A DATABASE-BASED RATHER THAN A LANGUAGE MODEL-BASED NLP METHOD

Anonymous authors
Paper under double-blind review

ABSTRACT

Language model pre-training for NLP tasks take natural language as the direct modeling object. However, we believe that natural language is essentially a way of encoding information (knowledge). Therefore, the study object of language should be the information encoded in language, and the organizational and compositional structure of the information described in language. Based on this understanding, we propose a database-based NLP method that changes the modeling object from language to the information encoded in language. On this basis, 1) sentence generation task is transformed into read operations implemented on the database, and some sentence encoding principles to be followed; 2) sentence understanding task is transformed into sentence decoding principles and a series of Boolean operations implemented on the database; 3) learning task can be achieved by writing operations. Our method is more closer to how the human brain processes information and has excellent interpretability and scalability.

1 INTRODUCTION

Enabling machines to understand and use natural language as humans do is the ultimate goal of NLP. Many language models have been developed for related NLP tasks. For example: Word2Vec Mikolov et al. (2013) and GloVe Pennington et al. (2014) model the correlations between words by constructing numerical representation of words (i.e., word vector) and expect to obtain a word-level understanding by computing the similarities between the word vectors. Seq2seq Sutskever et al. (2014) and Transformer Vaswani et al. (2017) are used for translation tasks, they model the mapping relations between words and the mapping relations between sentence structures in different languages. GPT Radford et al. (2018) and Bert Devlin et al. (2018), which pre-train language models on a large-scale corpus, aim to model the sequence features in the corpus.

However, language model-based approaches overlook the fact that language is just one way of encoding information (knowledge). Besides language, we also use other tools to encode and transmit information (knowledge) between each other, such as body gestures, faces, drawings, etc. Therefore, our research object should be the information (knowledge) encoded and represented by language, not just language. Here, we propose a database-based NLP method that aims to discover and reveal how information (knowledge) is organized and stored in human brains, then simulate the findings (structures) to build databases (models), and finally provide the solutions of NLG and NLU (including the learning part) tasks based on these models.

To summarize the contribution of this work:

- Our method changes the modeling object from language to the information (knowledge) represented by language, giving the model we construct excellent interpretability and scalability.
- We propose a brand new NLP approach that is different from rule-based and statistical model-based (i.e., language model) approaches, and it is more closer to the way the human brain processes information.
- Based on the different and deeper understanding of natural language, we create a new framework for solving NLP problems, as well as bringing new thought forn AI research.

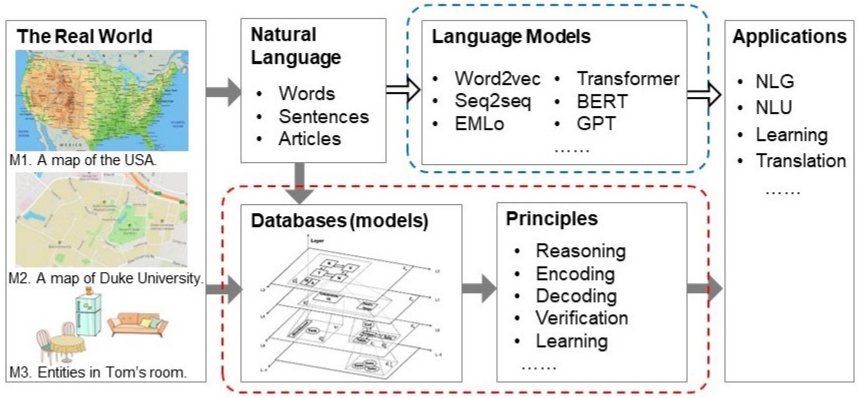


Figure 1: Differences in modeling objects. Language model-based methods take natural language as the object to construct the model. Database-based methods take both the information in the real world and the organizational and compositional structure of the information described in natural language as the objects to construct the model.

The United States *is south of* Canada.	Duke University *is in* North Carolina.
The cat is *in* Tom's room.	The table *is in front of* the fridge.
The cat *is on top of* the fridge.	The sofa *is next to* the fridge.

