WORD EMBEDDING AND POSITIONAL ENCODING: CROSS – LINGUISTIC ANALYSIS

CHULUUNDORJ BEGZ

University of the Humanities, Mongolia
E-mail: chukabegz@humanities.mn

Abstract - Positional encodings are techniques used in neural networks to provide information for understanding the structure of sequence in a text and to handle a data more effectively. Positional encodings added to the input embedding's present effective techniques to develop text extraction models. The integration of word embeddings and positional encodings has vital importance for developing text extraction techniques in Mongolian language.

Keywords - Positional Encoding, Embedding Layer, Syntax Constructions, Associative Network, Attention Window, Graph, Vector Model, Sinusoidal Function

I. INTRODUCTION

Relation between morphology and word order in different languages presents the specific object of analysis in computational linguistics.

Positional encoding vary across languages with different word orders. Effect of word order in English and Mongolian languages illustrates the diversity of morphological systems leveraging the varying importance of positional encoding across languages.

Original word order of a sentence in attention mechanism serves as a basis to process information sequentially with the nuances and dependencies between words. Relative distances between word depending on sequence length in different languages illustrate a difference in the cohesion between words.

Vector based positional embedding presents effective techniques to analyze a sequence order through linear transformations. The even and odd positions in the sequence can be represented as a series of sine and cosine functions.

Linear transformation serves as effective model to encode relative distances in word embedding. Graph and embedding matrix allow to compute dependencies between different words in the text. Embedding can bridge the gap between typologically different languages.

II. COMPARATIVE ANALYSIS: THE SYNTAX AND WORD ORDER

Encoding and embedding are fundamental elements of NLP and LLM. Developing encoding technique and vector - based embedding in typologically diffenent languages plays critical roles in offering promising applications of AI in varions fields.

Basic elements of human verbal reasoning, order of mental operations employed in verbal structures are the object of analysis in multiple dimensions. Verbal structures embedded in human mind, operations with these structures present methodological significance for analysis of human cognitive mechanism.

The order of mental operations has importance for description of basic rules and constraints for combining structural and semantic features into words and combining words into a potentially infinite number of phrases and sentences. (Tai Linzen, T.Florian Jaeger. Cognitive science. Vol.40.2016. Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions.) Relation between morphology and word order in different languages presents the specific object of analysis in computational linguistics.

Some researchers express an idea about representation of human brain in the framework of Euclidean distance. The central tenet of this idea is that brain is a tensorial system and central nervous system acts as a metric tensor which determines the relationships between contravariant and covariant vectors. This idea leads to space-time representation of human cognition as a quantum model of cognition. (Sisir Roy, 2004. Quantum processes and functional geometry: New perspectives in brain dynamics.)

According to the researchers, metric space (Euclidean, Minkowski, Hausdorff, Hilbert etc.) is important starting point for interpretation of psycho-cognitive operations as operations of mental grammar, rules of verbal thinking.

Word order has been used in distributional vector models (Bringing order to neural word embeddings with embeddings augmented by random permutation. Trevor Cohen, Dominic Widdows 2018. Procesdings on the 22 nd conference on NLL.)

In high-dimensional vector space semantic relationships between words is point of analysis in word embedding.

Embedding as a nearest neighbor classification is based on similarity measures between objects.

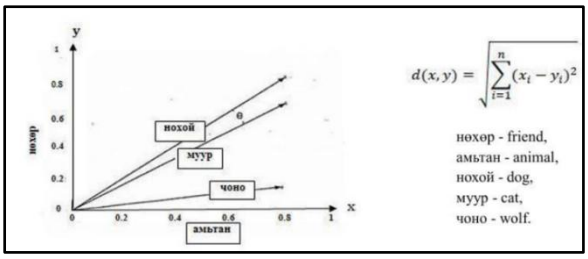


Fig 1. Cosine similarity between the vectors

Distance between words in the sentence reflects ratios of co-occurrence probabilities to encode meaning components in word vector space. Embedding based similarity metrics reliable in comparative analyses of human judgements in typologically different languages.

Common semantic and syntactic properties serve as a basis of associative network of words in human mental vocabulary. Classification of verb and noun classes as an expression of dynamic and static properties, presents product of mental operations similar to factorization. Factorization in human mind must be based on denotative, objective and also connotative, subjective features. For example, reference of the word "нохой-dog" is associated with the class of home animals and also has connotative link with "human friend".

