (Sub-Tree Parse) to eliminate noisy instances. Accordingly, a neural model is designed to take the generated sub-tree as input and perform an entity-wise attention operation to filter more important features. The neural network is initialized using the prior insights obtained from the transfer learning approach. In another study, Nayak et al. [2021] used a self-ensemble filtering technique to omit noisy data from the train set. RH-Net [Wang 2020] is a distantly supervised relation extractor that benefits from the advantages of **Hierarchical Relational Search (HRS)** and RL and incorporates these techniques to select high-quality relation instances. The primary role of the HRS module is to identify semantic relationships.

Yuan et al. [2019] presented a distant-supervised method empowered by an attention mechanism to select the most important words in sentences. The system consists of a linear attention simulation mechanism to detect the importance of a word in the text and a novel embedding approach to determine the relevance of the words in bags.

Some researchers have targeted deep few-shot learning for relation extraction. Sainz et al. [2021] redefines the relation extraction process as an entailment task with straightforward verbalizations of relations. A textual entailment engine has been developed that uses manually generated verbalizations. Baldini Soares et al. [2019] presented one of the most potent relation extraction models based on few-shot learning. They claim to have surpassed human accuracy. Ye and Ling [2019] propose a multi-level matching algorithm powered by an aggregation network architecture.

Recently, the advantages of the transfer learning paradigm have been utilized in deep learning models. Pretrained language models like BERT and GPT are the result of these developments. As two examples of BERT-based models, Zhao et al. [2019] and Wei et al. [2019] achieved the best results on SemEval-2010 Task 8 and the NYT datasets, respectively. In the work of Peng et al. [2020], an entity-masked contrastive pretraining model for relation extraction has been proposed. The proposed framework is a context-aware system that combines the BERT language model, CNNs, and an MTB (Matching the Blanks) approach. BERT-Side [Moreira et al. 2020] introduced a distantly supervised neural relation extraction system that employs side information using the BERT model. In BERT-Side, the sentence embedding is obtained from the BERT language model, and a built-in attention mechanism determines the importance of the words.

He et al. [2020] present an approach based on PLU (Positive and Unlabeled Learning) to boost the efficiency of distantly supervised relation extraction. The model utilizes RL to determine the positiveness of a sentence against a particular relation and then generates unlabeled, positive bags. Two different representations are defined: one for positive bags and another for unlabeled bags. Finally, the proposed model merges these bags to predict relations at the bag level. Table 6 depicts the technical characteristics of the explained methods.

5 RELATION EXTRACTION DATASETS

This section introduces the most famous relation extraction datasets, followed by a comparison and analysis of the datasets.

5.1 TACRED [Zhang et al. 2017]

With 106,264 samples created over English newswire and online text from the corpus utilized in the annual **TAC Knowledge Base Population (TAC KBP)** challenges between 2009 and 2014, TACRED is a large-scale relation extraction dataset. The Stanford NLP Group created it. Samples in TACRED are tagged as "no-relation" if there is no specified relation maintained, and they encompass 41 relation patterns utilized in the TAC KBP challenges (e.g., per: schools attended and org: members). These samples were generated by mixing crowdsourcing with the human annotations accessible through the TAC KBP tasks. The LDC's human annotations, the TAC KBP relation categories, and Mechanical Turk crowdsourcing were used to create the annotations. Each year of the slot-filling task evaluation, 100 entities (individuals/organizations) were provided as queries (i.e., subjects), for which the competing systems were expected to locate relevant relations and object

Table 6. Technical Characteristics of the Explained Methods of Machine Learning and Deep Learning Methods

Author(s)	Category	Technical Description	Dataset	F1-Score
[Zeng et al. 2014]	CNN based	Convolutional deep neural network	SemEval 2010	82.7
[Nayak and Ng 2019]	CNN based	CNN-based global feature extraction + multi-factor attention mechanism	NYT10	56.6
[Santos et al. 2015]	CNN based	CNN-based relation classifier	SemEval 2010	84.1
[Zhou et al. 2016]	RNN based	Attention-based BiLSTM networks	SemEval 2010	84.0
[Yang et al. 2018]	RNN based	Adaptive boosting LSTM + attention mechanism	Freebase + NYT	54.0
[Nguyen and Verspoor 2019]	RNN based	BiLSTM-CRF-based entity recognition + deep biaffine attention layer for RC	CoNLL04	69.6
[Adel and Strötgen 2021]	RNN based	Two stacked LSTMs + an enriched attention mechanism	ACE 2005	75.7
[Adel and Strötgen 2021]	RNN based	Two stacked LSTMs + an enriched attention mechanism	TACRED	68.3
[Wei et al. 2019]	BERT based	A novel hierarchical binary tagging framework for joint extraction	SemEval 2010	87.5
[Peng et al. 2020]	BERT based	An entity-masked contrastive pretraining model	TACRED	69.5
[Huang et al. 2022]	Language model-based	Joint semantic embedding	TACRED	75.5
[Cohen et al. 2020]	Language model based	Span prediction based method	TACRED	74.8
[Cohen et al. 2020]	Language model based	Span prediction based method	SemEval 2008	91.9
[Park and Kim 2021]	Language model based	Curriculum learning + graph attention network	TACRED	75.0
[Park and Kim 2021]	Language model based	Curriculum learning + graph attention network	Re-TACRED	91.4
[Wu and He 2019]	BERT based	Enriched pretrained language model with entity information	SemEval 2010	89.25
[Zhao et al. 2019]	BERT based	Entity pair graph for RC	SemEval 2010	90.2
[Baek and Choi 2022]	BERT based	Graph neural network (GCN) + SpanBERT	TACRED	75.4
[Wang 2020]	Distantly supervised	HRS + RL	NYT	84.6
[Nayak et al. 2021]	Distantly supervised	Self-ensemble noise filtering	NYT	54.0
[Han et al. 2018b]	Distantly supervised	Hierarchical attention + knowledge graphs	NYT	81.6
[Baldini Soares et al. 2019]	Deep few-shot	Best configuration: 5 ways, 1 shot	Fewer	88.9
[Ye and Ling 2019]	Deep few-shot	Best configuration: 5 ways, 5 shots	FewRel	92.66
[Sainz et al. 2021]	Deep few-shot	A textual entailment engine that uses manually generated verbalizations	TACRED	69.0
[Zhang et al. 2018]	Graph neural networks	Graph convolution + pruned dependency trees	TACRED	67.1
[Christopoulou et al. 2019]	Graph neural networks	Edge-oriented graphs + iterative algorithms	CDR, GDA	63.6
[Zhang et al. 2018]	Graph neural networks	GCN (graph coevolution network) + dependency trees	TACRED	68.2
[Hoffmann et al. 2011]	Multi-instance learning	Weak-supervised learning + a probabilistic multi-instance learning	Freebase + NYT	60.5

entities. Every one of the phrases evaluated in the TAC KBP assessment and a sample of additional sentences from the assessment corpus known as TACRED includes the question entities.

5.2 DocRED [Yao et al. 2019]

A relation extraction dataset called *DocRED* (Document-Level Relation Extraction Dataset) was created using Wikidata and Wikipedia. The dataset's documents are each annotated by humans with references to identified entities, coreference data, intra- and inter-sentence relations, and supporting documentation. The dataset also offers vastly dispersed remotely supervised data and human-annotated data. DocRED, the most significant human-annotated dataset for document-level relation extraction from plain text, annotates identified entities and connections. This dataset includes 5,053 Wikipedia articles, 132,375 entities, and 56,354 relational facts. The dataset contains large-scale, remotely supervised data from 101,873 articles and human-annotated data.

5.3 SemEval-2010 Task 8 [Hendrickx et al. 2019]

SemEval-2010 Task 8 is a multi-way classification dataset for identifying and labeling semantic relations between pairs of incompatible nominals. Employing NLTK and text BLOB packages, the dataset is standardized using standard NLP methods. The data is disseminated as a NumPy array and is represented using the word-vector model. The learning algorithm, test set, and independent verification set are all included in the frozen NumPy file. SemEval-2010 Task 8's objective was to provide a testbed for automated semantic RC.

