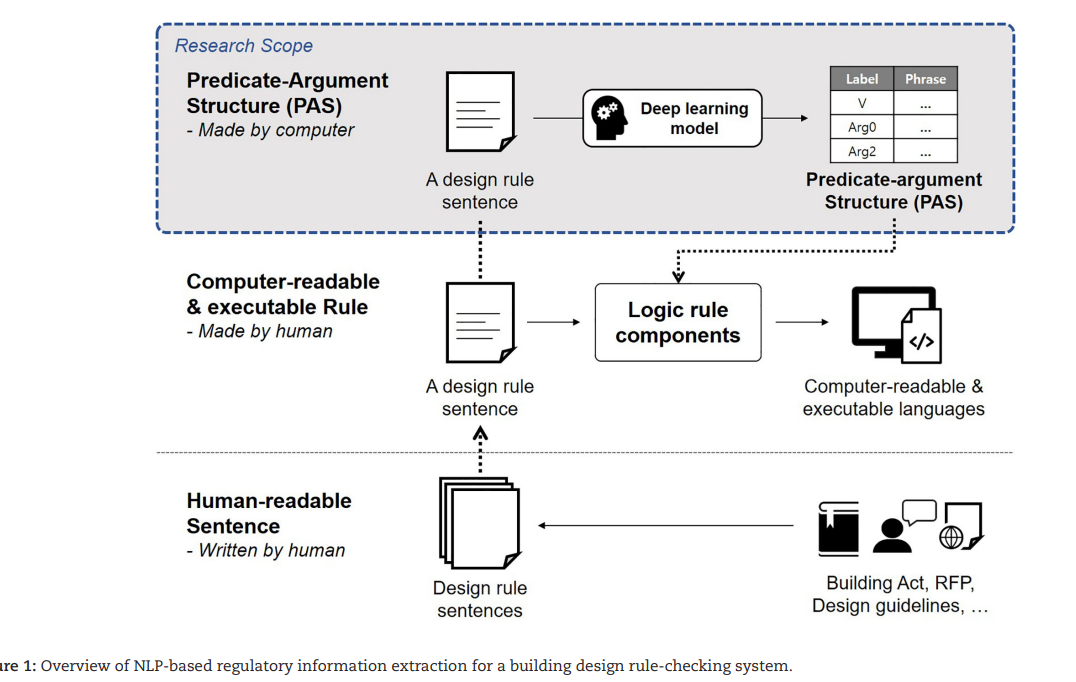
RESEARCH ARTICLE

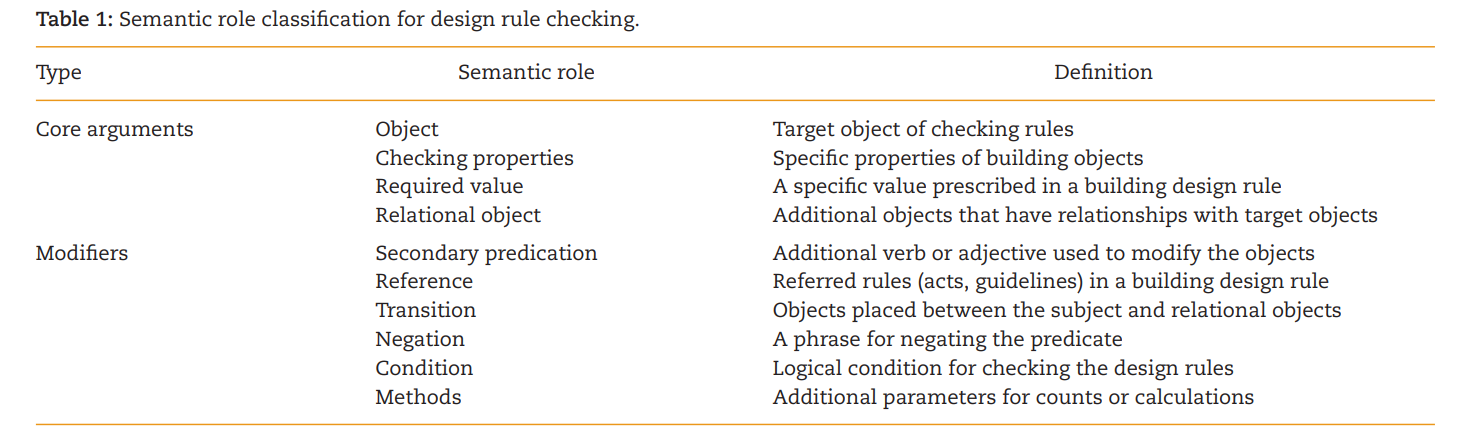
1. Deep learning-based extraction of predicate-argument structure (PAS) in building design rule sentences