Differences in human perception of an object and its properties are reflected in a sequence of deep and surface structures in verbal cognition. Intrinsic dimensions are reflected in a sequence of attributive components in relation to noun (object or subject). Some languages describe different color, shape, and motion. Gradable adjectives such as "big" tend to highlight dimensions of relations between objects as well as intrinsic dimensions. (Catriona Silvey, Simon Kirby Kenny Smith, Word meanings evolve to selectively preserve distinctions on salient dimensions, 2015.)

Association-based semantic models must be effectively applied to analysis of differences in word embeddings in typologically different languages. In that way analysis of human mental associative mechanism must be used for developing positional embedding and encoding techniques.
Relationship between word order and positional encoding offers a way to encode morphology and syntax features in multilingual semantic space.
Semantic similarity of keywords or terms related to the same node and logico-semantic coherence between keywords are based on different associations in human cognition.
In the sentence "Хүүшатартоглов" (The boy played) chess each word represented as a vector:

query – тоглов (played), key – хүү (boy), value – шатар (chess).
The query vector tells us what "тоглов" is seeking to understand: who or what is playing (тоглов<хүү, шатар). Key vector for "хүү" represents the information these tokens hold, value vector represents the actual content. We will focus on "тоглов" seeking information from "хүү" and "шатар". To determine how relevant "хүү" and "шатар" to "тоглов", we calculate the keys of the other tokens using the softmax function to convert these similarity scores into attention weights. These weights determine how much attention "тоглов" should pay to "хүү" and "шатар". Using these attention weights, we compute the weighted sum of the value vectors from "хүү" and "шатар".

Differences in language syntactic typology significantly influence the attention window and are reflected in positional encoding:

He examined the car.
Тэрмашинаашалгав.
He painted the wall blue.
Хүүширээгцагааналарбудав.

These embeddings are processed by attention layers with respect to typological differences in the languages.
The differences refer to neural oscillations implicated in human herbal cognition enabling different types of binding. Oscillatory signals can be represented as a weighted sum of sine and cosine waves. Some differences must be interpreted in the frame work of attention transformation.
Embedding matrix with rows representing tokens and columns corresponding to latent features provides more power to describe a difference in semantic similarities of the syntax structures in typologically different languages.
Positional encoding as a sinusoidal encoding in combination with positional embedding captures both similarities and relationships in syntactic structures.
The wave or signal coming from the vertical coordinate, or the a signal from horizontal coordinate must be used to modeling syntax structures in typologically different languages in relation to attention structure. (Vaibhav Sharma, Attention: A mathematical walkthrough 2025.) Decomposing the oscillatory signal into a bunch of sinusoidal signals presents effective technique for vector embedding.
Paradigmatic and syntagmatic relations between words evoke different associations in human mental space (smog-smoke, fog, emission; smog-health, cancer). The differences in association present interest for analysis in terms of covariant and contravariant transformations and multimodal embedding.

Original word order of a sentence (SVO and SOV) in attention mechanism serves as a basis to process information sequentially with the nuances and dependencies between words. Relative distances between words depending on sequence length in different languages illustrate a difference in the cohesion between words.

Languages differ in word order variation in preference to topicalization hierarchies (Michael Hahn. Yang Xn. Crosslinguistic word order variation reflects evolutionary pressures on dependency and information locality. 2022) Positional encoding vary across languages with different word orders. Vector based positional encoding presents effective techniques to analyze a sequence order through linear transformations. The even and odd positions in the sequence can be represented as a series of sine and cosine functions. Effect of word order in English and Mongolian languages illustrate the diversity of morphological systems leveraging the varying importance of positional encoding across languages.

Vector-based analysis of sequence regularities in above named structures proposes that implicit statistical knowledge in working memory brings to light the relevance of intrinsic and extrinsic features of an object to verbal cognition.

Giving	Багш хүүд ном өгөв.
Taking	Хүү багшаас ном авав.
Throwing	The girl threw the ball to the boy.
Catching	The girl caught the ball from the boy.

Difference in distance between the components in syntax constructions SVO and SOV is reflected in vector representation of words in the embedding space. In Mongolian language direct object animacy and causative verbs are more relevant for contextual embedding.

Би нэг ном авсан.

I bought a book

Word Embedding v_0 v_1 v_2 v_3

Positional encoding matrix. p_0 p_1 p_2 p_3

Би	0	p_{00}	p_{01}	p_{02}	p_{03}
нэг	1	p_{10}	p_{11}	p_{12}	p_{13}
ном	2	p_{20}	p_{21}	p_{22}	p_{23}
авсан	3	p_{30}	p_{31}	p_{32}	p_{33}

Table 1. Relative positioning

Difference of the position information is reflected in attention window. Perception of a sentence depends on the attention transformer parts of sentence directly interacting across distant positions. In that's way sine and cosine functions play important role in modeling syntax structures as a sentence where words or tokens interact across distant positions.