5.4 Wiki-KBP [Ellis et al. 2012; Ling and Weld 2012]

The Wiki-KBP dataset uses a large number of sentences extracted from about 780,000 Wikipedia articles. To generate the training set, 1.5 million sentences are sampled automatically [Liu et al. 2017]. The test set is populated using manual annotation during the 2013 KBP slot-filling task. This dataset provides 153,966 instances and 13 different relation types [Chen et al. 2020].

5.5 FewRel and FewRel 2.0 [Han et al. 2018c]

FewRel is a dataset for classifying few-shot relationships. It contains 70,000 words in natural language, indicating 100 relationships taken from Wikipedia and annotated by crowd workers. Each sentence's relation is first determined by distance supervision, then processed by crowd workers. The train set (64 relations), validation set (16 relations), and test set (20 relations) are the three subgroups that make up the dataset. FewRel 2 is derived from FewRel by adding a new test set in a different domain.

5.6 NYT-H [Zhu et al. 2020]

In NYT-H, distantly supervised labeled training data is used for distantly supervised relation extraction, and many annotators are recruited to categorize test data. NYT-H may be utilized as a benchmark for distantly supervised relation extraction. With the help of deep neural networks, the field of distantly supervised relation extraction has made a lot of progress in recent years. Entity pairs from knowledge bases may be effectively aligned to sentences via distant supervision to create vast amounts of annotated data. These distant supervision generated datasets always feature inaccurate labels, which provide erroneous assessment scores throughout testing and may cause confusion among the investigators. A novel dataset called *NYT-H* has been offered to address this issue. The NYT-H dataset provides a much bigger test set than the earlier datasets, allowing researchers to conduct a more precise and continuous review. NYT-H is based on the NYT10 [Riedel et al. 2010] dataset.

5.7 WebNLG [Gardent et al. 2017]

The WebNLG corpus consists of triplet sets that use natural language text to describe facts (entities and their relations with one another). The corpus includes up to seven triplet sets, each with at least one reference text. The test set is divided into two groups: seen, which contains inputs generated for entities and relations associated with DBpedia groups visible in the training data, and unseen, which includes inputs retrieved for entities and relations associated with five overlooked groups. The dataset was first utilized for the WebNLG natural language generation task, which entails transferring sets of triplets to text and includes producing referring expressions, aggregate expressions, surface realizations, and sentence segmentation. The opposing task of triplet extraction also uses the corpus.

5.8 SciERC [Luan et al. 2018]

The SciERC dataset holds 500 scientific abstracts annotated with coreference clusters, relations, and academic entities. The Semantic Scholar Corpus extracted the abstracts from 12 AI conference/workshop sessions from four AI groups. SciERC benefits from SemEval 2017 Task 10 and SemEval 2018 Task 7 datasets, adds cross-sentence relations via coreference linkages, and expands entity types, relation types, and relation coverage.

5.9 ACE 2005 [Walker et al. 2006]

The LDC created the ACE 2005 Multilingual Training Corpus, which has more than 1,800 documents of mixed-genre text in English, Arabic, and Chinese with annotations for entities, relations, and events. The **Automatic Content Extraction (ACE)** technology assessment conducted in 2005 reflects educational data collection in those languages. Newswire, broadcast media, broadcast conversation, weblogs, forum discussions, and informal telephone communication are among the genres. The data was annotated with cooperation from the ACE Project, LDC, and other parties. The ACE program's goal was to create technology for automated content extraction that would assist in automatic text analysis.

An English-only version of the ACE 2005 dataset is also available [Mani et al. 2008].

5.10 BLURB [Gu et al. 2022]

The BLURB (Biomedical Language Understanding and Reasoning Benchmark) dataset provides biological NLP resources. It is a benchmark for PubMed-based biomedical NLP applications. BLURB uses 13 generally accessible datasets for six different tasks. To avoid overemphasizing tasks with large datasets (like NER), BLURB presents the macro average of all tasks as the primary score. BLURB's primary purpose is to reduce barriers to involvement in biomedical NLP and contribute to expediting advancement in this vitally important area for humanity and society.

5.11 BioRED [Luo et al. 2022]

The advancement of relation extraction systems in biomedicine is severely constrained because most current benchmarking datasets for biomedical relation extraction only concentrate on relationships of a single type (e.g., protein-protein interactions) at the sentence level. A total of 600 PubMed summaries were used to create BioRED, a pioneering biomedical relation extraction corpus with several entity types (e.g., gene/protein, illness, and chemistry) and relation pairings (e.g., gene-disease, chemical-chemical) at the document level.

Table 7 reveals the technical comparison of the mentioned datasets.

Each of the datasets introduced has its restrictions. Table 8 outlines several of the most significant limitations of relation extraction datasets.

Table 7. Comparison of Most Common Relation Extraction Datasets

Dataset	#Rel.	#Ins.	Other Statistics	Updated
TACRED	42	106,264	#Ent. Pair: 5,530; #Triple: 5,600; #Ent.: 2,999; #Sent.: 2,294; #Ins. in Test Set: 2,717; #Ins. in Test Set w/o NA: 3,325; % Neg.: 79.5%	Aug. 30, 2020
SemEval-2010 Task 8	19	10,717	#Ent. Pair: 10,233; #Triple: 10,281; #Ent.: 7,858; #Sent.: 10,674; #Ins. in Test Set: 2,717; #Ins. in Test Set w/o NA: 2,717; % Neg.: 17.4%	Feb. 29, 2020
ACE 2003–2004	24	16,771	#Words in English, Chinese, Arabic: 158,000, 154,000, 151,000	Feb. 15, 2006
ACE 2005–English	6	7,120	#Ent. Pair: 5,530; #Triple: 5,600; #Ent.: 2,999; #Sent.: 2,294	Jan. 22, 2008
DocRED (human annotated)	96	63,427	#Doc.: 5,053; #Word: 1,002k; #Ent.: 132,375; #Sent.: 40,276; #Fact: 56,354	June 17, 2019
DocRED (distantly supervised)	96	1,508,320	#Doc.: 101,873; #Word: 21,368k; #Ent.: 2,558,350; #Sent.: 828,115; #Fact: 881,298	June 17, 2019
Wiki-KBP	13	153,966	#Ent. Pair: 131,534; #Triple: 133,050; #Ent.: 40,415; #Sent.: 23,884; #Ins. in Test Set: 2,209; #Ins. in Test Set w/o NA: 316	N/A
FewRel	100	70,000	#Rel. in train set, validation set, and test set: 64, 16, 20	Dec. 8, 2021
NYT-H	22	667,806	#Ent. Pair: 375,829; #Triple: 377,393; #Ent.: 69,063; #Sent.: 320,668; #Ins. in Test Set: 9,955; #Ins. in Test Set w/o NA: 9,955	Oct. 26, 2020
NYT-10	58	742,748	#Ent. Pair: 375,914; #Triple: 377,495; #Ent.: 69,063; #Sent.: 320,711; #Ins. in Test Set: 172,448; #Ins. in Test Set w/o NA: 6,444	Nov. 16, 2010
SciERC	N/A	4,716	#Entities: 8,089; #Relations/Doc 9.4; #Coref links: 2,752; #Coref-clusters: 1,023	Aug. 15, 2018
BioRED	N/A	6,503	20,419 entity mentions; 3,869 unique concept identifiers	July 31, 2022

Note: % Neg. denotes the percentage of negative examples (no relation).

Table 8. Limitations of Relation Extraction Datasets

Method	Limitations
TACRED, SemEval-2010 Task 8, ACE 2003–2004, ACE 2005–English	• Restricted to a specified domain and entity types • Unbalanced distribution of relation types • Lack of cross-sentence relations
DocRED (human annotated)	• No multi-label or multi-class relations present in the dataset
DocRED (distantly supervised)	• Noisy because the distant supervision method is used to construct the dataset
Wiki-KBP	• Dataset limited to a specified domain and suffers from lack of multi-label or multi-class relations
FewRel, NYT-H, NYT-10, SciERC, BioRED	• Restricted to a particular domain • Lack of multi-label or multi-class relations

6 EVALUATION OF RELATION EXTRACTION

The nature of the approach (supervised or unsupervised) and the kind of dataset determine how well the evaluation of the relation extraction task is performed. This section will take a quick look at evaluating supervised, semi-supervised, and unsupervised approaches.

