Word Vectors for Criticism of a Korean Film - 'Decision to Leave' by Chanwook Park

Kwangho Ko and Juryon Paik *

Posted Date: 24 November 2023

doi: 10.20944/preprints202311.1558.v1

Keywords: Analogy operation; chanwook park; 'decision to leave'; film criticism; korean film; language model; LSTM; similarity operation; word vector

Article

Word Vectors for Criticism of a Korean Film - 'Decision to Leave' by Chanwook Park

Kwangho Ko [1] and Juryon Paik [2,*]

[1] Dept. of Smart Mobility, Pyeongtaek University, S. Korea; kwangho@ptu.ac.kr
[2] Dept. of Data Information and Statistics, Pyeongtaek University, S. Korea; jrpaik@ptu.ac.kr
* Correspondence: jrpaik@ptu.ac.kr

Featured Application: Sentiment Analysis, Topic Extraction, Document Classification, Search and Retrieval.

Abstract: This research presents a novel approach to film analysis, leveraging word vectors to objectively evaluate movie critiques. Focusing on Director Chanwook Park's award-winning film, 'Decision to Leave', the study employs word vectors derived from the movie's script of Korean text. Traditional critiques often emphasize contrasting elements and themes, but their subjective nature poses challenges in objective validation. To address this, we trained a language model using LSTM on the film's script, obtaining word vectors that capture the essence of the narrative. These vectors were then used to perform various text analyses, including similarity and analogy operation for the keywords suggested by the critiques. By comparing the semantic relationships in the critiques with those derived from the word vectors, we could objectively validate the critiques' assertions. Furthermore, we visualized the word vectors in a two-dimensional space, confirming the spatial relationships of key terms highlighted in critiques. The study underscores the potential of word vectors in providing a more objective lens for film analysis, bridging the gap between traditional film criticism and data-driven insights.

Keywords: Analogy operation; chanwook park; 'decision to leave'; film criticism; korean film; language model; LSTM; similarity operation; word vector

1. Introduction

Language models are constructed by learning the repetitive sequences of words in the training corpus. During this learning process, an embedding layer is utilized to transform input words into vectors. These word vectors sequences are then passed through recurrent neural networks and fully connected layers to optimize the parameters necessary for selecting appropriate words following the input sequence. Therefore, a language model trained on a given corpus can be regarded as a system capable of receiving a specific length of sequence as input and predicting the suitable word to follow the sequence based on the learned context. Furthermore, the word vectors represent the characteristics of the vocabulary in the training corpus as numbers corresponding to the dimensionality of the vectors.

Using such language models or word vectors, one can analyze text or documents similar to the training corpus. For instance, if movie review texts are used as the training corpus, you can perform sentiment analysis on review texts similar to the training corpus. Additionally, you can extract the main topics or themes or select key keywords. Through sentiment analysis, topic extraction, and keyword selection, you can classify various documents based on specific criteria and quickly search for desired documents. Most previous researches primarily utilize word vectors for these practical purposes as follows:

Sentiment Analysis: Language models, like the one used here, can be fine-tuned for sentiment analysis. For example, by training the model on movie review texts, it learns to recognize the sentiment expressed in text — whether it's positive, negative, or neutral. This is a valuable application for understanding public opinion, customer feedback, or any text data with sentiment [1–3].

 doi:10.20944/preprints202311.1558.v1

Topic Extraction: Language models and word vectors can help identify the main topics or themes within a large corpus of text. By analyzing the relationships between words and phrases, these tools can automatically group similar content. This is particularly useful for organizing and summarizing large volumes of textual data, such as news articles or academic papers [4–6].

Keyword Selection: Identifying keywords or key phrases is another essential task in text analysis. Language models can assist in this by highlighting words that frequently appear in a specific context or have a strong association with a particular topic. These keywords can be used for indexing, search engine optimization, or summarization [7–9].

Document Classification: Through sentiment analysis, topic extraction, and keyword selection, documents can be classified based on specific criteria. For instance, you can categorize customer reviews as positive or negative, news articles by subject, or research papers by field. This classification helps in organizing and retrieving documents efficiently [10,11].

Search and Retrieval: Once documents are classified and keywords are identified, you can quickly search for and retrieve specific documents that meet your criteria. This is valuable in applications like information retrieval, content recommendation, and data mining [12–14].

In practice, these techniques are widely used for various purposes, such as understanding customer sentiments in product reviews, organizing news articles into sections, recommending content based on user preferences, and more. By leveraging language models and word vectors, organizations can gain valuable insights from their textual data, leading to better decision-making and improved user experiences.

In this study, we propose a technique that utilizes word vectors obtained during the language model training process using movie scripts as a corpus to objectively evaluate movie critiques. The purpose of movie critiques is to analyze films, make recommendations to consumers, and provoke deeper insights among the audience, often focusing on aspects like the narrative development, selection and placement of props or settings, and character traits.

All the elements intensely analyzed in movie critiques constitute the components of the film text. In cases where movie scripts are employed as the corpus for language model training, we can analyze the targets of these critiques using the word vectors that can be acquired. Here's how this analysis can be performed specifically:

Keyword Selection: Begin by selecting key terms from the critiques. Then, calculate the similarity between these keywords and words in the corpus using vector similarity operations.

Semantic Analysis: Evaluate whether the meanings and symbols attributed to the selected keywords in the critiques align with the characteristics of the words identified through similarity calculations.

Relation Analysis: Define relationships between these keywords as presented in the critiques and employ vector operations to identify words that align with these relationships.

Visualization: Project the word vectors onto a two-dimensional plane and assess whether closely related keywords in the critiques are indeed proximate in the plane, verifying if the direction aligns with the critiques' claims.

To achieve these objectives, we conducted an analysis of the Korean script of 'Decision to Leave', the 11th film by director Chanwook Park, famous for the Korean movie 'Oldboy'. We analyzed various critiques of 'Decision to Leave' to understand the meanings and symbols associated with key terms. We trained a language model with LSTM(Long Short Term Memory) scheme, using the movie script text as the corpus. During the training process, we utilized the acquired word vectors to perform similarity and inference operations related to the selected keywords. Additionally, we reduced the dimensionality of the word vectors and visualized the positions and directions of the keywords on a plane, comparing them with the critiques' assertions. The analysis focused on assessing whether the characteristics of the words align with the critiques' claims.

In the following chapter, 'Related works', we conducted an analysis of critiques related to 'Decision to Leave'. Subsequently, in the chapter 'Result and Analysis', we compared and analyzed the results of

 doi:10.20944/preprints202311.1558.v1

similarity and inference operations concerning the main keywords in the critiques. We also verified the positions and directions of the keywords visualized on the plane. Through this comparative analysis, we proposed the utility of word vectors for movie critiques.

