

# Discovering Mise-en-scène in Movies by Analyzing Scripts\*

O-Joun Lee<sup>[0000–0001–8921–5443]</sup> and Jin-Taek Kim\*\*

Future IT Innovation Laboratory, Pohang University of Science and Technology  
77, Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do, Republic of Korea 37673  
{ojlee112358, jintae}@postech.ac.kr

**Abstract.** This study aims to analyze mise-en-scène in movies. Although the mise-en-scène includes all the entities that appear on the scene, we focus on metaphorical meanings of the entities. When entities are expressions of the same metaphorical symbol, lexical meanings of the entities will be relevant to each other. Also, if a metaphorical symbol is significant for storytelling, the symbol will appear on most of the scenes. Therefore, we find groups of entities, which share similar lexical meanings and appear all over the movie. As a preliminary study, we have applied these approaches on terms in movie scripts with simple natural language processing techniques. Although we have not conducted evaluation yet, we anticipate that the proposed method will be helpful for analyzing artistic and topical intentions of movie directors.

**Keywords:** Computational Narrative Analysis, Mise-en-scène, Movie Script Analysis, Metaphorical Symbol.

## 1 Introduction

Mise-en-scène literally indicates things located on the scene [2,11]. Simply speaking, this concept covers everything that appears on the scene; e.g., characters, props, backgrounds, etc. Although a few existing studies [1,4] have been conducted on the mise-en-scène in films, they focus on mise-en-scène as physical expressions; e.g., lighting, composition, camera work, etc. These studies can not cover meanings of the expressions, which can be metaphorical symbols located by directors. In a movie ‘Parasite (2019),’ vertical movements (e.g., walking up the stairs) signify changes in social classes.

In this study, we deal with the mise-en-scène as the metaphorical symbol. We cannot interpret meanings of each symbol but attempt to discover which entities in a film have metaphorical intentions. Thus, different from our previous studies [5,6,7,8,10], which focus on interactions between characters, this study concentrates on entities that appears

---

\* Copyright © by the paper’s authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). In: J.-T. Kim, J. J. Jung, E. You, O.-J. Lee (eds.): Proceedings of the 1st International Workshop on Computational Humanities and Social Sciences (Computing4Human 2020), Pohang, Republic of Korea, 15-February-2020, published at <http://ceur-ws.org>

\*\* Corresponding author.

in scenes. Also, as a preliminary study, we analyze terms in movie scripts rather than detect entities in videos.

To discover metaphorical symbols, we first preprocess the movie script. Based on formality of the script, we extract terms that were used for describing backgrounds, events, or characters. Since the metaphorical symbol is expressed by various terms, we cluster terms in the script according to their semantic relevancy by using WordNet [14]. The clusters of terms are candidates of the metaphorical symbol. Then, we attempt to distinguish the metaphorical symbol from other clusters of terms that were used in their lexical meanings.

## 2 Movie Scripts

Movie scripts are highly formalized documents, as displayed in Fig. 1. Various information (e.g., scene headings, action descriptions, names of characters, dialogues, background descriptions, etc.) is written according to a consistent convention. When the metaphorical symbol is placed as spatial backgrounds (e.g., stairs or basements), it will be depicted in the background description or even in the scene heading. If the spatial background is related to behaviors of characters, it can be described in the action description (e.g., walking up the stairs). Also, if the symbol is a prop, it will be depicted on the dialogue or action description.

Thus, we extract terms from scene headings, background descriptions, action descriptions, and dialogues in movie scripts. Then, we compose sets of terms that occurred in each scene. It is difficult to find which terms have metaphorical meanings. However, the metaphorical meaning also originates from the lexical meaning. Therefore, terms related to a metaphorical symbol will have similar lexical meanings and appear on most of the scenes.

## 3 Discovering Metaphorical Symbols

However, the metaphorical symbols are normally abstract concepts, and terms are annotations of their expressions in the background, dialogue, and behavior. In the case of ‘Parasite (2019),’ Joon-ho Bong uses vertical movements as metaphors, and the movements were expressed mainly by stairs and basements. Thus, terms in the script are difficult to be directly metaphorical symbols. We have to collect terms that frequently appear. Then, we discover groups of terms that can be used for expressing the same metaphorical meaning.