6.1 Evaluation of Supervised Methods

Relation extraction is a classification challenge in supervised techniques; therefore, measures including precision, recall, and F-measure are used to assess effectiveness [Bach and Badaskar 2011]. The following define these metrics:

$$Precision = \frac{Number\ of\ correctly\ extracted\ relations}{Total\ number\ of\ extracted\ relations} \tag{1}$$

$$Recall = \frac{Number\ of\ correctly\ extracted\ relations}{Actual\ number\ of\ extracted\ relations} \tag{2}$$

$$F\text{-}Measure\ \ (F1) = \frac{2\ \times\ Precision\ \times\ Recall}{Precision\ +\ Recall} \tag{3}$$

As the F1-score is the harmonic average of the corresponding mean recall and precision values, it is commonly used in relation extraction evaluation.

6.2 Evaluation of Semi-Supervised Methods

The evaluation of semi-supervised algorithms is altered without labeled test data, but the primary metrics (accuracy, recall, and F-measure) are still used. Large volumes of data are often used to apply semi-supervised algorithms for relation extraction, frequently identifying several new patterns and associations. As a result, it is challenging to quantify precision and recall in part. The result is assumed to be typical of a small sample chosen at random, and any genuine relationships are carefully verified. The concept from Section 6.1 (Evaluation of Supervised Methods) is then utilized to produce the approximate precision estimate. Enormous volumes of data make it challenging to determine the precise number of entity-relation pairs and calculate recall when evaluating semi-supervised techniques [Bach and Badaskar 2011].

6.3 Evaluation of Unsupervised Methods

In NLP, the assessment remains a challenge, particularly when unsupervised methods are used. To cope with the overwhelming volume of information from digital resources, particularly the Internet, unsupervised approaches in the information extraction field are becoming increasingly important. They minimize several drawbacks of supervised or semi-supervised techniques, such as the requirement to specify the kind of relations to concentrate on statically, the requirement to annotate a sizable number of instances, or the requirement to give a sizable number of seeds [Wang et al. 2012].

Four tasks (specifically linguistic preprocessing, candidate extraction, relationship filtering, and relationship grouping) make up the whole connection extraction procedure. The extracted candidates are grouped in the relation clustering phase to gather close relationships. It is crucial to provide a better perspective of the actual relations among named entities. This stage makes use of both a clustering technique and a similarity metric. Therefore, clustering techniques and their evaluation are crucial for unsupervised information extraction methods [Wang et al. 2012].

6.4 Limitations of Common Metrics

Precision and recall are highly sensitive to minor variations but lack sensitivity to significant variations. Because precision and recall have limits, even small changes to an information retrieval system can make a big difference in how well it works. In response to this deficiency, Mizzaro et al. [2002] introduced the ADM (Average Distance Measure) metric.

In addition, the precision/recall method encounters an issue of double penalization for misclassification, where a system can suffer both precision and recall penalties if it misclassifies a person as a location, resulting in a missing person and a false location. This problem extends to other metrics that employ binary evaluation, where the result is either correct or incorrect. In contrast, cost-based evaluation is designed to mitigate this issue by assigning distinct weights to various errors in a flat structure, resulting in a scalar score rather than a binary one. A possible solution to the misclassification problem is to utilize a metric that assigns error weights based on the similarity between the given and target responses. This can be accomplished by incorporating positional and commonality measures [Hartmann et al. 2005].

In relation extraction systems, there is a clear tradeoff between precision and recall. A system can achieve 100% precision by identifying nothing, thus avoiding relationship mistakes. However, it can achieve 100% recall by identifying everything, thus avoiding missed relationships. The F-measure is often used with precision and recall as a weighted average to balance this tradeoff. Precision and recall are considered equally important when the weight is set to 0.5 [Hartmann et al. 2005].

Precision, recall, and F1-score put much weight on how well the model predicts the future. Even though accuracy is essential, it can be misleading if the model is too cautious and leaves out a lot of important relationships. In these situations, even though the model may detect some relations accurately, the recall score will be poor.

6.5 Other Proposed Evaluation Metrics

Deußer et al. [2022] proposed an adjusted F1-score based on a weighting scheme as a new metric for performance measurement in the relation extraction task. The purpose of defining this new metric is to calculate the actual efficiency of relation extraction. In many cases, predictions ignore parts of non-numerical entities. It leads to a zero F1-score for both relations:

$$o_i = \left| e_{i,pred} \cap e_{i,gt} \right|, \tag{4}$$

where $e_{i,pred}$ is a set containing all token identifiers and $e_{i,gt}$ is ground truth. Using Equation (4), true positives, false negatives, and false positives are calculated as the following (i and j are two entities, and r demonstrates a relation between entities i, j):

$$TP_r = \frac{1}{2}\left(\frac{o_i}{n_{i,gt}} + \frac{o_j}{n_{j,gt}}\right), \tag{5}$$

$$FN_r = 1 - TP_r, \tag{6}$$

$$FP_r = \frac{1}{2}\left(\frac{n_{r,pred} - o_i}{n_{i,pred}} + \frac{n_{j,pred} - o_j}{n_{j,pred}}\right), \tag{7}$$

where $n_{i,.} := |e_{i,.}|$. We can use these equations to compute traditional measures such as recall, precision, and F1-score. Accordingly, the adjusted F1-score can be calculated by

$$F_{1,r} = 2 \times \left(\frac{precision_r \times recall_r}{precision_r + recall_r}\right). \tag{8}$$

7 REVIEW OF RELATION EXTRACTION METHODS IN PERSIAN

Compared to languages like English, there is a need for more research on semantic relation extraction in Persian. This section aims to introduce and analyze the most notable relation extraction methods used in the Persian language, with most studies utilizing the PERLEX dataset introduced in Section 5.1. We will begin by reviewing the most significant studies conducted using Persian text corpora and non-standard datasets before delving into a detailed introduction and comparing the research conducted using PERLEX.

7.1 Corpus-Based Methods

Shamsfard and Barforoush [2004] proposed a method based on an ontology engineering approach. Building general-purpose ontologies is a time-consuming task. Hasti, as an automatic ontology builder, aims to resolve this problem. Only a small part called the *kernel* is created manually in this system. It is a flexible system against environmental changes, which has a drawback: there is a possibility of a sharp decrease in speed in some queries and the first steps of the learning process.

In the work of Nasser et al. [2019], a distantly supervised method has been proposed for relation extraction from a sizable Persian corpus. Distant supervision has advantages, such as domain independence and the ability to populate a knowledge base, compared to supervised and unsupervised methods. This work uses an extensive knowledge base named *FarsBase*, which contains millions of relation instances. It adapts existing entities in these relations with Persian Wikipedia articles to create a dataset in a distantly supervised way. The relation extraction strategy relies on a piecewise CNN.

In the work of Farahani et al. [2021], ParsBERT, a BERT-based pretrained language model for Persian, has been proposed. It is a monolingual BERT with a higher performance for Persian text analysis than other multilingual language models. ParsBERT has achieved better results in important NLP tasks like relation extraction, sentiment analysis, and NER.

Atarod and Yari [2020] proposed a distantly supervised approach for the relation extraction task in Persian texts. They tried to extract relationships from Persian Wikipedia articles using a distant supervision based algorithm. This model comprises three components: preprocessing, pattern extraction, and relation extraction. The pattern extraction component extracts patterns at the entity level by mapping the relation samples gathered from Wikidata as a knowledge base. To extract relationships, all elicited patterns were matched with existing sentences in plain text.

FarsBase-KBP [Asgari-Bidhendi et al. 2021] is a hybrid architecture that combines six extractors (four for extracting information and two for extracting relations) to identify and extract semantic relations. A fusion module is built on top of the extractors, which increases the model's accuracy in predicting relationships between entities.