Table 1: Examples of sentences that describe the spatial position of target entities in the real world

2 METHODOLOGY

As shown in the red box in Figure 1, our method consists of two parts: 1) databases (models) and 2) principles. The databases part is devote to find out how information (knowledge) is organized and stored in the human brain, accordingly to simulate these structures to build models that can be used to generate databases. The principles part simulate how the information (knowledge) is processed and utilized in the human brain to accomplish related human-like intelligent activities.

2.1 BACKGROUND

As we learned in neuroscience, people constantly receive information from the outside world through their eyes, ears, noses, and other neural pathways Mark F. Bear (2004). This fact creates information (knowledge) gaps between people. Humans use language as one of the tools to reduce the gap, mainly by encoding the information (knowledge) in the human brain into language and exchanging it with each other. There are several steps prior to generating language, such as determining what is to be output, reasoning (i.e., pre-processing the meta-knowledge or information), encoding, etc. We should note that different types of information (knowledge) encoded in language must be modeled and processed according to their different nature and characteristics. In this paper, we only take the spatial position information (or knowledge) as the example to demonstrate our methodology.

People exchange the spatial position information (knowledge) of real-world entities by encoding them in sentences, as shown in Table 1. Looking at these sentences, we can see that they have the same structure: (Entity 1) + (...) + (Spatial relation) + (Entity 2).) + (spatial relation) + (entity 2), where "entity 1" is the **target entity** whose spatial position we want to describe by the sentence, "entity 2" is a **helper entity** that helps to locate the target entity, and "spatial relation" describes the spatial relation between the target entity and the helper entity. Table 2 shows three types of spatial relations commonly used in languages: 1) spatial range relations, 2) spatial directional relations, and 3) spatial distance relations. The spatial direction relations can be further divided into 2.1) absolute direction relations and 2.2) relative direction relations according to the different reference systems.

The above findings in language reveal how the knowledge (i.e.,the spatial position of real-world entities) is organized and stored in the human brain. We can also see that people are used to using

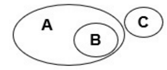
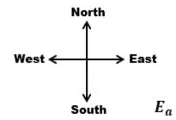
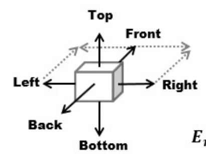
Spatial relations		Lexical representations	Reference system
1.Range relations		**Inside**: in, at... **Outside**: outside of...	
2. Directional relations	2.1 Absolute directional relations	**East**: east of... **West**: west of... **North**: the north side of ... **South**: the south side of...	
	2.2 Relative directional relations	**Top**: on, above, over, on top of... **Bottom**: below, under, beneath... **Left**: left of... **Right**: the right side of... **Front**: before, in front of... **Back**: behind, back of...	
3.Distance relations		by, near, next to, beside...	

Table 2: Classification of the relative spatial relations between entities, and lexical representations of the relative spatial relations.

entities with a relatively stable spatial position (immovable entities) as the helper entities. The immovable entities and the spatial relations between them form a stable system that we will use to construct our model.

2.2 Database (Model Architecture)

We construct a TGHM (**T**ree **G**raph **H**ybrid **M**odel) to describe and store the spatial position of real-world entities. In a TGHM, the real-world entities are abstracted as nodes; the spatial relations between the nodes are abstracted as directed edges E. The TGHM is a data structure that consists of two types of basic models: Tree and Graph.

2.2.1 Tree Model

We use a tree model to describe the **spatial range relations** (E_s) between entities. E_s is consist of two opposite directions, i.e., $E_s = \left\{\xrightarrow{inside}, \xleftarrow{outside}\right\}$. For example, we use the tree in Figure 2 to describe the spatial range relations between entities "North Carolina", "Duke University", "Tom's room","Table", "Cat", etc. The tree in Figure 2 can also be written in tabular form as shown in Table 3. In a tree, the child nodes with the same parent node should be spatially independent of each other, which means, there is no spatial range inclusion relation between them, if not, the child node must be moved up or down until all the child nodes are spatially independent of each other.