2. Related Works

2.1. The Movie - Story

The film 'Decision to Leave', directed by Chanwook Park, won the Best Director award at the 75th Cannes Film Festival in 2022. 'Decision to Leave' deeply explores the themes of suspicion and love between a female suspect and a male detective. The female lead character is portrayed as someone who immigrated from China and is suspected of assisting in her mother's euthanasia, which raises doubts about her innocence. On the other hand, the detective suffers from insomnia but is renowned for his meticulous investigative skills. The film weaves a suspenseful and intimate love story between these two characters as their journey unfolds from the mountain to the sea, spanning two different settings: the urban city of Pusan and the rural seaside village of Leepho.

In the film, the story unfolds as follows:

In the city of Pusan, often referred to as Korea's second city, Detective Jang investigates the case of a man named Kee who fell to his death at a place called Guso Mountain's Bigum Rock. During his investigation, Detective Jang summons Song, Kee's Chinese wife, who is not fluent in Korean, to the police station for questioning. Detective Jang inexplicably develops a strange fondness for Song. He orders premium sushi lunch boxes from Shimasushi to eat while continuing his investigation.

As the investigation progresses, it is revealed that Kee had subjected his wife, Song, to violence, including leaving his initials as tattoos on her private parts. This revelation leads to suspicions that Song might be the murderer of her husband.

Detective Jang continues to monitor Song, who works as a helper for elderly grandmothers in the neighborhood. He observes her visiting various elderly people on different days, taking care of them. He notices her solitary evening meals, where she enjoys ice cream, practices the Korean language by watching historical dramas on TV, and gradually becomes more and more fascinated by her.

As the story unfolds, an anonymous letter surfaces, implicating Kee in a bribery case. It makes the case closed as a suicide, and Song is no longer considered a suspect. Instead, Detective Jang's relationship with Song deepens. Jang cook a hot stir-fried rice dish for Song, and she, in return, teaches him breathing techniques and guided imagery to help him cope with his insomnia, eventually helping him fall asleep.

On a rainy morning, Song and Jang take a stroll near a temple. Their mutual fondness for each other continues to grow. However, when Jang visits an elderly lady's house on a Monday instead of Song, who needed to attend a funeral, he stumbles upon a health app on her smartphone that records stair-climbing data. This data shows that she had climbed 138 flights of stairs. This discovery prompts Jang to reevaluate the case.

In a pivotal scene, it is revealed that Song, in fact, disguised herself as if she were visiting the elderly lady's house on a Monday. She had secretly climbed to the summit of the mountain, where her husband, Kee, was located. Song ambushed him and pushed him off the cliff, leading to his fatal fall. Detective Jang later realizes the truth but decides to keep the case closed.

A powerful dialogue occurs during this scene, (all dialogue is quoted from the movie script) :

(Song, to Jang, who is questioning the incident at the cliff)

(Song) Please don't talk about our situation like that.

(Jang) Our situation, what is it? Is it me lingering near your house every night? Listening to your every breath as you sleep soundly? Whispering to you, holding you close, making you happy?

That's dignity, right? Do you know where dignity comes from? It's my pride. I was a proud

 doi:10.20944/preprints202311.1558.v1

police officer. But I ruined the investigation because of my obsession with you. I've completely collapsed.
I gave your grandmother a new phone, the same model, and she has no idea. Throw the old phone into the sea.
Sink it deep so no one can find it.
(After Jang leaves, Song enters 'collapse' into her smartphone dictionary. The definition reads 'fall apart and break.' The scene fades out with Song's expression reflecting her own collapse.)

In a later part of the story, Detective Jang, now living in the quiet coastal village of Leepho with his wife Ann, encounters the remarried Song couple by chance. A few days later, the body of Song's husband, Lim, is discovered, and Jang becomes suspicious of Song as the possible murderer. However, it is revealed that Sha, a scam artist who lost all of his assets by Lim, is the actual culprit.

But as the investigation unfolds, it becomes evident that Song was involved in euthanizing Sha's mother. Jang realizes that Song had set up a scheme to implicate Lim in the murder in order to protect Jang. In a phone conversation between Jang and Song:

(Referring to an audio file Song stored)
(Jang) You have that file too, don't you? Tell me, what's in it?
(Song) Your voice, confessing your love to me.
(Jang) Me?
(Song) I was so happy about it. But Lim found out.
(Jang) When did I ever say I love you?
(Song, in Chinese) The moment you said you loved me, your love ended, and my love began.
Get rid of Lim's phone. Throw it into the deep sea.
(Jang) When did I ever tell you I love you?
(The call ends, and Jang plays the recorded audio file from Song's smartwatch)
(Song, only voice) Please don't talk about our situation like that.
(Jang, voice) Our situation, what is it? Is it me lingering near your house every night? Listening to your every breath as you sleep soundly? Whispering to you, holding you close, making you happy? I've completely collapsed. Throw the old phone into the sea. Sink it deep so no one can find it.

In this scene Detective Jang, realizing that these words were actually a confession of love, hurriedly searches for Song. However, he arrives at the beach only to find Song digging a hole, soon to be engulfed by the encroaching tide. The scene is set against the backdrop of an old Korean pop song, 'Fog', which was popular in the 1970s. The lyrics of the song go like this:

Walking alone on this foggy, densely shrouded street
Somewhere, there's only one shadow of your once tender figure
When I think, there's nothing but passing memories
Yet my heart still aches with longing
Where could that person have gone?
Lonely, endlessly, in the midst of the fog
When I turn around, a faint voice blocks my way

 doi:10.20944/preprints202311.1558.v1

Oh, wind, walk through this fog

Where could that person have gone?

Open your eyes in the midst of the fog, hide your tears

The famous Korean folk song 'Fog,' heard from the front sea of the coastal village of Leepho, known for its fog, poignantly illustrates the image of a woman endlessly walking in search of the 'shadow of her once tender figure,' a central motif of the film. However, the woman in 'Fog' is also an active agent, asking the wind to clear the fog and commanding herself to open her eyes and hide her tears.

This active female subject is symbolically embodied in the character of Song in the film [15]. As the detective collapses, she interprets his desperate words, initially meant to discard evidence, as a 'love confession.' To become the eternal unresolved case of his love after their relationship collapses, Song voluntarily immerses herself in a pit on the beach to die.

2.2. The Movies - Critics

The film 'Decision to Leave' weaves a complex narrative that begins with a murder in the mountain and ends in suicide by the sea. It explores how suspicion can transform into love, how betrayed love can trigger another, all through a rich tapestry of imagery and symbolism.

The name 'Pu-San' where the crime in the mountain occurs is derived from Korean phonetics, with 'San' meaning 'Mountain.' Similarly, the coastal village of 'Lee-Pho' gets its name from 'Pho,' which means 'Port' in Korean. Thus, the film unfolds as a story that transitions its backdrop from 'Mountain' Pu to 'Port' Lee [16].

During their visit to Mountain Homi to scatter Song's mother's ashes, Song hands over Monday Grandma's phone, which is evidence from the case in Pusan, and says:

(Song) Use this to reopen the case. Go back before the 'collapse.' It seems like I came to Leepho because I want to become your unresolved case. Put my picture on the wall, to make you awake, and only think about me.