**INTRODUCTION:-**

**rule-checking process is widely implemented with four steps:**

1. rule interpretation and logical structuring
2. building model preparation
3. rule execution
4. reporting of the checking results

* The scope of this paper is to propose a deep learning-based PAS extraction process and validate the proposed process by implementing the process
* In this study, we propose using a Bi-LSTM model to extract required regulatory information from Korean Building Act sentences.
* The biomedical discipline is one of the most active areas for employing information extraction techniques
* 



**BackGround:-**

Previous studies on rule-making approaches

* Nora El-Gohary – Name popped up many time
* KBimLogic is a logic rule-based mechanism that was developed to translate Korean Building Act sentences into a computer-readable script language
* KBimLogic is composed of three logical elements:
  1. building objects and properties (noun phrases)
  2. methods for checking (predicates)
  3. logical relationships between sentences.
* KBimLogic enables architects and rule-checking experts who are not familiar with programming to translate natural language-based legislations into a computer-readable format, named KBimCode.
* Furthermore, KBimCode is managed with a meta-database that accumulates the translated script code data for each logical element. However, KBimCode database is limited in the specific corpus derived from the target sentences. The corpus data should be expanded manually like other logic rule-based approaches. Machine learning techniques can contribute to alleviating the time and cost for expanding the scope of KBimLogic

General information extraction process using the NLP technique

Here, mainly 2 techniques are used

* + 1. **Named Entity Recognition (NER) -** A named entity refers to a specific noun word that can be classified into specific categories, such as a person’s name, a country, an organization, or a numeric expression. Extracting and classifying entities from sentences can help to identify the exact information associated with a specific entity mentioned in a sentence
    2. **Semantic Role Labeling (SRL) -** SRL is more focused on the semantic relationship between entities and is regarded as a shallow semantic processing task. SRL extracts the semantic role of each entity based on the meaning of the predicates. The output of the SRL is a PAS that represents the relationship and semantic role of each argument.
* RNN, LSTM, and bidirectional LSTM (Bi-LSTM) have recently been used for sequence labeling tasks (including NER tasks) and have demonstrated outstanding performance

NLP-based information extraction for design requirement analysis

The words representing the building objects and their properties can be captured by a lexical analysis comparing the input words and the defined words in a database.

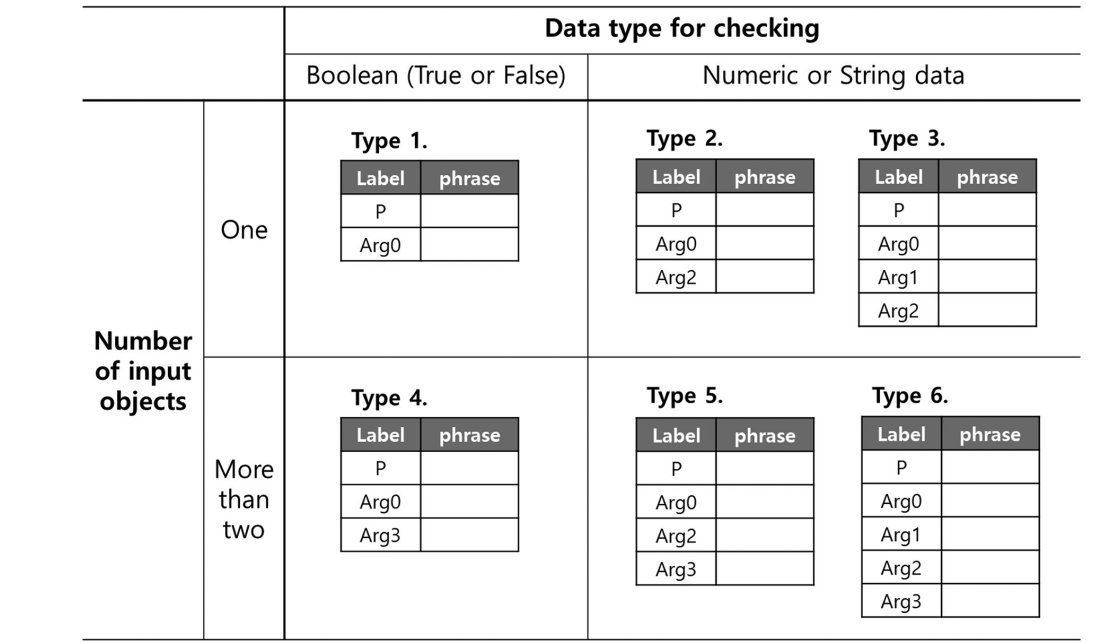
PAS Extraction in Building Design Rule Sentences for Design Rule Checking