Above named structures with a differences in positional encoding present an object for modeling with sinusoidal function. There is difference in a way to incorporate position information. linguistic coding of motion events, static locations of objects must be packaged differently in human memory. Fourier method is applied to decompose the oscillatory signal into a bunch of sinusoidal signals.

Simple way to associate token and positional embeddings is based on co-occurrence matrix in linear meaning. Some complex sentences also present interest in terms of variety of attention influence:

I told him that he was wrong.
I don't know whether it will rain or be sunny.
Батхэзээирэхийгбимэдэхгүйбайна.
Өчигдөрзавгүйбайсангэдгийгбитүүндхэлсэн.

Tensor models present effective way to interpret, complex sentence as a object-extracted relative clauses, subject-extracted relative clauses, left-branching and right-branching constructions.

Хүүзамбруузаасаналдаагаааойлгожээ.
Замбруузаасаналдаагаахүүойлгожээ.

Syntax structures related to prototype-plus-distortion phenomenon (central prototype and radial prototype categories) present special case for embedding in human mental space depending on typological differences of the language.

Эгчталхзүсэв. The sister cuts the bread.

Radial Categories	**Force**=voluntary agent: **Цас хаалга даржээ** - **Салхинд хаалга хаагджээ.** The wind closed the door.
	Instrument: Чулуу цонх хагалжээ. The hammer broke the window.
	Neutral: Дэмбээ ууланд авирав. Peter climbed the mountain.

There is an increased interest how typologically different languages encode morphology and syntax features across different layers. Language specific features as subject-verb agreement, flexibility of word order, morphological type require more parameters to represent. Cross-lingual abstractions of morphosyntactic concepts serve as a basis for disentangling latent grammatical concepts across typologically diverse languages.

III. TEXT SEGMENTATION WITH EMBEDDINGS

Embeddings have become the effective technique to convert text data into vectors. Methodology for conceptualizing the embedding space must be effectively applied to text analysis in Mongolian

language. (Interpreting embedding spaces by conceptualization, Adi Simhi, ShaulMarkoitch 2023).

In Mongolia text on the air pollution word vector for “Smog” is the same for the dimensions “Smog” and “health”. In the graph the vectors for associated words in each pair are measured for correlation.

Утаа

Утаатмананньажүйлдвэржсэнбүснутгаарнийтлэгб өгөөдэнэньхотуудынхувьдтанилхарагдацбайсаарб айна. Утаатмананихэвчлэннүүрстүлснээсүүсдэг. Оронгэрээхалаахболонхоолхийхзорилгоорнүүрстү лэхньхортхийнялгарлыгбийболгожбайна. Утаатмананньутааболонманангийннэгдэлюм. Утаатмананньагаарынбохирдлынаюултайтөрөлгэ жүзэгддэг. Мананхүмүүстижилтөрлийнамьсгалынасуудлыгб ийболгодогбайхадутааньуушгийгхортхавдарүүсгэ гчээрдүүргэжбайгаагаарааилүүаюултайюм. Уушгиныхортхавдарньнүүрсийгахуйнхэрэглээний зорилгоортүлэхтэйбасхолбоотойюм. Нүүрсньдэлхийнийтээрхэрэглэжбуйнүүрстөрөгчи йнхамгийнтомэхсурвалжбөгөөдбидний эрүүлмэнд эдзонхилохаюулзаналыгучруулжбуйгэжүздэг.

Smog

Smog is common in industrial areas, and remains a familiar sight in cities. The smog usually came s from burning coal. Burning coal inside the home for the purposes of heating and cooking produces gas emissions.

Smog is combination of fog and smoke. Smog refers to a dangerous type of pollution. While fog can cause similar breathing problems in people, smog is more dangerous as it fills the lungs with cancer causing carcinogens. Lung cancer also is associated with household combustion of coal.