PTRC [Torbati et al. 2013] is a temporal RC system for Persian texts to predict temporal relationships between two events. The proposed method is based on a support vector machine mechanism powered by SSKs (String Subsequence Kernels) and convolution trees. It can be used in downstream applications like QA and text summarization. PTRC has tried to address the problem of compound verbs and Persian's free order of words.

With the emergence of pretrained language models like BERT and GPT, and the impressive impact of these models on NLP tasks, including relation extraction, researchers have also employed them for the Persian language. Taghizadeh et al. [2021] released Sina-BERT, a BERT-based pretrained language model to boost the quality of relation extraction and sentiment analysis, question classification, and retrieval in the medical domain. ParsBERT [Farahani et al. 2021], a transformer-based model for the Persian language, has been developed for use in various NLP tasks, including relation extraction.

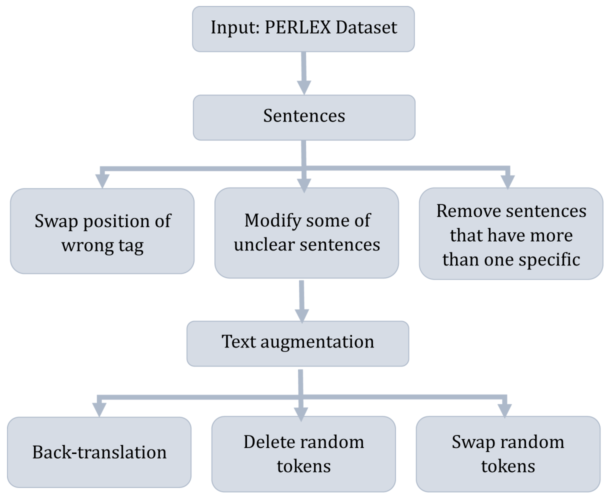

Fig. 1. Preprocessing and augmentation activities [Sartakhti et al. 2021].

7.2 PERLEX-Based Methods

In the work of Sartakhti et al. [2021], a data augmentation-based method was proposed for relation extraction in Persian, and they participated in the NSURL 2021 workshop. This study employed PERLEX [Asgari-Bidhendi et al. 2020] as the primary dataset and enhanced it by implementing certain preprocessing tasks. Another contribution of the research was increasing the size of the base dataset via data augmentation methods. Finally, the research has adopted two existing relation extraction models, including ParsBERT [Farahani et al. 2021] and multilingual BERT [Libovický et al. 2019], for relation extraction on the augmented version of the PERLEX dataset. Figure 1 illustrates text preprocessing and augmentation activities.

At the same time, Sartakhti et al. [2021] also introduced two other models, R-BERT+ParsBERT and RIFRE+ParsBERT, respectively. BERT+ParsBERT uses the famous Persian language model ParsBERT and has fine-tuned its parameters to the best values to perform relation extraction on the PERLEX dataset. The second model, BERT+ParsBERT, adopts the RIFRE [Zhao et al. 2021] model (a method that employs graph neural networks and a message passing mechanism for the relation extraction task) with the ParsBERT and multilingual BERT.

U-BERT [Jafari et al. 2021], a transformer-based relation extraction and classification method, uses the BERT language model over the Persian language. This model uses two approaches to improve accuracy. First, it oversamples the instances in smaller relation classes to cover inequality relation samples during the training process in different relation classes. Second, to soften the impact of the "Other" class, U-BERT defines a loss function using a pairwise ranking. The analyses confirm that the "Other" class is the noisiest.

At the same time as presenting U-BERT, Jafari et al. [2021] proposed another BERT-based approach named T-BERT. The proposed method extracts and uses syntactic features of sentences. Since the RC task is defined in terms of the shortest dependency path problem, T-BERT adds a new embedding layer at the input of the BERT architecture. For this purpose, the word vector is augmented with two specific tags: the POS tag and the dependency tree tag. In addition, a new layer has been added to the embedding portion of the BERT model. The output of the augmented layer is inserted into token embeddings and other embeddings.

In addition to the preceding, during the PERLEX dataset construction, Asgari-Bidhendi et al. [2020] presented six methods for extracting relationships from this dataset. Table 9 compares these six models with other methods introduced previously.

Table 9. Most Influential Methods Proposed for Relation Extraction in the Persian Language

Method	Description	Core Technologies	F1-Score on PERLEX
LRC + L2R LR	Utilizes features for training a logistic regression classifier with an L2R LR solver.	Logistic regression	57.42%
CNN on PERLEX	Uses: the method employs CNNs with kernel lengths ranging from 2 to 5, along with dropout and L2 regularization techniques to address the problem of overfitting.	CNNs	69.28%
Att-BLSTM	Employs single-layer BiLSTM with hidden state size 100 and recurrent/regular dropout, and the L2 regularization method to avoid overfitting.	BiLSTM	69.61%
BLSTM-LET	Employs a multi-head attention layer with four attention heads. The layer size is 50, and the hidden state is 300. To prevent overfitting, recurrent and regular dropout, and L2 regularization techniques, are utilized.	Attention mechanism + BiLSTM	70.79%
R-BERT on PERLEX	Fine-tunes the BERT-base, with hyperparameters set as follows: sentence length of 128, batch size of 16, Adam optimizer, and a learning rate of 3e-5.	Transformer-based language models (BERT)	75.31%
BERTEM-MTB	Fine-tunes the base BER, with hyperparameters set as follows: sentence length of 128, batch size of 16, Adam optimizer, and a learning rate of 3e-5.	Transformer-based language models (BERT)	77.66%
Sartakhti et al.	A data augmentation based method for relation extraction in the Persian language improves the PERLEX dataset by using some preprocessing tasks.	Transformer-based language models (BERT) + data augmentation	81.76%
U-BERT	Uses the ParsBERT language model and fine-tunes its parameters to perform relation extraction on the PERLEX dataset.	Transformer-based language models (BERT)	78.83%
T-BERT	Extracts and uses syntactic features of sentences and adds a new embedding layer at the input of the BERT language model.	Transformer-based language models (BERT)	76.97%
R-BERT + ParsBERT	A transformer-based approach for relation extraction and RC tasks that tries to boost the accuracy by oversampling the instances in smaller relation classes and defining a loss function using pairwise ranking to reduce the impact of the "Other" class.	Transformer-based language models (BERT)	79.11%
RIFRE + ParsBERT	Adopts the RIFRE [Zhao et al. 2021] model with ParsBERT and multilingual BERT.	Transformer-based language models (BERT)	83.82%

As can be seen in Table 9, the use of transformer-based language models, especially BERT [Devlin et al. 2018], and their different contributions have increased significantly in recent years. This is due to their high performance compared to previous methods. Before that, the most common methods were based on RNNs [Gupta et al. 2016], especially LSTM [Miwa and Bansal 2016], BiLSTM [Gupta et al. 2019], and BiLSTM along with the attention mechanism [Yuan et al. 2020].

8 PERSIAN DATASETS FOR INFORMATION AND RELATION EXTRACTION

This section introduces Persian language datasets for different tasks in the information extraction area. We categorize datasets into five groups: NER, relation extraction, dependency parsing, question answering, and knowledge bases.

8.1 Relation Extraction Datasets

PERLEX [Asgari-Bidhendi et al. 2020]

SemEval-2010 Task 8 is one of English's most widely used relation extraction datasets. In contrast, unlike English and other languages with abundant resources, Persian needs more resources for relation extraction. The PERLEX dataset represents the first Persian dataset for the relation extraction task and has been expertly translated. PERLEX is a parallel translation of each case in the SemEval-2010 Task 8 dataset. Nevertheless, this dataset is built on top of a previously published and extensively used dataset. Therefore, inferential comparisons are available between the outcomes of applying relation extraction approaches to this dataset and the English dataset. There are nine predefined relations, including "Entity-Destination," "Member-Collection," "Entity-Origin," "Cause-Effect," "Product-Producer," "Message-Topic," "Component-Whole," "Instrument-Agency," "Content-Container," and "Other." The "Other" class is used when there is no relationship between two entities.