In the scene at Mountain Homi, where the circumstances surrounding Sha's mother's euthanasia and the murder plot involving Lim are revealed, Song confesses her desire to become Jang's unresolved case. This moment signifies her willingness to become the eternal mystery in Jang's life, even if it means an obsession that keeps him up at night. She fulfills this desire by taking her own life by the sea in Leepho. Therefore, this movie can be seen as a story that transitions from a murder in the mountain, leading to suspicion and infatuation, and ultimately ending with a collapse of emotions as Song takes her life by the sea.

The imagery of mountain-sea, fall-murder, and pit-suicide is creatively varied throughout the movie. Song's apartment interior wallpaper features an overlapping image of mountains and the sea (see Figure 1), and the dress she wears during her encounters with her second husband Lim and in Leepho is turquoise that's consisted of blue and green (see Figure 2). In Korea, mountains are often represented as green and the sea as blue in color. Therefore, Song incorporates this complex mountain-sea imagery into her home and clothing choices [17].

 doi:10.20944/preprints202311.1558.v1

Figure 1. Song's apartment interior wallpaper.

Figure 2. Song's dress in Leepho.

The image of being submerged in the sea in a pit as the waves roll in is also expressed through Jang. In a scene where Song, while trying to help Jang, who suffers from insomnia, get some sleep, she says the following:

(Song) Close your eyes.

(Song gently covers Jang's eyes with her hand, and Jang smiles.)

Listen to my breathing, synchronize your breath with mine.

(Song leans closer to Jang, both closing their eyes.)

(Jang listens to Song's breathing. At some point, their breathing rhythms sync perfectly.)

(Song) Let's go to the sea. Dive into the water. Descend. Keep descending.

You're a jellyfish. No eyes, no nose, no thoughts.

(Song, in Chinese) Neither happy nor sad. No emotions.

Push away everything that happened today in the water, to me.

I'll take it all, and now you have nothing left.

(Song opens her eyes. Jang's face is right in front of her, and he's already asleep.)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2023 doi:10.20944/preprints202311.1558.v1

The image of a jellyfish descending deep into the sea, synchronized with Song's breathing as Jang falls asleep, serves as a symbolic representation in the movie. In this scene, the sea and the jellyfish symbolize Jang's descent and entry into a state of sleep [18]. This imagery aligns with Song's journey from the mountain, falling, and then digging a pit before being submerged in the water and ultimately succumbing, reflecting the overall narrative of the movie where Jang and Song's story flows and descends like water from the mountain to the sea.

The overall flow of the movie is symbolically represented in the design of the movie script's cover and the movie poster. The cover of the script features a page from the 'Book of Mountain & Sea', which is being transcribed in a scene from the film (Figure 3). The 'Book of Mountain & Sea' is a mythical collection from ancient China that includes various natural landscapes and bizarre creatures that inhabit them [19]. The cover of the script shows a flowing stream of water descending like a waterfall from the mountain, and along this stream, Korean text is visible. Here is a partial quote from the content:

Figure 3. The design of the movie script's cover.

Once you go east for 250 li (a Chinese unit of distance), you will find Mount Gireum. The peaks of this mountain are deeply hidden, and to those who do not wish to see them, they remain invisible. Worms living here are as long as a hundred-year-old pine tree, and with something sticky comeing out from their bodies, it doesn't slip, allowing them to climb the mountain. Their white bodies, folded a thousand times, make it hard to distinguish front from back, but when they extend their long mouths, people decide which way to run. When the worm meets a person, it wraps itself around with its long body, pierces with its mouth, and sucks blood and bones, so one must avoid it at all costs. Once the worm falls and dies, a swarm of twenty thousand golden flies rises from its burst head and finally scatters into the world.

 doi:10.20944/preprints202311.1558.v1

If you travel three hundred li to the south, you'll find Homi Mountain. This mountain, when unseen by people, can walk but if someone notices it, it freezes in place, becoming an ordinary mountain. This mountain is exceptionally quiet, and you can even hear the sound of trees growing. Yet, if a person ventures into the midst of these trees, they vanish and never return. At the mountain's peak, there is a small rock covered in white clay. This rock, flat like a go stone, appears irregular from a distance. When adorned with cinnabar, it emits a brilliant light, piercing the eyes. There is also a peculiar tree known as 'Baeggu'. Its inner bark is crimson, and its appearance resembles rice or millet when it blossoms. The sap from this tree is as sweet as barley malt, akin to barley syrup.

'Mount Gireum' represents the location where the gruesome murder occurs, and it's associated with the name 'Kee' who has fallen and died. Similarly, 'Homi Mountain' is depicted as the mountain where Song and Jang scatter Song's mother's remains. These adaptations from the original text provide symbolic connections to the film's plot and characters.

The portrayal of the parasitic creature falling and giving rise to a swarm of flies is metaphorically linked to Kee's fall and the aftermath of his death, particularly the scene where his body is left to decompose. This demonstrates how the film creatively draws from traditional symbolism to enhance its storytelling.

The description of Homi Mountain in the film can indeed be interpreted as symbolizing Song's belatedly blossoming love. The quiet mountain with the sound of trees growing and the notion that one cannot return once they enter among these trees may metaphorically represent Song's love, which began late after Detective Jang's breakdown [20].

As the love between Song and Jang develops amid the serene backdrop of their relationship, it mirrors the idea of the sound of trees growing. However, as Song enters into this fully bloomed love, she disappears, signifying her departure through her act of suicide by the sea in Leepho.

The narrative of descending from the mountains to the sea and the love story that unfolds along the way is well-reflected in the film's posters, with two prominent examples shown in Figure 4. In the poster featuring Detective Jang, the background is a mountainous landscape, while in the poster featuring the Chinese wife, Song, the background is the sea. This effectively encapsulates the film's structure, where Detective Jang investigates a murder in the mountains and falls in love, and Song completes this love story by ending her life in the sea. These posters provide a visual representation of the film's thematic progression and characters [21].

Figure 4. The film's posters.

 doi:10.20944/preprints202311.1558.v1

The town of Leepho, famous for its fog, is depicted with various events and elements in the story. In Leepho, there is a nuclear power plant where Detective Jang's wife, Ann, works. Safety must be the top priority at a nuclear power plant. Detective Jang's wife's full name is 'Jeong-Ann Ann,' which, in the Korean naming convention where the family name comes first, can be transliterated as 'Ann Jeong Ann'. In Korean, the phonetic sound 'Ann Jeong' means 'Stable' in English. Both 'Ann Jeong' when read from left to right and when read from right to left represent 'Stable.' This highlights that Detective Jang's wife is a stable individual working in a place known for safety. However, one might also consider that nuclear power plants, as safe as they are designed to be, can carry a certain level of risk [22].