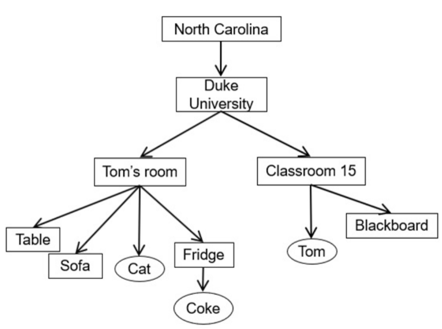


Figure 2: The tree that describes the spatial range relations between entities "North Carolina", "Duke University", 'Table", "Cat", etc.

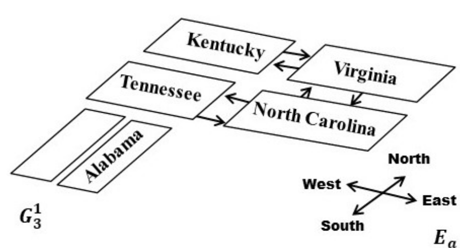


Figure 3: The graph that describe the absolute spatial directional relations between some entities in M1 in Figure 1.

E_s \ V	**North Carolina**	**Duke University**	**Tom's room**	**Classroom 15**	**Fridge**
$\xrightarrow{inside}$	Duke University	Tom's room, Classroom 15	Table, Sofa, *Cat*, Fridge	Blackboard, *Tom*	*Coke*
$\xleftarrow{outside}$	$\emptyset$	North Carolina	Duke University	Duke University	Tom's room

Table 3: The tabular form of the tree in Figure 2

2.2.2 Graph Model

We use graph models to describe the spatial directional relations between entities. The **spatial directional relations** can be future divided into **1) absolute directional relations** (E_a), which consists of four fixed directions, i.e., $E_a = \left\{\xrightarrow{east}, \xrightarrow{west}, \xrightarrow{north}, \xrightarrow{south}\right\}$, and **2) relative directional relations** (E_r), which consists of six fixed directions, i.e., $E_r = \left\{\xrightarrow{left}, \xrightarrow{right}, \xrightarrow{front}, \xrightarrow{back}, \xrightarrow{top}, \xrightarrow{bottom}\right\}$. Now, we can use the graph in Figure 3 to describe the absolute directional relations between some entities in M1 in Figure 1, and use the graph in Figure 4 to describe the relative directional relations between the entities in M3 in Figure 1. These two graphs can also be written in tabular forms.

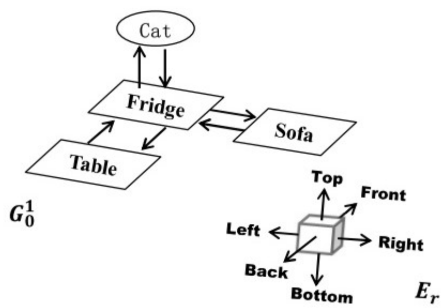


Figure 4: The graph describes the relative spatial directional relations between the entities in M3 in Figure 1

E_r \ V	**Table**	**Fridge**	**Sofa**
$\xrightarrow{left}$	$\emptyset$	$\emptyset$	Fridge
$\xrightarrow{right}$	$\emptyset$	Sofa	$\emptyset$
$\xrightarrow{front}$	Fridge	$\emptyset$	$\emptyset$
$\xrightarrow{back}$	$\emptyset$	Table	$\emptyset$
$\xrightarrow{top}$	$\emptyset$	*Cat*	$\emptyset$
$\xrightarrow{bottom}$	$\emptyset$	$\emptyset$	$\emptyset$

Table 4: The tabular form of the graph in Figure 4

2.2.3 TGHM

Finally, taking the nodes common to the tree and the graphs in Figures 2, 3 and 4 as connection points, we can integrate the tree and the graphs into the tree-graph hybrid model (TGHM) as shown in Figure 5. TGHM describes the spatial range relations between entities on the vertical structure (i.e., the inter-layer structure) ; and the spatial directional relations between entities on the horizontal structure (i.e., the intralayer structure). Each layer of a TGHM can accommodates multiple subgraphs. Usually, the E_a (absolute directional relations) is used as the reference frame of the whole layer, and the E_r (relative directional relations) is used as the reference frame in each subgraph. As shown in Figure 5, the subgraph G_0^2 and G_0^1 take E_r as their reference frame, and the layer L0 is using E_a as its reference frame. In a TGHM, the immovable nodes and the edges between them form a stable frame, which is a new reference frame in addition to the widely used numerical positioning system (e.g., GPS). TGHM also provides a bridge for information (knowledge) exchange between language and the numerical reference frame, as shown in Figure 6.