Coal remains the world’s single largest source of carbon pollution, presenting a major threat to our health.Graph of the text:

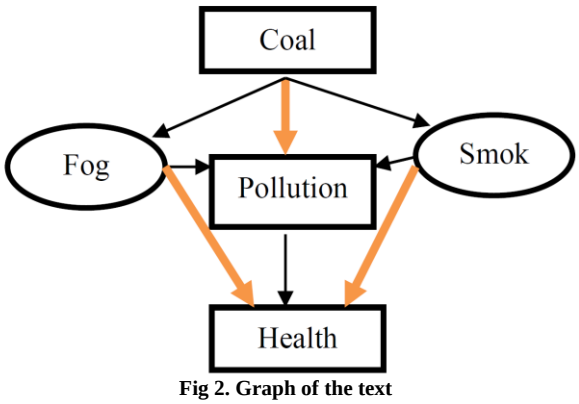
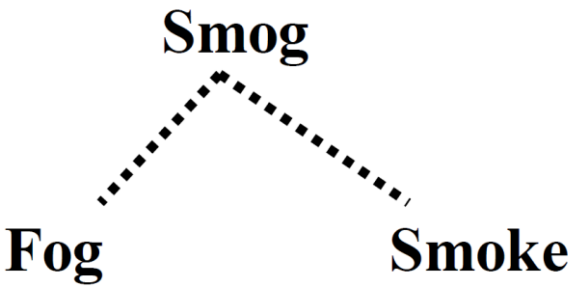


Fig 2. Graph of the text

Embedding layer maps each word to a vector. Words with similar meanings are positioned closer to each other:

Sources of pollution must be represented by a graph:



Embedding layer also allows other words “health, cancer an lung” into vectors. Distance between any two words can be interpreted as a measure of relative similarity leading to contextual embeddings within the text. Adding third dimension as innovative energy source must extend embedding space. Graph embedding represents effective technique to concider the semantic relationships between vertices. Relations between two or more concepts must be represented in terms of fuzzy sets.

Embedding based weight matrices allows to compute dependencies between different words in a sequence in a text.

Model based on vector multiplication precents effective technique to embedding nonlinear verbal structures where semantic transformations in combination with pragmatic meaning must be described in terms of cross product. This model serves as a basis to develop the transformer-based cross-lingual embedding models to process text in different languages.

IV. CONCLUDING REMARKS

The interaction between language and thought motivates new approaches to analysis of embedding and coding the manner of an action. Vector based model in combination with applying cosine and sine functions to encoding positions presents an effective technique to analyze SVO and SOV structures reflecting the differences in human verbal cognition. Contextualized word embeddings serve as a technique for representing underlying structure of the data beside the differences in the typology of languages.

Applying the positional embedding to modeling different ways of encoding complex syntax structures is important to alternate between cosine and sinusoidal functions, to distinguish between odd and even indices of word embedding dimension. Sinusoidal function is more applicable to positional embedding with multiple dimensions. Sensitivity to word order in typologically different languages with fixed or free word order is associated with encoding

the word order information in different layers. This idea must be applied to analysis other complex syntax structures in typologically different languages.
Embeddings generated from different languages play an important role in developing multilingual transformer models.

REFERENCES

[1] Adi Simhi, ShaulMarkovitch. (2013). Interpreting embedding spaces by conceptualization.

[2] Arthur C. Graesser, Danielle S.McNamara. (2011). Computational analysis of multilevel discourse comprehension. Topics in cognitive science, l3(2), p386.

[3] Catriona Silvey, Simon Kirby, Kenny Smith. (2015). Word meanings evolve to selectivelypreserve distinctions on salient dimensions.

[4] Michael Hahn. Yang Xu. (2022). Crosslinguistic word order variation reflects evolutionarypressures on dependency and information locality.

[5] Patrizio. E, Tressoldi, Lance Storm, Dean Radin. (2010). Extrasensory perception andquantum model of cognition. NeuroQuantoloty, 8(4), p585

[6] Peter D. Bruza., Zheng Wang., Jerome R, Busymeyer. (2015). Quantum cognition: a new theoretical approach to psychology. Trends in Cognitive Sciences, 18

[7] Shuiwang, J., YaochenXie., Hongyang Gao. (2020). A mathematical view of attentionmodels in deep learning. A&M University, Texas

[8] Roy, S., &Kafatos, M. (2004). Quantum processes and functional geometry: New perspectives in brain dynamics. Forma, 19, 69–84.

[9] Tai linzen. TFLorian Jaeger. (2016). Uncertainty and Expectation in Sentence Processing:Evidence from Subcategorization Distributions. Cognitive seience, 40(6), p1386.

[10] Trevor Cohan. Dominic Windows. (2018). Bringing order to neural word embeddings with embeddings augmented by random permutation. Procedings on the 22nd conference onNLL

[11] Vaibhav Sharma. (2015). Attention: A mathematical walkthrough

[12] Vaibhav Sharma. (2025). Scaling attention on transformers: sliding window & chunkedattention. p. 85

[13] Zuccon, G., Azzopardi, L.A., Rijsbergen, C.J. (2009). Semantic space: Measuring thedistance between different subspaces. University of Glasgow, p.230

★★★