The dense fog that Detective Jang has to pass through every time he commutes from Pusan to Leepho can symbolize the safety and danger of the Leepho nuclear power plant, where his wife, Ann, works. While driving through the fog, Jang sometimes blinks and barely makes it to Leepho, highlighting the precarious nature of the journey. The fog can provide a comforting feeling of being warmly enveloped, akin to a sense of safety, but it is actually symbolic of danger. Moreover, since the region experiences heavy fog, the humidity is high in Leepho, making it known for mold issues. Jang's observation of mold growing in the corner of the wall while having an affair with Ann also effectively represents these complex attributes of the fog. In essence, the fog in the story carries a dual symbolism - it can represent a deceptive sense of security but also serves as a metaphor for the underlying dangers and complications.

The presence of the sea creature 'turtle' in the context of Detective Jang's insomnia is indeed an interesting symbol. In Leepho, turtle is bred by a grandmother and thieves stole the turtles. Detective Jang pursues criminals who stole the turtles. Detective Jang accidentally releases the turtles and as he attempts to catch them again, he ends up being bitten by the turtle. In the Korean language, "Ja Ra"(the pronunciation of 'turtle') means "Sleep" as a command. After his breakdown and relocation to the seemingly safe Leepho, Detective Jang still struggles with insomnia, and the command "Sleep!" remains ineffective for him. Naturally, he ends up physically being bitten by the turtle commanding 'Sleep!' [23].

In contrast, the jellyfish used to induce sleep for Jang by Song is portrayed as sinking deep into the ocean, without eyes or a nose, effectively guiding him into a deep slumber. Even though they are both sea creatures, turtle, subject to the constant command to sleep, bites the hands of Jang, while the jellyfish, like Song, gently drifts into a profound sleep. This comparison serves as a poignant representation of their respective experiences with sleep and the consequences thereof.

In another scene from the movie, there's a significant moment where Song is served fried rice instead of ice cream as dinner. This scene, with the contrast between the cold ice cream and the hot fried rice, is also considered a symbolic representation. When Detective Jang, who's preparing the shrimp fried rice, explains that it's a "only one" Chinese dish that he knows how to cook, Song reacts with a sense of bewilderment:

(Jang cooks shrimp fried rice in a wok as Song watches)

(Jang) This is a 'only one' Chinese dish that I can make.

(After a while, Song stands in front of the wok and takes a spoonful of fried rice)

(She chews thoughtfully, while Jang anxiously waits for her reaction)

(Song) You call this Chinese food?

(She says to a disappointed Jang) It tastes good.

(Jang, relieved, scoops the rice onto a plate)

In fact, Jang's fried rice is Korean, not Chinese. Song says it's delicious, but there's a Korean/Chinese difference between Jang and Song in the first place. The fried rice is a nice, warm substitute for

 doi:10.20944/preprints202311.1558.v1

cold ice cream, but it's not the Chinese food Song is used to eating. In this subtle way, the fried rice foreshadows the catastrophe of the relationship between Jang and Song, as it slides down the "suspicion-pity-betrayal-collapse-love-suicide" path.[24]

To summarize the various and in-depth symbols and images expressed in the above analysis of the film, we can summarize them as follows.

(1) Mountain/Sea : murder/suicide, doubt/love, investigation/compassion, man/woman, detective/suspect

(2) Pusan/Leepoh: mountain/sea, incident/safety, city/rural

(3) Green/Blue: mountain/sea

(4) Jellyfish/Turtle: sleeping/unsleeping

(5) Fog: dual image of safe and dangerous

(6) Negative images: maggots, ice cream, nuclear power plant, mountain, turtle

(7) Positive images: fried rice, sea, jellyfish

(8) Dramatic images: collapse, unresolved case, abandonment in the depths of the sea

Although we can summarize various images and symbols in the movie like this, it is uncertain whether the devices, places, and characters expressed in the movie actually have such features. Therefore, it is necessary to determine the authenticity by analyzing the nature of the words indicated by each place/device/person using word vectors that can be obtained using a recurrent neural network, a type of deep learning technique. When expressed as a word vector, it is possible to judge the degree to which each word is similar, and when the word vector is reduced to two dimensions and displayed, its location can be visualized.

3. Result and Analysis

3.1. Language Model - Structure

A popular technique for building language models using deep learning is Long Short Term Memory (LSTM). LSTM techniques learn the sequence of words in a list, creating a word vector for the vocabulary that makes up the training corpus. These word vectors can be used to find words that are highly similar to each other, or words that form similar relationships. By performing similarity and analogy operations on these word vectors, various text analyses are possible [25,26].

The composition of the deep learning model used in this study is shown in Table 1. The dimensionality of the word vector and LSTM hidden layers is the same, 2000, and the sequence length (time_size) is 10 100. Two LSTM layers were constructed between the embedding layer and the dense layer, and dropout (ratio = 0.6) was applied to prevent overfitting.

For the Korean script text of the film "Decision to Leave" used as the corpus, the number of Korean vocabulary words is 5,909 and the number of words in the corpus is 24,368. The language model is constructed by learning the order of the 5,909 vocabulary words to form the 24,368 words in the corpus. The ratio of the number of words in the corpus to the number of vocabulary is important, and in the case of "Decision to Leave", the ratio is about 4.12. To explain the ratio, we can use the analogy of Lego blocks: the fewer the types of blocks used (vocabulary) and the larger the structure created by stacking them (corpus), the easier it is to learn the order of application of the blocks [27]. Naturally, if a language model is created using only one movie script, the sentence generation performance is low, but it is reasonable for the purpose of obtaining word vectors and utilizing vector operations as in this study [28].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2023 doi:10.20944/preprints202311.1558.v1

Table 1. Structure of Language Model.

Layer Structure		
Embedding	(vocab_size, wordvec_size, inpu_length=time_size)	
LSTM	(hidden_size, dropout=0.6, return_sequences=True)	
LSTM	(hidden_size, dropout=0.6)	
Dense	(vocab_size, activation='softmax')	
Model Summary		
Layer	Output Shape	Parameters
Embedding	(None, 10 100, 2000)	11,818,000
LSTM_1	(None, 10 100, 2000)	32,008,000
LSTM_2	(None, 2000)	32,008,000
Dense	(None, 5909)	11,823,909
(Total parameters : 87,657,909)		
Hyperparameters		
vocab_size	5,909	Number of vocabulary
wordvec_size	2,000	Dimension of word vectors
hidden_size	2,000	Dimension of hidden values of LSTM
time_size	100 ∼ 10	Length of sequence
corpus_size	24,368	Number of corpus

As Korean is a agglutinative language, the meaning of a sentence depends on the nouns/verbs followed by conjunctions. Therefore, it is necessary to separate the nouns/verbs and conjucntions through text preprocessing and use each as a token (morphemes) [29]. In this study, the Okt() class of KoNLPY, a Korean language preprocessing library, was used for preprocessing. The Okt() class is a preprocessing technique developed mainly for tokenizing Korean Twitter texts [30]. For example, the following sentence is preprocessed with Okt() and the result is as follows.