TGHM can be continuously extended upwards and downwards in the vertical structures to add new nodes, and continuously subdivided in the horizontal structure to add new nodes. Therefore, the TGHM could satisfy people's need to describe and store the spatial position of numerous entities in

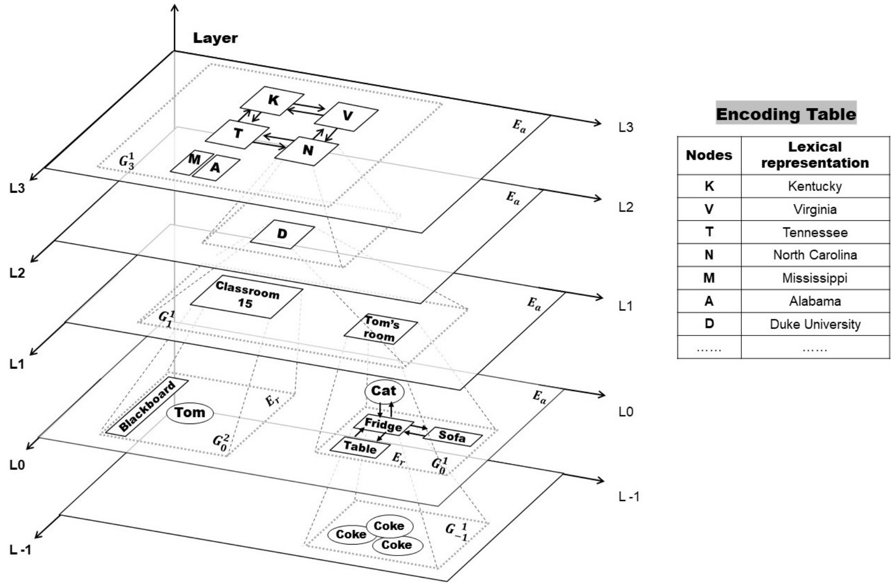


Figure 5: A perspective example of the TGHM, which describes the spatial range relations between entities in the vertical structure (inter-layer structure) and the spatial directional relations between entities in the horizontal structure (intralayer structure). The encoding table stores the lexical representations in different languages corresponding to each node in the TGHM. Here, we only list English as an example.

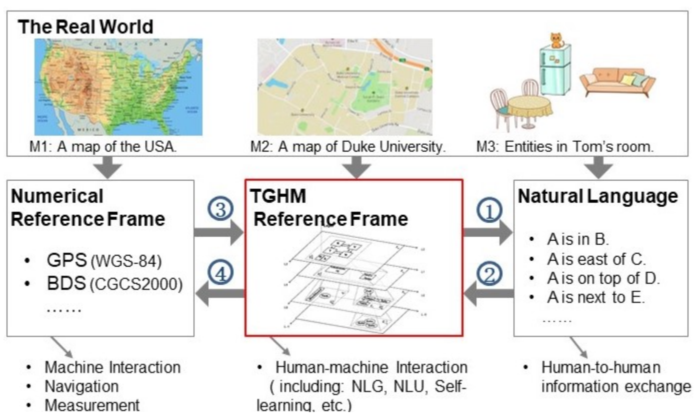


Figure 6: Three ways of describing (or encoding) the spatial position of real-world entities. The information exchange routes between different systems: ① sentences generation. ② sentences understanding.③ search for neighboring entities. ④ get the numerical position of the target entities.

the real world. When the spatial position of an entity changes, the corresponding data in the TGHM is simply modified accordingly. In addition, we can also build datasets to store the spatial position of the movable entities to record their footprint. Therefore, TGHM allows us to simulate how humans organize and store the spatial position of real-world entities in the brain, which means we can create memories for machines. Together with some corresponding data processing principles, the machine is able to process and utilize the data in the TGHM, and this is a practice of artificial intelligence.

2.3 Data Processing Principles

TGHM consistent with the characteristics of the structural-mechanistic kind of model Schölkopf & Bernhard (2015) which follow an underlying mechanistic understanding of reality. Thus, the TGHM can be used for many purposes. In this paper, we only present how the data in the TGHM has been processed and utilized in NLG and NLU (including learning part) tasks.