8.2 NER Datasets

FarsiYar PersianNER [Asgarian 2021]

The Persian-NER repository contains a body of standard-tagged information. The information is extracted from Persian Wikipedia and currently contains about 25 million tokens in the form of about 1 million sentences. This corpus is published as open source. The information is labeled based on five categories: person names, organization names, place names, events, and time and date expressions.

ArmanPersoNERCorpus [Poostchi et al. 2016]

The first named entity dataset in Persian that has been fully annotated is ArmanPersoNERCorpus. Only academic research purposes were considered when it was distributed. There are 7,682 Persian sentences in the dataset, which has 250,015 tokens. Three different folds are accessible for use as separate training and testing data. One token and associated named entity annotation are included on each file line. There is a new line after each sentence. Person, organization (including banks, ministries, embassies, teams, nations, networks, and publishers), place (including cities, towns, streams, oceans, gulfs, deserts, and mountains), infrastructure (including schools, colleges, research institutions, airplanes, railroads, buildings, highways, bridges, terminals, institutions, parks, aquariums, and theaters), and commodity (including books, newspapers, TV programs, and movies) are the six categories.

PEYMA [Shahshahani et al. 2018]

PEYMA is a sizable, labeled Persian NER dataset freely available for academic use. The writers investigated conventional NER datasets created for English to create such a standard dataset, and they discovered that practically all of these datasets were created using news articles. As a result, they gathered information from 10 news websites. After examining the standards for creating CoNLL and MUC standard English datasets, they later established their own rules, taking Persian linguistic norms into account, to guide annotators in tagging these papers. All terms in papers may be classified as person, place, institution, time, date, percentage, currency, or some other by following these rules (words that are not in any other seven classes).

```
"question": "فولاد مبارکه چند بار برنده جایزه شرکت دانشی را کسب کرده است؟",
"is_impossible": false,
"id": 2
 },
 {
  "answers": [
   {
     "answer_start": 413,
     "answer_end": 447,
     "text": "تندیس زرین جایزهٔ ملی تعالی سازمانی"
   }
  ],
```


Fig. 2. One of the samples of PeCoQ dataset [Etezadi and Shamsfard 2021].

8.3 Question Answering Datasets

PersianQA [Ayoubi and Davoodeh 2021]

The PersianQA (Persian Question and Answering) dataset is a language learning resource based on the Persian version of Wikipedia. More than 9,000 records make up the dataset that was gathered from the public. Each item may be an inquiry with no possible answer or one with one or more answers found in the context. Using impossible or unanswerable questions makes it feasible to build a system that "understands that it does not know the answers," like the SQUAD 2.0 dataset.

Additionally, the collection includes 900 test data. The dataset's crowd workers are all Persian born and raised speakers. It is also important to note that the Wiki categories used to compile the contexts are comprehensive (historical, religious, geographic, and scientific). Each context now has three questions that cannot be answered and seven pairs with a single response.

PeCoQ [Etezadi and Shamsfard 2021]

PeCoQ is a dataset for question answering tasks in the Persian language. It contains 10,000 complex questions and answers extracted from FarsBase [Asgari-Bidhendi et al. 2021]. These 10,000 questions were selected from the 127,000 questions created at the beginning. Each question has a corresponding SPARQL query and two paraphrases that linguists wrote. This dataset has different levels of complexity, such as multi-relation, multi-entity, temporal, and ordinal conditions.

PeCoQ has some drawbacks: (1) its main focus is on 2-hops (instead of hops), (2) it needs to improve in postprocessing the answers to ordinal and temporal questions, and (3) it has only a single constraint in existing questions.

Figure 2 depicts one of the samples of PeCoQ.

8.4 Knowledge Bases

FarsBase [Asgari-Bidhendi et al. 2021]

FarsBase is a Persian-specific knowledge base, unlike BabelNet and DBpedia, which need more support for the Persian language. It confronts the same challenging issues as other knowledge bases, such as updating and growing with current information. Some knowledge bases, like Wikidata, depend on human resources to annotate structured data and guard against the introduction of false knowledge instances. FarsBase has a lower level of community support than Wikidata, which highlights the need for a method to extract information automatically and prevent incorrect relation examples from being given to the knowledge base.

8.5 Dependency Parsing Datasets

UPDT [Seraji 2015]

UPDT (Uppsala Persian Dependency Treebank) is a corpus annotated with dependency-based syntax. The treebank was created using a bootstrapping process using open source data-driven dependency parser MaltParser and hands annotation verification. It comprises 6,000 sentences (151,671 tokens) of text in CoNLL format. In 2013, the complete treebank was made available. Meanwhile, the treebank's first release, which included a seed dataset of 225 sentences, happened in the fall of 2011.

Universal Dependencies [Nivre et al. 2020]

To facilitate the creation of multilingual parsers and the study of parsing and cross-linguistic learning, the **Universal Dependencies (UD)** project aims to establish cross-linguistic consistency in treebank annotation across different languages. The annotation process involves utilizing a combination of Google's universal part-of-speech tags, an interlingua for morphosyntactic tag sets, and a set of Stanford dependencies that is continually evolving. The overarching goal is to deliver a global catalog of groups and rules that will enable uniform annotation of comparable constructs across languages while allowing for language-specific expansions.

LSCP 0.5 [Khojasteh et al. 2020]

LSCP (Large Scale Colloquial Persian Dataset) is a hierarchically organized semantic taxonomy emphasizing multi-task comprehension of the informal Persian language as a broad topic. With its dependency linkages in syntactic annotation, part-of-speech tagging, emotion polarity, and automated translation of original Persian phrases in five other languages, LSCP comprises 120 million lines from 27 million casual Persian tweets (in English, Czech, German, Italian, and Hindi). This dataset has a total of 120 million sentences.

MULTEXT-East [Erjavec 2010]

MULTEXT-East language resources is a multilingual dataset used for research and development in language engineering. It focuses on the morphosyntactic level of language descriptions. This dataset covers a significant number of primarily Central and Eastern European languages. The fourth update of these resources includes six more languages, including Persian and XML-encoded morphosyntactic requirements. This extensively documented and publicly accessible dataset is exceptional due to its broad coverage of languages and the variety of encodings it provides.

9 CHALLENGES IN THE RELATIONSHIP EXTRACTION AND DATASET CREATION

This section addresses challenges associated with constructing relation extraction datasets and extracting semantic relations in Persian.

Table 2 outlines the challenges in extracting various types of semantic relations, irrespective of the language. In summary, these challenges are various document formats and long dependencies in document-level relation extraction, low efficiency in cross-sentence reasoning and lack of enough training data in sentence-level relation extraction, lack of end-to-end models and insufficient training data in n-ary relation extraction;, the complexity of existing dependencies and need for the conditional dimension in conditional relation extraction, low efficiency of few-shot relation extraction, immaturity of multi-dimensional relation extraction methods, the problem of complex structures in NRE methods, complex relations, and the problem of unknown entities and relations in overlapping relation extraction.

9.1 Specific Challenges of Persian Language in Relation Extraction

Persian is a widely spoken language in the Middle East, with more than 110 million speakers worldwide [Rahat et al. 2018]. However, Persian is considered a low-resource language due to the need for labeled datasets and corpora for NLP model training. The most significant challenge for relation extraction in Persian is the lack of high-quality and voluminous datasets [Asgari-Bidhendi et al. 2021]. As discussed in Section 7, almost all researchers use the PERLEX dataset. Although the PERLEX dataset plays a vital role in improving relation extraction methods in Persian, creating a comprehensive annotated dataset from the Persian language corpus is recommended.

Extracting semantic relations (and, in general, processing textual data) in the Persian language faces several challenges, such as its ambiguity-prone nature and the free-word-order nature of this language [Asgari-Bidhendi et al. 2020]. Applying context-aware approaches and using semantic augmentation (i.e., using language resources like WordNet to provide additional information about word relationships) are two approaches that can be useful in such cases. Other recommended options include efficient use of attention mechanisms, employing encoder-decoder architectures, and large language models.