Original sentence: 'Iam,' 'completely,' 'collapsed' (In Korean : '나는' '완전히' '붕괴되었어요')

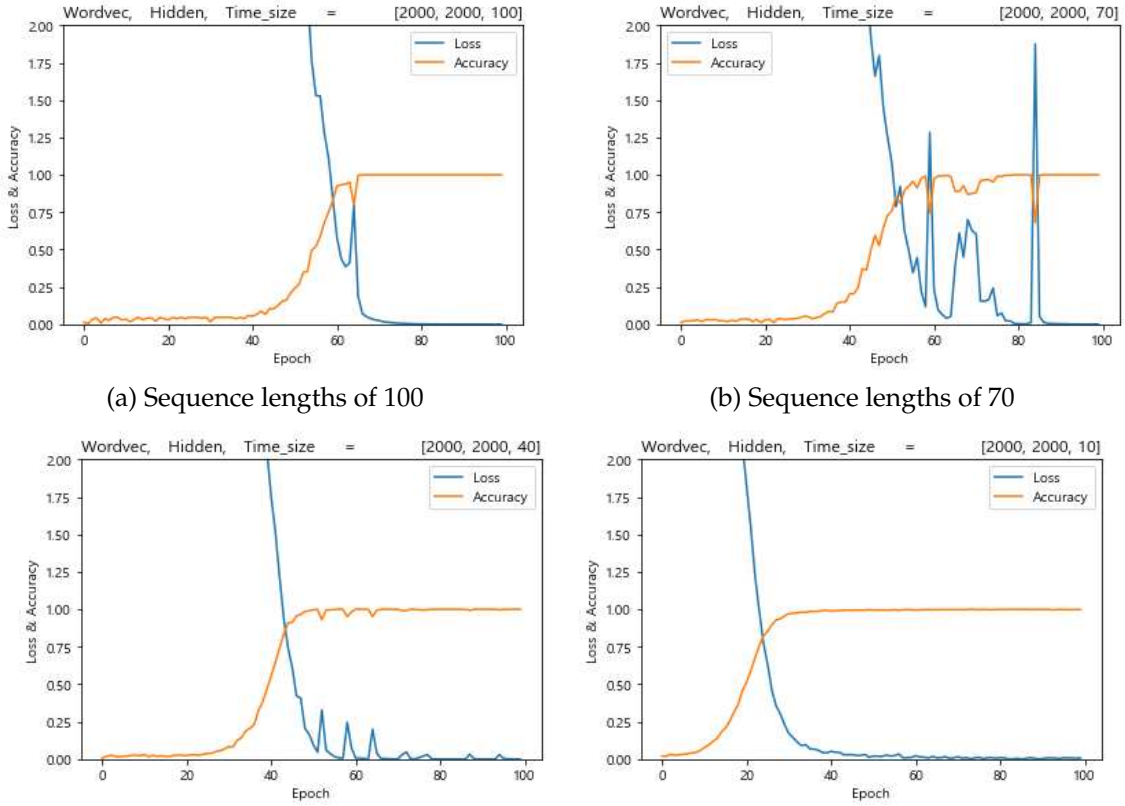
Preprocessed sentence: 'I' 'am' 'completely' 'collapse' '-d' (In Korean : '나' '는' '완전히' '붕괴' '되었어요')

You can see that the original sentence consisted of three words, but after preprocessing, it was divided into five tokens. It makes sense to treat 'am(는)' and '-d(되었어요)' as separate tokens because they are conjunctions attached to the noun 'I' and the verb 'collapse', respectively in Korean.

The sequence length, which corresponds to the number of words entered at a time, was chosen to be between 10 and 100. A sequence length of 10 means that 10 words are entered and the next word is selected from a vocabulary of 5,909 words. For the model used in this study, the number of parameters required to accurately select the next word following the input sequence is about 87 million. A language model uses this many parameters to accurately estimate the next word in a sequence.

The learning loss/accuracy graph as a function of sequence length is shown in Figure 5. We can see that for sequence lengths of 100, 70, 40, and 10, the learning accuracy approaches almost 100% at an epoch of 50-60. However, we can see that the loss/accuracy oscillates somewhat unsteadily as the sequence length increases. This means that the accuracy of estimating the next word from multiple words can change significantly as the sequence length increases. In this study, we mainly used word vectors generated by the embedding layer during the language model training process, so we analyzed the results of word vectors for sequence lengths 40 and 10, which are stable in learning.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2023 doi:10.20944/preprints202311.1558.v1



(a) Sequence lengths of 100

(b) Sequence lengths of 70

(c) Sequence lengths of 40

(d) Sequence lengths of 10

Figure 5. The learning loss/accuracy graph as a function of sequence length.

3.2. *Word vectors - Similarity*

In the movie 'Decision to Leave' analyzed earlier, the main keywords were mountain/sea, Pusan/Leepho, fog, collapse, jellyfish/turtle, pit, and ice cream. The words with high similarity to these words are summarized in Table 2. The cosine similarity between word vectors is calculated as follows.

$$\text{similarity}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{x_1 y_1 + \cdots + x_n y_n}{\sqrt{x_1^2 + \cdots + x_n^2}\sqrt{y_1^2 + \cdots + y_n^2}}$$

($\mathbf{x}$, $\mathbf{y}$: Word vectors ;; x_i, y_i : Elements of the word vectors)

The closer the cosine similarity result is to 1, the more similar the direction of the two word vectors are [31]. In other words, in this study, we used the cosine similarity between the word vectors representing the nature of words as 2,000 numbers to find and analyze words that are similar in direction or nature.

In Table 2, the words in the left column correspond to the sequence length 40, and the words in the right column correspond to the sequence length 10 of the language model. For sequence 40, the words similar to 'mountain' are 'investigation', 'pencil case', and 'left hand', while the words similar to 'sea' are 'trembling', 'investigation', 'left hand', and 'music'. For sequence 10, words similar to 'mountain' were generated: 'frame', 'pass by', 'paper holder', and 'wait', and words similar to 'sea' were generated: 'letter', 'woman', 'wave', and 'suicide'.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2023 doi:10.20944/preprints202311.1558.v1

Table 2. Similar words to the keywords recommended by word vectors.

Keyword	Time_size	
	40	10
Mountain	Investigation	Frame
	Left hand	Pass by
	Pencil case	Paper holder
	Top	Wait
Sea	Tremble	Letter
	Investigation	Women
	Left hand	Wave
	Music	Suicide
Pusan	Doctor	Lunchbox
	Attached	Sushi
	Fast	Fall
	Fills	Running
Leepho	Heart	Suddenly
	Shaking	Computer
	Dining table	Polygraph
	Left hand	Smartphone
Fog	Investigation	Suddenly
	Shaking	Flask
	Left hand	Playback
	Expression	Leepho
Collapse	Investigation	Ace
	Shaking	Same
	Left hand	Rise
	Razor	Immersed
Jellyfish	Music	Dies
	Investigation	Tasteless
	Shaking	Full
	Expression	Stealing
Turtle	Sitting	Adam's apple
	What	Came
	Laughter	Rises
	Taking out	Midnight
Pit	Taking out	Ignition
	Sitting	Dining table
	Puts down	Stopping
	Sigh	Bouquet
Icecream	Laughter	Soft
	Taking out	Computer
	Welcoming	Crazy
	Prayer	Without

Since the mountain is the place where the murder takes place and the investigation begins, the similarity between 'mountain' and 'investigation' is natural, but the words 'pencil case', 'left hand', 'frame', 'paper holder', and 'wait' are somewhat unnatural. In the case of 'pencil case', it is the place where Jang kept the crow's feathers brought to Song by the stray cat that she fed. When the cat catches the crow to reward Song for feeding him, she mumbles to herself in Chinese, which Jang, who is spying on her, records and then listens to the Korean translation on his smartphone.