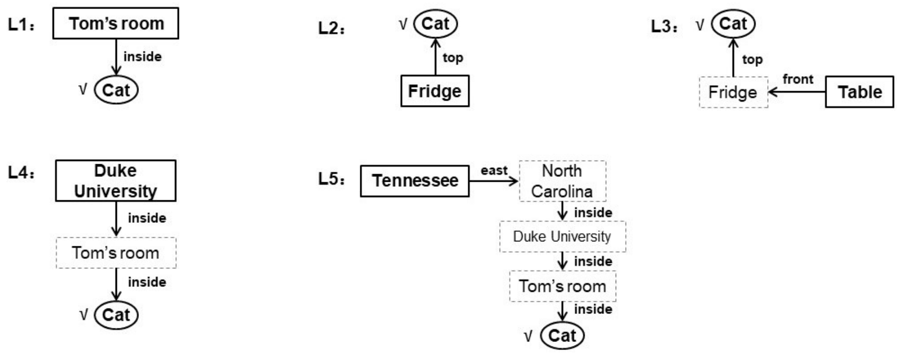


Figure 7: If we take the entity "Cat" as the target, then we can find the above 5 data chains in the TGHM to help locate the entity "Cat". The nodes marked with "✓" are the target nodes.

2.3.1 Natural Language Generation

Since, language is a tool used to convey information (knowledge). The NLG task can be further broken down into **subtask 1**- determining the information (knowledge) that needs to be conveyed, and **subtask 2**- encoding that information (knowledge) into sentences.

- DATA READING

In this paper, the knowledge to be conveyed is the spatial position of the target entity. Following the language expression, we use a helper entity and the spatial relations between the helper and the target entity to describe the spatial position of the target entity. For example, if we want to describe the spatial position of the entity "Cat", we first need to find the corresponding node (target node) of the entity "Cat" in TGHM in Figure 5, then find the helper nodes that have a spatial relation with the target node, such as the nodes "Tom's room", "Fridge", "Table", "Duke University", "Tennessee" and so on, then we can get 5 corresponding data chains as shown in Figure 7, which are composed of the target node, the helper node, and the spatial relations between them. Each of these 5 data chains can describe the spatial position of the entity "Cat", but their precision is different. If we sort these 5 data chains by precision, we can get the following result: $L2 > L3 > L1 > L4 > L5$. However, when generating a sentence, precision is not the only goal we are pursuing, if we want to describe the spatial position of the entity "Cat" to a particular person, we also need to know how much this person knows about the spatial position of the 5 candidate helper nodes? and what is the person's requirement for descriptive precision? So that we can filter out the appropriate one accordingly. Here, we will skip this part and go straight to the sentence encoding part. Filtering out the appropriate helper node and reading out the data chain is the goal of **subtask 1**.

- ENCODING PRINCIPLES

Although the principles for encoding a data chain into a sentence vary slightly from language to language, but the following parts are requisite: 1) the target and the helper nodes in the data chain, 2) the spatial relation between the target node and helper node, and 3) judgment of the spatial relation.

Existence judgment of the spatial relation: A specific spatial relation may or may not exist between two nodes. In language we use "be" and "be not" to describe these two statuses. For example, the words "is" and "is not" in row 4 of Table 5 describe whether the spatial relation in row 3 exist or not.

Reasoning of the spatial relations: In the TGHM, the spatial relation between any two nodes can be calculated. We have summarized some reasoning principles according to how humans process them; see examples below:

- *Elimination operation:* e.g., $\xrightarrow{inside} + \xleftarrow{outside} = \emptyset$, $\xrightarrow{left} + \xrightarrow{right} = \emptyset$, $\xrightarrow{north} + \xrightarrow{south} = \emptyset$...
- *Union operation:* e.g., $\xrightarrow{inside} + \xrightarrow{inside} = \xrightarrow{inside}$, $\xrightarrow{east} + \xrightarrow{east} + \xrightarrow{north} = \xrightarrow{northeast}$...
- *Hybrid operation:* when a data chain contains both spatial range relations and spatial directional

	Main Parts \ Data Chain	L1	L1*	L2	L2*
1	Target node	The cat	The cat	The cat	The cat
2	Helper node	Tom‘s room	Tom‘s room	the fridge	the fridge
3	Spatial relation (E)	in	*on top of*	on top of	*in*
4	Existence judgment of the E: • True • False	is	*is not*	is	*is not*

Table 5: Examples of the requisite parts for encoding a data chain.