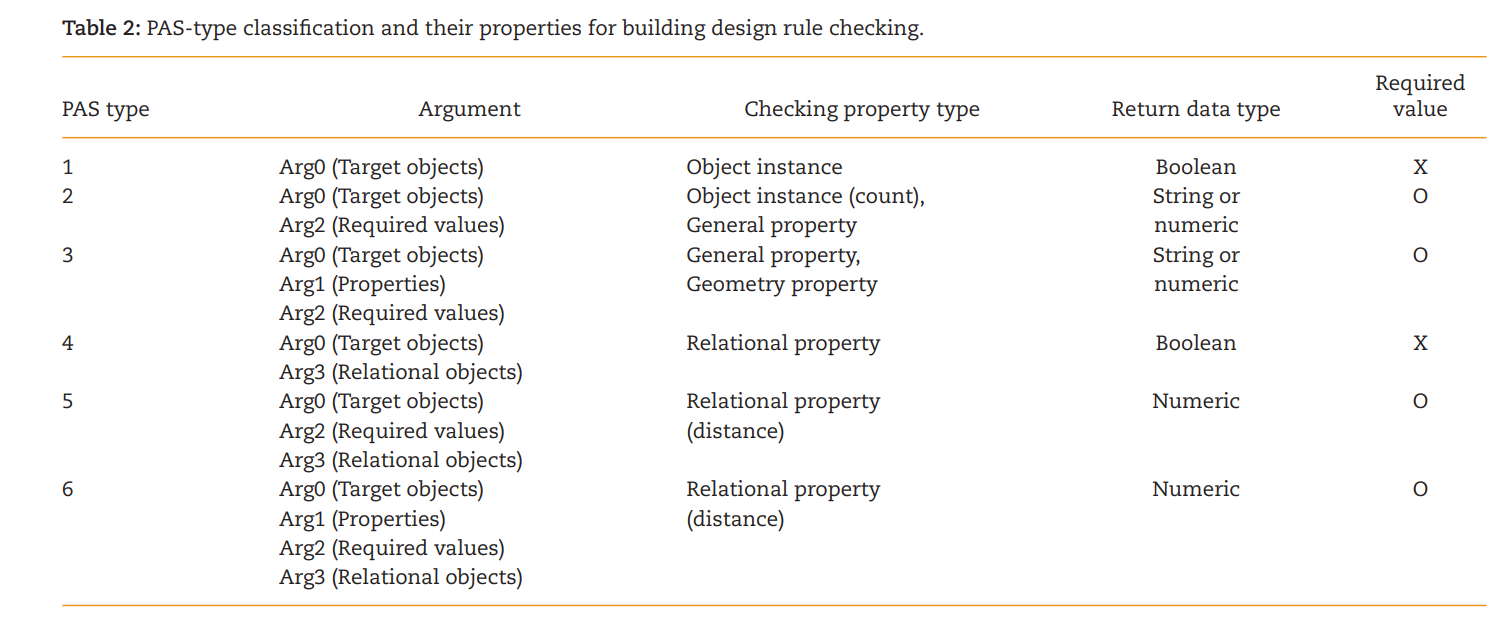
3.1 Classification of regulatory information in building design rule

The arguments are classified into core arguments (numbered arguments) and modifiers to construct the semantic structure. PropBank has five numbered arguments (agent, patient, instrument, starting point, and ending point) and several other modifiers.