> (Jang listening to smartphone) I can't control what you kill to eat, but if you're giving me a gift to thank me for feeding you, that's fine, really. If you really want to get me a gift, bring me the heart of that kind detective. I'd like some.
>
> (Jang puts his hand on his left chest, heart prickling.)

It's awkward because it's a literal translation of what Song said in Chinese. Song is telling the cat to bring the love of the kindly Detective Jang. Jang feels Song's affection and is touched to the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2023 doi:10.20944/preprints202311.1558.v1

core. At this point, Jang saves the crow's feather and keeps it in his pencil case, so the pencil case can be seen as a symbol of Song and Jang's attraction to each other. There is a possibility that the 'pencil case' that holds the crow's feather, which expresses the process of changing from murder/suspicion to liking/compassion, is computed as a word with high similarity to the 'mountain'. By using the word vector operation, you can notice these details that are easy to miss and enjoy the movie more deeply.

It is also difficult to think of 'paper holder' as a word with high similarity to 'mountain', but when we see the scene where paper holder appears, we can understand the result of the similarity calculation. This is the scene where Jang is reviewing the documents that record what Song recorded on her smartwatch.

> (Jang takes the paper holder out of his briefcase and reads the recording of Song's smartwatch).
>
> (The day of the fire alarm) (Song's voice) You're wearing Derbys, your face is a little burnt. I can see his beard. He used to shave morning and night. Is he lazy? He looks weak. Because you can't see me? (Song laughs)
>
> (Looking at Lim to go outside and smoke) (Song's voice) I smoked next to him when he was making that weird Chinese food. Lim couldn't stand the smoke, and he was always talking about love in words. I remember the nights Jang spied on me, and I felt like I was being watched over by a man I trusted.

In order to see Jang, who has been transferred to Leepho, Song sets off the fire alarm inside the police station building, and her observations and recordings of Jang evacuating the building are documented. Also, when her second husband, Lim, is nagging her to go outside and smoke a cigarette, she mumbles to herself that she remembers Jang cooking fried rice for her. Since it is a paper holder that organizes papers that expresses Song's nostalgic feelings, it may have a similarity to 'mountain' as 'pencil case'.

'Left hand' refers to Song's left hand when she was first investigated. This word appears frequently because Jang observes the scar on the back of her left hand and the fact that she wears a wedding ring. Since we see the detective Jang looking at the suspect Song's left hand, 'left hand' may also have a high similarity to 'mountain' where the murder/investigation begins.

A word with a high similarity to 'sea' feels more natural in Sequence 10. Sequence 40 shows almost the same results as 'mountain'. However, 'music' might have a high similarity because the song/music 'Fog' is heard in the last scene where Song commits suicide in the sea. In Sequence 10, the words 'woman', 'waves', and 'suicide', which were computed as high similarity words to 'sea', are all suitable as high affinity words to the sea.

Examining high-similarity words gives you a broader and more complex view of the different emotions and meanings represented by the movie's keywords. It's hard to imagine that 'paper hold' or 'pencil case' would have a high affinity for 'mountain', but that's what the language model's word vector similarity computation reveals.

Every location and prop used in a movie is placed for a mise-en-scene that the director intends [32]. The pencil case where the man first hears his lover's crush in an awkward translation, and where he keeps the feather that symbolizes it, is obviously a device that exudes a certain meaning. The paper holder in which the transcript of the lover's intimate monologue is organized is also a deeply emotional device. The word vector similarity computation reveals the effects of these details that are difficult to grasp through superficial movie viewing.

3.3. Word vectors – Analogy

Word vectors can be used to infer words that are similar to the relationship between certain words. For example, in the relation 'King : Queen = Father : ?', you can find words for 'Father' that are similar

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2023 doi:10.20944/preprints202311.1558.v1

to the 'King-Queen' relation [33]. This is called an analogy operation. In the case of 'Decision to Leave', the contrasting word pairs are 'mountain-sea', 'Leepho-Pusan', and 'jellyfish-turtle', so the results of the analogy operation for these word pairs are summarized in Table 3.

Table 3. Corresponding words to the relations of the representative words by analogy operation.

Image/Keyword	Time_size	
	40	**10**
(City – Character)	Investigation	Rises
	Music	Ignition
Leepho – Pusan	Throw away	Force
| |	Bedroom	Question
	Left hand	Sensor
Fog – ?	Ear	Chasing
(Places – Action)	Bandage	Husband
	Pajamas	Song
Mountain – Sea	Put in	Marriage
| |	Mind	Prison
	Pending	House
Collapse – ?	Laughter	Stir
(Animals)	Taking out	Cremation urn
	Sits	Rises
Jellyfish – Turtle	Laughter	Alley
| |	Puts down	Couple
	Doubt	Jewel
Maggot – ?	Sigh	Adam's apple

For sequence length 40, words such as 'investigation', 'music', 'bedroom', and 'left hand' were computed as words corresponding to 'fog', which is similar to the relationship corresponding to 'Leepho-Pusan'. For sequence 10, the following words were computed: rises, ignition, force, question, sensor, chasing. Since fog is a key characteristic of the fishing village of Leepho, words that describe the city of Pusan should be computed. A murder in the mountain of the city of Pusan, an 'investigation' into the murder, and 'question' asked during the investigation are all considered appropriate words.

However, bedroom, music, left hand, etc. are somewhat awkward, so let's look at the meaning of these words through the script. The word music refers to the trending song "Fog"" on the smartphone of the elderly Monday, whom Song visits and cares for, or to Mahler's symphonies, which Song's first husband Kee often listened to. The bedrooms are variously located in Jang's house in Pusan, Ann's house, and Song's house, but they are intimate spaces where Song puts Jang to bed and where Jang has sex with his wife, Ann.

Similar to the relationship between Leepho and fog, the pairings of 'Pusan' are probably words that mean murder/investigation/observation/fascination, which are characteristics of Pusan. The left hand, music, and bedroom can all be seen as objects associated with a place where a murder/investigation begins or takes place. Song's left hand with a suspicious wound, Song's left hand with and without a wedding ring, Jang's bedroom in Pusan where Song puts Jang to sleep, and the music dictating the song 'Fog' can be seen as words that satisfy these relationships.

The results of analogies such as 'mountain-sea' and 'jellyfish-turtle' in Table 3 show a similar trend. In this way, word vectors make it easy to find words that form similar relationships to the main pairing words by applying analogy operation. Through these relationships, you can get a deeper sense of the meaning of the various locations and props in the movie [34].