The English language typically structures relation clauses in the order of (*subject, relation-phrase, predicate*), resulting in less ambiguity in defining the boundary between *subject* and *predicate*. In contrast, the Persian language commonly uses the structure of (*subject, predicate, relation-phrase*). The "Ezfe" morpheme is used in Persian to join words together to form noun phrases. The distinction between the *subject* and *predicate* is ambiguous because the morpheme needs to be written in the text [Rahat et al. 2018]. Additionally, to create the related expression, the words may occasionally need to be rearranged. The potential for changing the object or predicate within the sentence is one of the difficulties. The explicit subject is frequently omitted; instead, the subject must be identified by examining the final component of the verb.

In addition, some verbs in this language can take on a complicated form consisting of multiple words, nouns, and prepositions linked to the verb. Ignoring any component of such verbs can result in associations with no real purpose [Rahat et al. 2018]. Our experience demonstrates that machine learning based methods using a sizable Persian text corpus are more effective than alternative approaches for addressing the issues caused by these aspects of the Persian language.

9.2 Dataset Creation Challenges

Relation extraction datasets are usually created in one of the following ways: (1) using distantly supervised methods or (2) using a hand-labeling approach [Asgari-Bidhendi et al. 2020]. The hand-labeling approach requires extensive human intervention because all existing instances of relationships must be labeled manually. This is a challenging, costly, and time-consuming process. Moreover, the quality of the datasets produced in this way depends on the human factors' accuracy. The two most famous datasets created in this way are SemEval-2010 Task 8 and TACRED.

However, in distantly supervised generated datasets, a knowledge base is used to identify instances of relationships between entities. NYT-10 is an example of a dataset made in this way. Distant supervised methods for dataset generation utilize knowledge bases like Wikidata [Vrandečić and Krötzsch 2014], DBpedia [Lehmann et al. 2015], YAGO [Hoffart et al. 2013], and Freebase [Bollacker et al. 2008] to annotate relations. Using these methods has become an accepted and common approach. However, the primary issue is generating a significant volume of noisy data, which impedes the acquisition of comprehensive knowledge. These methods require comprehensive knowledge bases containing detailed information about target relations [Claro et al. 2019].

In Table 10, we summarize the pros and cons of these two approaches.

Table 10. Advantages and Disadvantages of Dataset Creation Approaches for Relation Extraction

Approach	Disadvantages	Advantages
Distantly supervised	• Noisy labels • Low generalizability	• No need for human experts • High dataset generation speed • Suitable for downstream knowledge base applications • A vast amount of generated data
Hand labeling	• Expensive and time-consuming • Strong dependence on expert staff	• Flexibility in the generation process

10 CONCLUSION AND FUTURE DIRECTIONS

Relation extraction is the process of finding the proper relationships between two or more entities. In this article, we summarized the latest trend and progress of relation extraction tasks. We comprehensively reviewed different approaches and types of relation extraction in English and most of the proposed methods in Persian. Previous research was categorized, analyzed, and compared from different technical perspectives. The advantages, weaknesses, and limitations of each category were also explained. Due to the importance of data in relation extraction (especially deep learning based approaches that require a vast amount of data for training), we reviewed relation extraction datasets in English and Persian in detail. The technical characteristics and statistics of the datasets were also introduced and compared. In addition, we investigated the standard metrics of relation extraction evaluation along with some novel evaluation metrics and approaches. These metrics are suggested for unsupervised, supervised, and semi-supervised approaches. Finally, the obstacles and difficulties of dataset creation, the general challenges of relation extraction, and its specific challenges in Persian were discussed.

The future work can include a comprehensive review of bilingual and multilingual relation extraction methods. In addition, the review and analysis of pretrained language models and transformer-based methods can be beneficial. It is recommended to study and compare language models other than BERT and GPT, such as RoBERTa, XLNet, and T5. The models can be compared based on the number of parameters, the time needed for pretraining and execution, the number of layers, and how well they work in different domains. The major challenge is that you will need to fine-tune these models. In addition, sufficient data are not available in all domains to test models.

Additionally, we reviewed the presented methods for extracting relations from Persian text using the PERLEX dataset. This dataset is considered the most widely used and most important in the Persian language for the relation extraction task. Future work should introduce and classify the methods developed on Persian corpora like Farsi Wikipedia. Our investigations show that efforts are being made to build Persian datasets in the relation extraction field. It is suggested to identify and track these activities and introduce new datasets in future work.

REFERENCES

L. Abualigah, K. H. Almotairi, M. A. A. Al-qaness, A. A. Ewees, D. Yousri, M. A. Elaziz, and M. H. Nadimi-Shahraki. 2022. Efficient text document clustering approach using multi-search Arithmetic Optimization Algorithm. *Knowledge-Based Systems* 248 (2022), 108833. https://doi.org/10.1016/J.KNOSYS.2022.108833

L. Abualigah and M. Altalhi. 2022. A novel generalized normal distribution arithmetic optimization algorithm for global optimization and data clustering problems. *Journal of Ambient Intelligence and Humanized Computing*. Early access, (2022).

L. M. Abualigah, A. T. Khader, and E. S. Hanandeh. 2018. A new feature selection method to improve the document clustering using particle swarm optimization algorithm. *Journal of Computational Science* 25 (2018), 456–466. https://doi.org/10.1016/j.jocs.2017.07.018

H. Adel and J. Strötgen. 2021. Enriched attention for robust relation extraction. *arXiv Preprint arXiv:2104.10899* (2021).

Y. A. Alhaj, A. Dahou, M. A. A. Al-Qaness, L. Abualigah, A. A. Abbasi, N. A. O. Almaweri, M. A. Elaziz, and R. Damaševičius. 2022. A novel text classification technique using improved particle swarm optimization: A case study of Arabic language. *Future Internet* 14, 7 (2022), 194. https://doi.org/10.3390/fi14070194

E. Asgarian. 2021. Persian-NER: The large corpus of labeled persian named entity recognition. Retrieved April 18, 2023 from https://github.com/Text-Mining/Persian-NER.

M. Asgari-Bidhendi, B. Janfada, and B. Minaei-Bidgoli. 2021. FarsBase-KBP: A knowledge base population system for the Persian Knowledge Graph. *Journal of Web Semantics* 68 (2021), 100638.

M. Asgari-Bidhendi, M. Nasser, B. Janfada, and B. Minaei-Bidgoli. 2020. PERLEX: A bilingual persian-english gold dataset for relation extraction. Retrieved April 18, 2023 from http://arxiv.org/abs/2005.06588.

S. Atarod and A. Yari. 2020. A distant supervised approach for relation extraction in Farsi texts. *International Journal of Web Research* 3, 2 (2020), 1–8.

M. Aydar, O. Bozal, and F. Ozbay. 2020. Neural Relation Extraction: A Survey. Retrieved April 18, 2023 from http://arxiv.org/abs/2007.04247.

M. Y. Ayoubi and S. Davoodeh. 2021. PersianQA: A Dataset for Persian Question Answering. Retrieved April 18, 2023 from https://github.com/sajjjadayobi/PersianQA.

N. Bach and S. Badaskar. 2011. *A Review of Relation Extraction.* Carnegie Mellon University.

H.-R. Baek and Y.-S. Choi. 2022. Enhancing targeted minority class prediction in sentence-level relation extraction. *Sensors* 22, 13 (2022), 4911.

L. Baldini Soares, N. FitzGerald, J. Ling, and T. Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. *arXiv e-Prints arXiv:1906.03158* (2019).

K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data.* 1247–1250.

D. Chen, Y. Li, K. Lei, and Y. Shen. 2020. Relabel the noise: Joint extraction of entities and relations via cooperative multiagents. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* 5940–5950. https://doi.org/10.18653/v1/2020.acl-main.527

F. Christopoulou, M. Miwa, and S. Ananiadou. 2019. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. *arXiv Preprint arXiv:1909.00228* (2019).