Data Chain	Target Node	Existence Judgment of the E	Spatial Relation	Helper Node
L1	The cat	is	in	Tom‘s room
L2	The cat	is	on top of	the fridge
L3	The cat	is	in front of (next to)	the table
L4	The cat	is	in	Duke University
L5	The cat	is	on the east side of	Tennessee
L1*	The cat	*is not*	*on top of*	Tom‘s room
L2*	The cat	*is not*	*in*	the fridge

Table 6: Examples of sentence encoding for the data chains in Figure 7. All the above sentences are 100% correct, but some of them might be regarded as the right nonsense, and won't be adopted in practice due to their low precision in locating the target entity.

relations, the relations in the upstream of the data chain is dominant, e.g., $\xrightarrow{inside} + \xrightarrow{top} = \xrightarrow{inside}$, $\xrightarrow{east} + \xrightarrow{inside} = \xrightarrow{east}$...

If there is only one edge in a data chain, we can encode it directly, such as the data chains L1 and L2. If there is more than one edge in a data chain, e.g., the data chains L3, L4, and L5, we can perform the reasoning principles to get the results below.

- L3: $\xrightarrow{front} + \xrightarrow{top} = \xrightarrow{upfront}$; L4: $\xrightarrow{inside} + \xrightarrow{inside} = \xrightarrow{inside}$; L5: $\xrightarrow{east} + \xrightarrow{inside} *3 = \xrightarrow{east}$.

Distance relations: In some cases, e.g.: 1) the spatial distance between the target entity and the helper entity is very close, or 2) it is not necessary to provide the exact position of the target entity, then we can use the spatial distance relations as an alternative, just like the sentence L3 in Table 6

You may argue that the sentences we generated are too simple. However, at the initial stage of language appearance, it is just some simple words and short sentences. With the development of human beings, more and more information is encoded in language, then sophisticated words and long sentences emerged. Therefore, it is a good start to launch our research with some simple words and sentences.

- ENCODING PRINCIPLES FOR PROCESSING REQUESTS

Sentences encode not only the data chain to be conveyed, but also the processing requests for that data chain. According to the implicit processing requests in the sentences, we divided sentences into following three categories: 1) data description sentence (i.e., declarative sentence), 2) data verification sentence (i.e., the yes-no question sentence), 3) data searching sentence (i.e., WH-question sentence).

Data description sentences implicit the processing request that listeners are expected to store the information (knowledge) in their databases. For example, teachers expect the students to remember what was taught in the class, and authors expect the readers to understand and remember the ideas shared in the book, and so on.

Figure 8: The data chain L6 and its three different cases.

Data Chain		Target Node			Helper Node	
L6		Duke University	is	in	North Carolina	.
L6-1	Is	Duke University		in	North Carolina	?
L6-2	Which state is	Duke University		in		?
L6-3		Which University	is	in	North Carolina	?

Table 7: Comparison of sentence structures that encode different information processing requests. (English only)

Data verification sentences implicit the processing request that listeners are expected to help verify whether *the spatial relation* described in the sentence is true; and feedback the verification result as the response. For example, in the case of the data chain L6-1 in Figure 8, speakers are not sure whether the spatial relation "inside" between the node (Duke University) and the node (North Carolina) exists? They could express the processing request that ask listeners to help verity whether the "inside" edge exists by moving the word "Is" to the beginning of the sentence and adding a question mark at the end of the sentence, as shown in Table 7.

Data searching sentences, in which listeners are expected to search for the missing information (knowledge) replaced by WH words in their databases and return the search result as the response. Take data chains L6-2 and L6-3 in Figure 8 as examples, speakers can use the word "which" to replace the missing parts and adjust the structure of the sentences, as shown in rows L6-2 and L6-3 in Table7, to express their expectation that the listener can help to search for the missing parts and return the search results.