3.4. Word vector - Visualization

When a 2000-dimensional word vector is reduced to two dimensions by a dimensionality reduction technique, each word vector can be displayed on a plane [35]. Dimensionality reduced word vectors using principal component analysis is shown in Figure 6, which displays the words

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2023 doi:10.20944/preprints202311.1558.v1

Pusan, Leepho, sea, mountain, fog, and collapse, which are the main keywords of the movie. This is a 2D representation of the word vector corresponding to sequence 10.

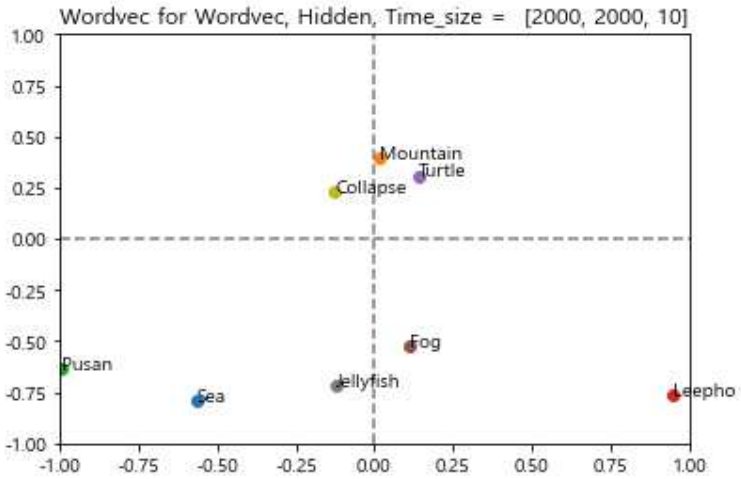


Figure 6. Word vectors in 2D plane.

You can see that 'sea' and 'mountain' are in opposite directions (180 degrees). We can see that the positions of two words with contrasting meanings and symbolism in the movie are actually opposite. The words 'jellyfish' and 'turtl" are also opposite. 'Pusan' and 'Leepho' are not 180 degrees apart, but they are quite far apart. We can see that 'collapse' is closer to 'mountain', and 'fog' is closer to 'Leepho' or 'sea'.

The close proximity of the words 'mountain', 'turtle', and 'collapse' reminds us of the scene in which the insomniac detective is bitten by 'turtle' as the love that began with the murder on the mountain is betrayed/collapsed. The close proximity of the words 'fog', 'jellyfish', and 'sea' reminds us of the jellyfish sinking to the bottom of the sea as Jang is buried by the fog, and sinking into sleep.

By dimensionally reducing the word vector and plotting it on a plane, you can visualize the similarity between words in the direction of the vector. You can also visualize the nature of the words by easily identifying their location on the plane. By analyzing the distribution of keywords in a movie, you can intuitively understand and enjoy the meaning of the movie.

4. Conclusion

This article analyzes Director Chanwook Park's Korean film 'Decision to Leave' using word vectors. The film won the Palme d'Or at the 2022 Cannes Film Festival and is a masterpiece about the suspicion and love between a female suspect and a male detective. The movie starts with a murder in the city's mountain, then shifts to a fishing village as the love disintegrates, and ends with a suicide at sea.

Referring to existing film criticism, the film was analyzed by focusing on places and props that have contrasting meanings, such as Pusan/Leepho, sea/mountain, turtle/jellyfish, fried rice/ice cream, etc. Pusan, the city, and mountain form a group, and Leepho, the village, and sea form a contrasting atmosphere. The names and characteristics of the places show the contrasting nature, and the design of the movie posters and the cover of the script also reveal this characteristic.

The cover of the movie script shows transcription flowing like water from a mountain. Like the plot of the movie, the story starts in the mountain and flows to the sea. In the movie poster, the mountain is placed in the background of the male protagonist and the sea in the background of the female protagonist, expressing the contrasting nature of the mountain and sea.

Although these traditional and critical approaches are logical, it is difficult to objectively prove such meaning or structure in practice. In this study, we tried to verify the main contents derived from movie criticism using word vectors obtained using LSTM, a recurrent neural network technique for

training language models. We trained a language model using movie script Korean texts as a corpus, and obtained 2000 dimensional word vectors in the process. A word vector is a vector of 2000 numbers representing the nature of the words that make up the corpus for training.

Using the 2000-dimensional word vector, we computed words with high similarity to the main keywords, and found words with similar meanings and feelings to each keyword. However, there were also a number of words that were considered to have completely different meanings, and the context of the script in which they were used revealed their similarity in meaning and tone.

We also analyzed the results of analogy operation using word vectors, and again, we were able to find words that form a similar relationship to the relationship between keywords. Similar to the similarity operation, disparate words were also calculated, but by looking at the context, we could see the similarity of meaning and mood.

We applied principal component analysis to the word vector and plotted the dimensionally reduced words on a plane to analyze the location of the words. As a result of plotting the main keywords Pusan/Leepho, sea/mountain, jellyfish/turtle, etc. on the plane, it was found that words with contrasting personalities were close to 180 degrees in direction, and words with similar moods were clustered together.

Through similarity and analogy operations using word vectors and visualization of dimensionally reduced vectors, we were able to objectively confirm existing critical inferences. In some cases, disparate words were derived, but it was a reasonable result when looking at the context of the movie text. This allowed us to discover meanings and symbols that are often missed in traditional criticism. We believe that through the semantic analysis process using these word vectors, we can enjoy deeper movie meaning and symbolic.

Author Contributions: Conceptualization, K.K. ; methodology, K.K. and J.P. ; software, K.K. and J.P.; validation, K.K.; formal analysis, K.K.; investigation, K.K. and J.P.; resources, K.K.; data curation, K.K. and J.P.; writing—original draft preparation, K.K.; writing—review and editing, J.P. and K.K.; visualization, K.K.; supervision, J.P.; project administration, J.P.; funding acquisition, J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2021R1F1A1064073).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Juliadi, R. N.; Puspitarani, Y. Supervised model for sentiment analysis based on hotel review clusters using rapidMiner,*Sinkron* 2022, 7(3), 1–10.
2. Agrawal, M.; Moparthi, N. R. A comprehensive survey on aspect based word embedding models and sentiment analysis classification approaches, *IOS Press Ebooks* 2021, 12(4), 123–135.
3. Shalehanny, S. ; Triayudi, A. ; Handayani, E. T. E. Public's sentiment analysis on shopee-food service using lexicon-based and support vector machine, *Jurnal Riset Informatika* 2021, 4(1), DOI:10.34288/jri.v4i1.287.
4. Di Giovanni, M.; Tasca, T.; Brambilla, M. DataScience-Polimi at SemEval-2022 task 8: stacking language models to predict news article similarity, In Proceedings of the SemEval-2022, Seattle, United States, 14-15 July 2022, 1229–1234.
5. Raza, S.; Reji, D. J.; Ding, C. Dbias: detecting biases and ensuring fairness in news articles, *Int. J. Data Sci. Anal.* 2022, DOI:10.1007/s41060-022-00359-4.
6. Stegmair, J.; Prybutok, V. Trust in google - a textual analysis of news articles about cyberbullying, *Inf. Syst. J.* 2022, 25, 45–63, DOI:10.28945/4894.