D. B. Claro, M. Souza, C. Castellã Xavier, and L. Oliveira. 2019. Multilingual open information extraction: Challenges and opportunities. *Information* 10, 7 (2019), 228.

A. D. N. Cohen, S. Rosenman, and Y. Goldberg. 2020. Relation classification as two-way span-prediction. *arXiv Preprint arXiv:2010.04829* (2020).

T. Deußer, S. M. Ali, L. Hillebrand, D. Nurchalifah, B. Jacob, C. Bauckhage, and R. Sifa. 2022. KPI-EDGAR: A novel dataset and accompanying metric for relation extraction from financial documents. *arXiv:2210.09163* (2022). http://arxiv.org/abs/2210.09163

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv Preprint arXiv:1810.04805* (2018).

J. Ellis, X. Li, K. Griffitt, S. M. Strassel, and J. Wright. 2012. Linguistic resources for 2013 knowledge base population evaluations. In *Proceedings of the 2013 TAC KBP Workshop.*

T. Erjavec. 2010. MULTEXT-East version 4: Multilingual morphosyntactic specifications, lexicons and corpora. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10).*

R. Etezadi and M. Shamsfard. 2021. PeCoQ: A dataset for Persian complex question answering over knowledge graph. *arXiv:2106.14167* (2021). http://qald.aksw.org/.

H. Fadaei and M. Shamsfard. 2010. Extracting conceptual relations from Persian resources. In *ITNG2010 – Proceedings of the 7th International Conference on Information Technology: New Generations (ITNG'10).* 244–248. https://doi.org/10.1109/ITNG.2010.191

M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri. 2021. ParsBERT: Transformer-based model for Persian language understanding. *Neural Processing Letters* 53 (2021), 3831–3847. https://doi.org/10.1007/s11063-021-10528-4

C. Gardent, A. Shimorina, S. Narayan, and L. Perez-Beltrachini. 2017. Creating training corpora for NLG micro-planning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Long Papers) (ACL'17).* 179–188. https://doi.org/10.18653/v1/P17-1017

H. Gharagozlou, J. Mohammadzadeh, A. Bastanfard, and S. S. Ghidary. 2022. RLAS-BIABC: A reinforcement learning-based answer selection using the BERT model boosted by an improved ABC algorithm. *Computational Intelligence and Neuroscience* 2022 (2022), 7839840. https://doi.org/10.1155/2022/7839840

Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, and H. Poon. 2022. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare* 3, 1 (2022), Article 2, 23 pages. https://doi.org/10.1145/3458754

P. Gupta, S. Rajaram, H. Schütze, and T. Runkler. 2019. Neural relation extraction within and across sentence boundaries. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 1 (2019), 6513–6520.

P. Gupta, H. Schütze, and B. Andrassy. 2016. Table filling multi-task recurrent neural network for joint entity and relation extraction. In *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers (COLING'16).* 2537–2547.

X. Han, Z. Liu, and M. Sun. 2018. Denoising distant supervision for relation extraction via instance-level adversarial training. *arXiv Preprint arXiv:1805.10959* (2018a).

X. Han, P. Yu, Z. Liu, M. Sun, and P. Li. 2018b. Hierarchical relation extraction with coarse-to-fine grained attention. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.* 2236–2245.

X. Han, H. Zhu, P. Yu, Z. Wang, Y. Yao, Z. Liu, and M. Sun. 2018c. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP'18).* 4803–4809. http://arxiv.org/abs/1810.10147.

T. Hang, J. Feng, Y. Wu, L. Yan, and Y. Wang. 2021. Joint extraction of entities and overlapping relations using source-target entity labeling. *Expert Systems with Applications* 177 (2021), 114853. https://doi.org/10.1016/j.eswa.2021.114853

J. Hartmann, P. Spyns, A. Giboin, D. Maynard, R. Cuel, M. C. Suárez-Figueroa, and Y. Sure. 2005. *D1.2.3 Methods for Ontology Evaluation.* EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB Deliverable D, 1. University of Karlsruhe.

Z. He, W. Chen, Y. Wang, W. Zhang, G. Wang, and M. Zhang. 2020. Improving neural relation extraction with positive and unlabeled learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 5 (2020), 7927–7934.

I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. Ó. Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz. 2019. SemEval-2010 Task 8: Multi-way classification of semantic relations between pairs of nominals. In *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval'19).* 33–38. http://arxiv.org/abs/1911.10422.

J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence* 194 (2013), 28–61.

R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.* 541–550.

J. Y. Huang, B. Li, J. Xu, and M. Chen. 2022. Unified semantic typing with meaningful label inference. *arXiv Preprint arXiv:2205.01826* (2022).

Y. Y. Huang and W. Y. Wang. 2017. Deep residual learning for weakly-supervised relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP'17).* 1803–1807. http://arxiv.org/abs/1707.08866.

M. M. Jafari, S. Behmanesh, A. Talebpour, and A. N. Ghomsheh. 2021. Improving pre-trained language model for relation extraction using syntactic information in Persian. In *Proceedings of the 2nd International Workshop on NLP Solutions for Under Resourced Languages (NSURL'21).* 31.

H. Jiang, Q. Bao, Q. Cheng, D. Yang, L. Wang, and Y. Xiao. 2020. Complex relation extraction: Challenges and opportunities. *arXiv:2012.04821* (2020). http://arxiv.org/abs/2012.04821.

H. A. Khojasteh, E. Ansari, and M. Bohlouli. 2020. LSCP: Enhanced large scale colloquial Persian language understanding. *arXiv Preprint arXiv:2003.06499* (2020).

J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, et al. 2015. Dbpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.

J. Libovický, R. Rosa, and A. Fraser. 2019. How language-neutral is multilingual BERT? *arXiv:1911.03310* (2019). http://arxiv.org/abs/1911.03310.

X. Ling and D. S. Weld. 2012. Fine-grained entity recognition. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence.*

L. Liu, X. Ren, Q. Zhu, S. Zhi, H. Gui, H. Ji, and J. Han. 2017. Heterogeneous supervision for relation extraction: A representation learning approach. arXiv preprint arXiv:1707.00166.

T. Liu, X. Zhang, W. Zhou, and W. Jia. 2018. Neural relation extraction via inner-sentence noise reduction and transfer learning. *arXiv Preprint arXiv:1808.06738* (2018).

Y. Liu, W. Hua, and X. Zhou. 2019. Extracting temporal patterns from large-scale text corpus. In *Databases Theory and Applications.* Lecture Notes in Computer Science, Vol. 11393. Springer, 17–30. https://doi.org/10.1007/978-3-030-12079-5_2

Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP'18).* 3219–3232. http://arxiv.org/abs/1808.09602.

L. Luo, P.-T. Lai, C.-H. Wei, C. N. Arighi, and Z. Lu. 2022. BioRED: A rich biomedical relation extraction dataset. *Briefings in Bioinformatics* 23, 5 (2022), bbac282. https://www.teamtat.org.

I. Mani, J. Hitzeman, J. Richer, and D. Harris. 2008. *ACE 2005 English SpatialML Annotations.* Linguistic Data Consortium, Philadelphia, PA.

M. Mintz, S. Bills, R. Snow, and D. Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. 1003–1011.

M. Miwa and M. Bansal. 2016. End-to-end relation extraction using LSTMs on sequences and tree structures. *arXiv Preprint arXiv:1601.00770* (2016).

S. Mizzaro, V. Scienze, and L. Rizzi. 2002. A New Measure of Retrieval Effectiveness (or: What's Wrong with Precision And Recall). Retrieved April 18, 2023 from https://users.dimi.uniud.it/~stefano.mizzaro/researchpapers/ADM.pdf.

L. Mohammad and Q. Abualigah. 2018. *Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering*. Studies in Computational Intelligence (Vol. 816). Springer, Cham, Switzerland.

J. Moreira, C. Oliveira, D. Macêdo, C. Zanchettin, and L. Barbosa. 2020. Distantly-supervised neural relation extraction with side information using BERT. In *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN'20).* 1–7.