2.3.2 Natural Language Understanding

The sentence understanding task consists of two parts: a) understanding of the **processing requests** implicit in a sentence; and b) understanding of the **specific knowledge** conveyed in the sentence, which includes all the requisite parts listed in Table 5. Whereas we only provide the model (TGHM) to describe the spatial relation between entities, here we only introduce the principles for understanding 3) *the spatial relation* and 4) *the existence judgment of the spatial relation*. The understanding of 1) *the target nodes* and 2) *the helper nodes* requires other databases, which are beyond the scope of this paper and will be introduced in other papers in the future.

- DECODING PRINCIPLES

Extract the processing requests: The specific processing requests are expressed by the specific sentence structures, specific feature words, and specific punctuation. These can be used to classify the sentences and extract the processing requests accordingly.

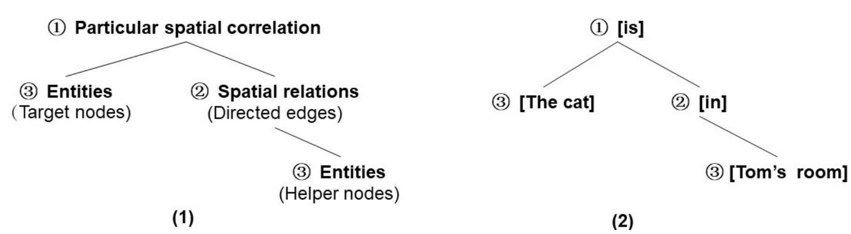


Figure 9: (1) General tree structure of sentences. (2) The sentence tree of sentence L1 in Table 6.

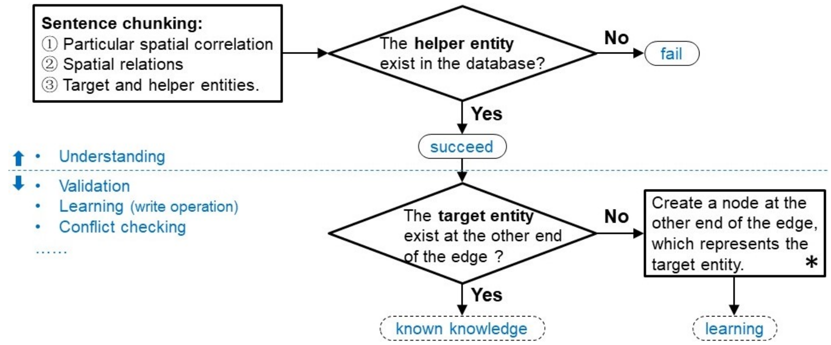


Figure 10: Processing flowchart of the data description sentences.

Sentence chunking: Listeners need to chunk the sentence and extract the requisite parts of the specific knowledge. Considering the difference in the number of words and phrases used to represent each class of the requisite parts, the most efficient way is to chunk the sentences following the order shown in the left part of Figure 9, and gain a sentence tree (see the example shown in the right part in Figure 9).

- UNDERSTANDING OF THE SPCIFIC KNOWLEDGE

Validation: In the process of understanding the specific knowledge, listeners need to verify each parts of the sentence tree in their TGHMs according to the flowchart shown in Figure 10. For example: in the case of the sentence tree in Figure 9, listeners should first verity whether the helper entity “Tom‘s room” exists in their TGHMs, if the helper entity exists, go ahead; if not, it means that the listeners cannot get the position of the target entity “The cat” through the helper entity “Tom‘s room”, so the understanding mission fails. If the helper entity “Tom‘s room” exists, the listeners can further verify whether the target entity “The cat” exists at the other end of the "inside" edge, if the target node “The cat” exists, it means that it is known knowledge to the listeners.

Learning: If the target node "The cat" not exists, the listeners can create a node at the other end of the “Inside” edge, to store the spatial position of “The cat” in their TGHMs, this is a learning process.

Conflict checking: Furthermore, suppose the specific knowledge described in the sentence conflicts with the knowledge (or data) stored in their TGHMs, a conflict check is required, which may create a new NLG task.