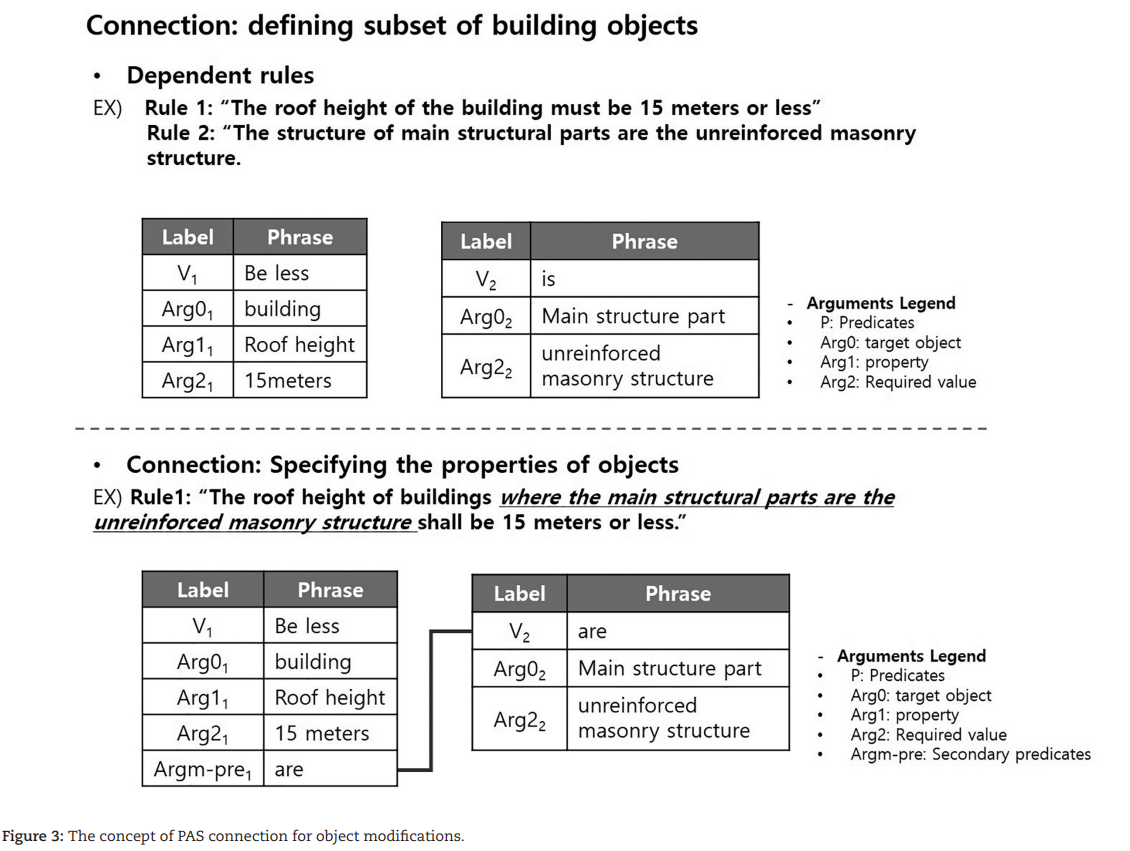
3.2. PAS for building design rule sentences

We identified the six PAS types by analyzing the syntactic constituent of building design rule sentences and their semantic meaning (Fig. 2; Table 2):