M. Nasser, M. Asgari, and B. Minaei-Bidgoli. 2019. Distant supervision for relation extraction in the Persian language using piecewise convolutional neural networks. In *Proceedings of the 2019 5th International Conference on Web Research (ICWR'19).* 96–99.

T. Nayak, N. Majumder, and S. Poria. 2021. Improving distantly supervised relation extraction with self-ensemble noise filtering. *arXiv Preprint arXiv:2108.09689* (2021).

T. Nayak and H. T. Ng. 2019. Effective attention modeling for neural relation extraction. *arXiv Preprint arXiv:1912.03832* (2019).

D. Q. Nguyen and K. Verspoor. 2019. End-to-end neural relation extraction using deep biaffine attention. In *Proceedings of the European Conference on Information Retrieval.* 729–738.

J. Nivre, M.-C. de Marneffe, F. Ginter, J. Hajič, C. D. Manning, S. Pyysalo, S. Schuster, F. Tyers, and D. Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC'20).* 4034–4043. http://arxiv.org/abs/2004.10643.

S. Park and H. Kim. 2021. Improving sentence-level relation extraction through curriculum learning. *arXiv Preprint arXiv:2107.09332* (2021).

S. Pawar, G. K. Palshikar, and P. Bhattacharyya. 2017. Relation extraction: A survey. *arXiv:1712.05191* (2017). http://arxiv.org/abs/1712.05191.

H. Peng, T. Gao, X. Han, Y. Lin, P. Li, Z. Liu, M. Sun, and J. Zhou. 2020. Learning from context or names? *An Empirical Study on Neural Relation Extraction. arXiv Preprint arXiv:2010.01923* (2020).

H. Poostchi, E. Z. Borzeshi, M. Abdous, and M. Piccardi. 2016. PersoNER: Persian named-entity recognition. In *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers (COLING'16).* 3381–3389.

M. Rahat, A. Talebpour, and S. Monemian. 2018. A recursive algorithm for open information extraction from Persian texts. *International Journal of Computer Applications in Technology* 57, 3 (2018), 193–206.

S. Riedel, L. Yao, and A. McCallum. 2010. Modeling relations and their mentions without labeled text. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases.* 148–163.

O. Sainz, O. L. de Lacalle, G. Labaka, A. Barrena, and E. Agirre. 2021. Label verbalization and entailment for effective zero- and few-shot relation extraction. *arXiv:2109.03659* (2021). http://arxiv.org/abs/2109.03659.

C. N. dos Santos, B. Xiang, and B. Zhou. 2015. Classifying relations by ranking with convolutional neural networks. *arXiv Preprint arXiv:1504.06580* (2015).

M. S. Sartakhti, R. Etezadi, and M. Shamsfard. 2021. Improving Persian relation extraction models by data augmentation. In *Proceedings of the 2nd International Workshop on NLP Solutions for Under Resourced Languages (NSURL'21).* 32–37. http://arxiv.org/abs/2203.15323.

M. Seraji. 2015. *Morphosyntactic Corpora and Tools for Persian.* Acta Universitatis Upsaliensis.

M. S. Shahshahani, M. Mohseni, A. Shakery, and H. Faili. 2018. PEYMA: A tagged corpus for Persian named entities. *arXiv Preprint arXiv:1801.09936* (2018).

M. Shamsfard and A. A. Barforoush. 2004. Learning ontologies from natural language texts. *International Journal of Human-Computer Studies* 60, 1 (2004), 17–63.

R. Socher, B. Huval, C. D. Manning, and A. Y. Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning.* 1201–1211.

N. Taghizadeh, E. Doostmohammadi, E. Seifossadat, H. R. Rabiee, and M. S. Tahaei. 2021. SINA-BERT: A pre-trained language model for analysis of medical texts in Persian. *arXiv Preprint arXiv:2104.07613* (2021).

M. E. Torbati, G. Ghassem-Sani, S. A. Mirroshandel, Y. Yaghoobzadeh, and N. K. Hosseini. 2013. Temporal relation classification in Persian and English contexts. In *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP'13).* 261–269.

D. Vrandečić and M. Krötzsch. 2014. Wikidata: A free collaborative knowledgebase. *Communications of the ACM* 57, 10 (2014), 78–85.

C. Walker, S. Strassel, J. Medero, and K. Maeda. 2006. ACE 2005 multilingual training corpus. *Linguistic Data Consortium, Philadelphia* 57 (2006), 45.

J. Wang. 2020. RH-Net: Improving neural relation extraction via reinforcement learning and hierarchical relational searching. *arXiv Preprint arXiv:2010.14255* (2020).

W. Wang, R. Besançon, O. Ferret, B. Grau, and W. Wang. 2012. Evaluation of unsupervised information extraction. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*. 552–558. https://hal.archives-ouvertes.fr/hal-02282037.

Z. Wei, J. Su, Y. Wang, Y. Tian, and Y. Chang. 2019. A novel hierarchical binary tagging framework for joint extraction of entities and relations. *arXiv Preprint arXiv:1909.03227* (2019).

S. Wu and Y. He. 2019. Enriching pre-trained language model with entity information for relation classification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management.* 2361–2364.

T. Wu, X. Li, Y.-F. Li, R. Haffari, G. Qi, Y. Zhu, and G. Xu. 2021. Curriculum-meta learning for order-robust continual relation extraction. *arXiv:2101.01926* (2021). http://arxiv.org/abs/2101.01926.

Y. Xu, L. Mou, G. Li, Y. Chen, H. Peng, and Z. Jin. 2015. Classifying relations via long short term memory networks along shortest dependency paths. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.* 1785–1794.

D. Yang, S. Wang, and Z. Li. 2018. Ensemble neural relation extraction with adaptive boosting. *arXiv Preprint arXiv:1801.09334* (2018).

Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou, and M. Sun. 2019. DocRED: A large-scale document-level relation extraction dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL'19).* 764–777. http://arxiv.org/abs/1906.06127.

Z.-X. Ye and Z.-H. Ling. 2019. Multi-level matching and aggregation network for few-shot relation classification. *arXiv Preprint arXiv:1906.06678* (2019).

C. Yuan, H. Huang, C. Feng, X. Liu, and X. Wei. 2019. Distant supervision for relation extraction with linear attenuation simulation and non-IID relevance embedding. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 1 (2019), 7418–7425.

Y. Yuan, X. Zhou, S. Pan, Q. Zhu, Z. Song, and L. Guo. 2020. A relation-specific attention network for joint entity and relation extraction. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI'20).* 4054–4060.

D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers (COLING'14).* 2335–2344.

X. Zeng, S. He, K. Liu, and J. Zhao. 2018. Large scaled relation extraction with reinforcement learning. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence.*

S. Zhang, D. Zheng, X. Hu, and M. Yang. 2015. Bidirectional long short-term memory networks for relation classification. In *Proceedings of the 29th Pacific Asia Conference on Language, Information, and Computation.* 73–78.

Y. Zhang, P. Qi, and C. D. Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. *arXiv Preprint arXiv:1809.10185* (2018).

Y. Zhang, V. Zhong, D. Chen, G. Angeli, and C. D. Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP'17).* 35–45. https://nlp.stanford.edu/pubs/zhang2017tacred.pdf.

K. Zhao, H. Xu, Y. Cheng, X. Li, and K. Gao. 2021. Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction. *Knowledge-Based Systems* 219 (2021), 106888.

K. Zhao, H. Xu, J. Yang, and K. Gao. 2022. Consistent representation learning for continual relation extraction. In *Findings of the Association for Computational Linguistics: ACL 2022.* Association for Computational Linguistics, 3402–3411. http://arxiv.org/abs/2203.02721.

Y. Zhao, H. Wan, J. Gao, and Y. Lin. 2019. Improving relation classification by entity pair graph. In *Proceedings of the Asian Conference on Machine Learning.* 1156–1171.

P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).* 207–212.

T. Zhu, H. Wang, J. Yu, X. Zhou, W. Chen, W. Zhang, and M. Zhang. 2020. Towards accurate and consistent evaluation: A dataset for distantly-supervised relation extraction. *arXiv abs/2010.16275* (2020).