- RESPONDING TO THE PROCESSING REQUESTS

Strictly speaking, responding to the processing requests implicit in a sentence is not a sentence understanding task, but a sentence generation task. Here, we briefly introduce the principles for responding the different processing requests. **For a data description sentence**, as shown by the part below the dotted line in Figure 10, the response can be further divided into validation, learning, conflict checking, etc., as introduced in the previous section. **For a data verification sentence**, the response is to return the verification results to the speakers. Take the sentence L6-1 in Table 8 as an example, in the listener’s TGHM, if the edge represented by the word “in” can be found between the node “Duke University” and node “North Carolina”, the listener can reply “Yes, it is” as feedback to the speaker. If not, the listener can reply “No, it is not” as the feedback. **For a data searching sentence**, the response is to return the searching results to the speakers. Take the sentence L6-2 in Table 8 as an example, in the listener’s TGHM, if there is a node at the other end of the edge (represented by the word “in”), the search mission succeeded, and the listener can give the speaker the lexical representation of that node. If not, the listener can reply “I don‘t know” or “I don‘t have a clue” to the speaker, to let him or her know that the search mission failed.

2.4 Conceptual Interpretation

In this paper, we briefly introduce the new method mainly at the conceptual and practical levels. For the new concepts in the database-based method, we first give a rough interpretation at the theoretical level for a better understanding. For example: 1) the TGHM, data chains, and the processing requests implicit in sentences are implicit knowledge; 2) the TGHM can also be explained as a semantic representation. 3) In the TGHM, the structure consists of computable edges can also be seen as the reasoning path, which giving our method the algebraic capacity to understand and generate a potentially infinite number of novel combinations from known components Malkus et al.. 4) the reasoning process is also a practice of systematic generalization Baroni (2017). 5) Since, we have reclassified sentences according to their implicit processing requests, the previous classification of NLP tasks, e.g., dialogue and question answering, will be replaced. 6) The arrangement of the requisite parts in a sentence (see Table 6) and the sentence structures for encoding different data processing requests (see Table 7) are called syntax.

3 Effectiveness and Related Work

The best thing about the database-based NLP method is that its effectiveness is innate. Databases are the cornerstone of the whole method, which factually describe how real-world information (knowledge) is organized in the human brains. Furthermore, we can think of the database as the axiom set, the NLG task is to derive propositions (i.e., sentences) from the axiom set; the NLU task is to verify sentences (propositions) with the axiom set, in which involves the validation and conflict checking processes. In the NLU process, if a given proposition is known to be true, the new information (knowledge) brought by the proposition can be written into the database, further expanding the database (axiom set), which is a learning process. Translation tasks can also be performed using the database, the encoding table (see Figure 5), and corresponding encoding principles. The NLG, NLU, learning and translation processes summarized above simulate how the information (knowledge) is processed in the human brain.

Somewhat surprisingly, the edges in TGHM and the computability of these edges correspond to the characteristics of an algebraic structure. We are also working with mathematicians to give a theoretical argument for the TGHM from the perspective of abstract algebra, in parallel with refining and standardizing the relevant operations.

4 Conclusion

In this work, we provided a new framework for solving NLP problems, and discussed its potential in other AI problems (e.g., learning, translation). So, what exactly can we learn from the study of language? As we have learned in neuroscience, humans receive information through neural pathways such as eyes, ears, mouth, nose, etc., and then send this received information to the brain for hierarchical processing and storage. Although we cannot directly observe how this information is processed and stored in human brains, but, a small proportion of the information is encoded as natural language for external output. Thus, we can take natural language as a window to explore how the information (knowledge) is stored and processed in the human brain, which will lead to a brand new direction in AI research.

References

Brenden Lake Marco Baroni. genera1ization without systematicity on the compositiona1 ski11s of sequence to sequence recurrent networks. 2017.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

D. S. Malkus, M. E. Plesha, and R. J. Witt: Concepts. [18] chomsky, n.: Syntactic structures. mouton, the hague, 1957 (reprint berlin and new york, 1985).

Michael A. Paradiso Mark F. Bear, Barry W. Connors. *Neuroscience: Exploring the Brain, 2nd ed.* Higher Education Press, Beijing, 2004.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding with unsupervised learning. 2018.

Schölkopf and Bernhard. Artificial intelligence: Learning to see and act. *Nature*, 518(7540):486–487, 2015.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.