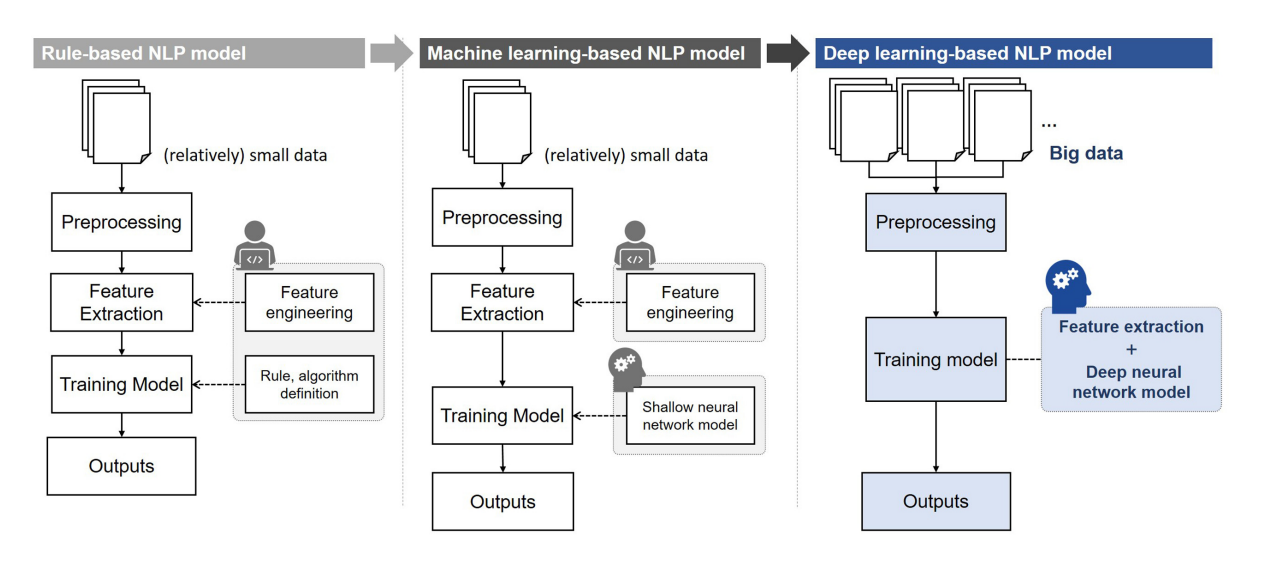


1. PAS Type 1 is a structure that only requires the target objects (Argument 0). Building design rule sentences that exclusively verify the existence of objects are classified in this type. Using natural language, these types of sentences are expressed with an S + V + O structure or as imperative sentences, such as “architects have to install something” or “install walls.” This structure is a translated structure for checking the installation or existence of building objects
2. AS Type 2 is a structure composed of target objects (Argument 0) and the required values (Argument 2). In this case, a sentence has a syntactic structure composed of a complement and a subject. The building design rule sentences that regulate the number of objects or specify the general properties such as material and functional usage are translated into Type 2. The specific checking property is not mentioned in this type of sentence, but can be inferred by the semantic meaning of the required values.
3. PAS Type 3 has specific checking properties (Argument 1) with a target object (Argument 0) and its required value (Argument 2). Similar to Type 2, Type 3 is also translated from the sentence that regulates the specific value of the properties. Some required values for general properties imply which properties are related to the values. However, geometric properties must be specified in sentences because quantitative values for geometric properties can be used without any distinction. The geometric property is used to determine how to implement the low-level algorithm that calculates the required properties. Target objects and their properties are usually expressed in the possessive form in Korean sentences, such as “interior finishing of wall” and “width of door.” These phrases must be separated to objects and their properties to clarify the purpose of sentence
4. PAS Type 4 is a structure for checking the relationship between different objects, which requires a pair of building objects to process the checking (Argument 0 and Argument 3). This paper focused on the physical relationships such as inclusion, connection, and adjacency. The results of checking relationships are Boolean (true or false) values, thus not needing the required value for checking.
5. S Type 4 is a structure for checking the relationship between different objects, which requires a pair of building objects to process the checking (Argument 0 and Argument 3). This paper focused on the physical relationships such as inclusion, connection, and adjacency. The results of checking relationships are Boolean (true or false) values, thus not needing the required value for checking.



3.3. Deep learning-based PAS extraction

The proposed method was developed with a assumption that input sentences are entered by human users and it does not guarantee that the given sentence is related to building design. Accordingly, the proposed process must be able to classify whether the given sentence is architecture related first and recognize the semantic role of each word



Neural network-based vector representation enables computer to learn the semantic meaning of each word by analyzing the concurrent word’s information. It also enables to infer how to translate the compound words or words from out of corpus. Through these features, neural network-based representation algorithms automate a feature extraction and reduce the dimension of vector representation. This eliminates the need for manual feature engineering and enhances the computation efficiency.

4. Development of Deep Learning and NLP-Based PAS Extraction Process