7. Suissa, O.; Zhitomirsky-Geffet, M.; Elmalech, A. Around the GLOBE: numerical aggregation question-answering on heterogeneous genealogical knowledge graphs with deep neural networks, *J. Comput. Cult. Herit.* 2023, 16(3), Article No.51, 1—24, DOI:10.1145/3586081.
8. Mitra, B. Neural methods for effective, efficient, and exposure-aware information retrieval, *ACM SIGIR Forum* 2021, 55(1), 1—24, DOI:10.1145/3476415.3476434.
9. Bekhuis, T. Keywords, discoverability, and impact, *J. Med. Libr. Assoc.* 2015, 103(3), 119–120, DOI:10.3163/1536-5050.103.3.002.
10. Yousef, M. W.; Alali, A. Analysis and evaluation of two feature selection algorithms in improving the performance of the sentiment analysis model of arabic tweets, *Int. J. Adv. Comput. Sci. Appl.* 2022, DOI:10.14569/ijacsa.2022.0130683.
11. Kalaiarasu, M.; Kumar, C. Sentiment analysis using improved novel convolutional neural network (SNCNN), *Indonesian J. Comput. Cybern. Syst.* 2021,15(1), 55–64, DOI:10.15837/ijccc.2022.2.4351.
12. Tian, Z. Application of artificial intelligence system in libraries through data mining and content filtering methods, *J. Phys.: Conf. Ser.* 2021, 1952 042091, DOI:10.1088/1742-6596/1952/4/042091.
13. Zhang, C.; Swami, A.; Chawla, N. SHNE: representation learning for semantic-associated heterogeneous networks, In Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Melbourne VIC Australia 11 - 15 February 2019, 690–698, DOI:10.1145/3289600.3291001.
14. Rak, T.: Żyła, R. Using data mining techniques for detecting dependencies in the outcoming data of a web-based system, *Appl. Sci.* 2022, 12, 6115, DOI:10.3390/app12126115.
15. Lee, J. The film 'Decision to Leave', the trans-identity of death towards existence, *Korean Association for Visual Culture* 2022, 41(6), 143–162, DOI:10.21299/jovc.2022.41.6.
16. Park, J. Archetypes and symbols in <Decision to Leave>: Jungian psychoanalysis, *J. of Korean Language and Culture* 2023, 80, 237–261.
17. Kim, M. Semiotic analysis and figurative understanding of film: focusing on the Park Chan-Wook's film 〈Decision to Leave, *J. of the Korea Contents Association* 2023, 23(2), 155-–163, DOI:10.5392/JKCA.2023.23.02.155.
18. Li, M.: Ahn, S. A study of visual representation in the movie <Decision to Break Up> - based on Lacan's psychoanalytic approach, *J. of Cultural Product & Design* 2022, 71(12), 133–143. DOI:10.18555/kicpd.2022.71.012.
19. Yi, S. The <Decision to Leave> on the mountain and the sea, *J. Res. Chin. Novels* 2023, 69(1), 1–22, DOI:10.17004/jrcn.2023..69.001.
20. Son, S. They decide to leave because they can't get together: sexual difference and discrepancy in a movie <Decision to Leave>, *Film Studies* 2022, 94, 67-103, DOI:10.17947/FS.2022.12.94.67.
21. Jin, S. Chinese women in Korean cinema: representation and ethics of women in the movie 'Decision to Leave', *J. of Chinese Literature* 2023, 52, 207–230, DOI:10.35822/JCLLT.2023.01.52.207.
22. Shim, E. Self-rerlexivity in park chan-wook's <Decision to Leave>, *Comp. Lit.* 2023, 89, 89–108, DOI:10.21720/complit89.04.
23. Kang, M. Boundary with the batter in <Decision to Leave>, *Multi-Cultural Contents Studies* 2023, 43, 123–139, DOI:10.15400/mccs.2023.03.43.05.
24. Lee, D. A study on the <Decision to Leave> - focusing on the suspension and formation of meanings, *J. of Lang. & Lit.*, 91, 349–378, DOI:10.15565/jll.2022.6.91.349.
25. Kusum, K.; Panda, S. Sentiment analysis using global vector and long short-term memory, *Int. J. Electr. Comput. Eng.* 2022, 414–422, DOI:10.11591/ijeecs.v26.i1.
26. Setyanto, A.; Laksito, A.; Alarfaj, F.; Alreshoodi, M.; Kusrini,; Oyong, I.; Hayaty, M.; Alomair, A.; Almusallam, N.; Kurniasari, L. Arabic language opinion mining based on long short-term memory (LSTM), Appl. Sci. 2022, 12(9), 4140, DOI:10.3390/app12094140.
27. Kitaoka, N.; Chen, B.;aObashi, Y. Dynamic out-of-vocabulary word registration to language model for speech recognition, *J. Audio, Speech, Music Proc.* 2021, Article no.4, DOI:10.1186/s13636-020-00193-1.
28. Syed, M. H.; Chung, S. MenuNER: domain-adapted BERT based NER approach for a domain with limited dataset and its application to food menu domain, Appl. Sci. 2021, 11(13), 6007, DOI:10.3390/app11136007.
29. Cho, D.; Lee, H.; Kang, S. An empirical study of korean sentence representation with various tokenizations, *Electronics* 2021, 10(7), 845, DOI:10.3390/ELECTRONICS10070845.
30. Park, Y.; Shin, Y. Semantic network analysis of newspaper articles related to agro-healing, *J. People Plants Environ.* 2022, 25(2), 167–176, DOI:10.11628/ksppe.2022.25.2.167.

31. Bingyu, Z.; Arefyev, N. The document vectors using cosine similarity revisited, Proceedings of the 3rd Workshop on Insights from Negative Results in NLP, Association for Computational Linguistics, Dublin, Ireland, 26 May 2022. 129—133, DOI:10.18653/v1/2022.insights-1.17.
32. Guthrie, J. You're tearing me apart: deconstructing mise-en-scène in rebel without a cause, International Summit of the Music & Entertainment Industry Educators Association, Nashville, United States, 21-23 March 2019, 5–12, DOI:10.25101/19.13.
33. Ethayarajh, K. Rotate king to get queen: word relationships as orthogonal transformations in embedding space, In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 3 - 7 November 2019, 3503-–3508, DOI:10.18653/v1/D19-1354.
34. Yanan, Q.; Fuqiang, T. Keyword extraction for film reviews based on social network analysis and natural language technology, *J. Tech. Inform.* 2020, DOI:10.1051/e3sconf/202018903019.
35. Latif, S. H. A. ; Alwan, A. S. ; Mohamed, A. M. Principal component analysis as tool for data reduction with an application, *EUREKA: Phys. Eng.* 2022, 5, 184–198, DOI:10.21303/2461-4262.2022.002577.