If a set of terms are expressions of the same metaphorical symbol, they will be related in terms of their lexical meanings. For example, the stairs and basements commonly have meanings related to heights. Although the expressions can have multi-layered metaphors, the film is one of the most commercial and popular media. To reveal the semantic relevancy, we use the WordNet [14]. The WordNet is an ontological database of English words, which contains not only definitions of the words but also relations between the words. These relations include hypernyms, hyponyms, coordinate terms, meronyms, holonyms, troponyms, etc.

```

7 INT. ELEVATOR, APARTMENT BUILDING - CONTINUOUS 7
//Scene heading

Joker steps onto the wheezing elevator, harsh fluorescent
lights, graffiti on the walls. As the door closes, he hears--
//Action and background description

SOPHIE (OS) //Name of a speaker
Wait!! //Dialogue

He puts his foot out with some panache to stop the closing
door-- He's a romantic at heart. Ding.
//Action description

And SOPHIE DUMOND (late 20's), tired eyes, hands filled with
grocery bags, steps onto the elevator with GIGI, her 5-year-
old daughter.

SOPHIE
Thank you.
(realizing)
Of course it's you,-- everyone else
in this building is just so fucking
rude.

Joker nods "thanks." Holds his breath, hoping he doesn't
start to laugh.

Floors dinging as the elevator rises. //Background description

```

Fig. 1: A part of a script of 'Joker (2019).'

For the terms in a movie script, we conduct stemming and removing stop words (e.g., articles, pronouns, auxiliary verbs, punctuation marks, etc.). Then, we measure distances between the terms by using the WordNet. The distance is measured with two approaches. First, we use length of the shortest path between two terms on the WordNet. The relations in the WordNet indicate semantic relationships between terms. Thus, if we can reach from a term to another term in fewer hops, we can assume that the two terms are semantically more related.

Second, a few terms are not directly related in their semantics, while they are conceptually relevant. Although we imagine 'height' from both of stairs and basements, the two terms are not much close on the WordNet. Thus, we embed the terms in movie script based on terms in their definition by using SkipGram in Word2Vec [13,5]. We do not conduct SkipGram based on sentences in the script, since the script is close to the everyday language rather than pellucid writings. The vector representations of terms enable us to conveniently compare semantics of the terms. In this study, we use the

Euclidean distance to compare the vectors. When a term has multiple homonyms, we are difficult to determine meaning of the term in the current context. Thus, we measure distances for all the homonyms and aggregate the distances by the arithmetic mean.

After normalizing these distances into  $[0, 1]$ , we cluster these terms using the  $k$ -NN (Nearest Neighborhood) clustering. The cluster model is composed to minimize semantic distances between terms in each cluster. We determine the number of clusters ( $k$ ) by maximizing average distances between clusters (external adjacency) and minimizing average distances among terms in each cluster (internal compactness). However, not all the clusters will correspond to metaphorical symbols. Most of them might be related to genres, insignificant subjects, or daily vocabularies.

We distinguish metaphorical symbols from other clusters, which consists of terms used in their lexical meanings, by using a few heuristics. First, the metaphorical symbol is to deliver the director’s intention in a euphemistic and sophisticated way. Thus, the symbol will be used in most of the scenes in a movie. Second, stories in narrative multimedia are led by conflicts of characters (mainly, around the protagonists) [12,16]. The conflict is a process that the characters try to recover normality of their lives against oppressive situations. This process forces the characters to change themselves, and the director tells artistic or topical intentions through the change. Thus, in lots of movies, the metaphorical symbols allude to purposes, personalities, or situations of characters [2,11]. And, the characters, which are implied by the symbols, might be significant characters.

Therefore, we assess each term cluster based on how many scenes terms in the cluster appeared in. Also, to consider significance of characters, we use node centrality of the characters on the character networks, as with our previous studies [3,9,10,15]. This can be formulated as:

$$S(C_n) = \sum_{\forall s_{\alpha,l} \in \mathbb{C}_\alpha} I(s_{\alpha,l}, C_n) \times \mathcal{C}(C_n) \quad (1)$$

where  $s_{\alpha,l} \in \mathbb{C}_\alpha$  indicates the  $l$ -th scene in a movie ( $\mathbb{C}_\alpha$ ),  $I(s_{\alpha,l}, C_n)$  is an indicator function for whether terms in  $C_n$  occurred on  $s_{\alpha,l}$ , and  $\mathcal{C}(C_n)$  refers to the summation of node centrality for characters that appeared in  $s_{\alpha,l}$ . Then,  $S(C_n)$  is the significance of the  $n$ -th term cluster. For composing character networks, CharNetBuilder<sup>1</sup> is used, and scenes in the script are segmented by the scene headings. The centrality of characters is measured by an average of the degree, betweenness, and closeness centrality. Finally, we can choose term clusters which have distinctively higher significance than the others, by using thresholds or clustering.

## 4 Conclusion

This study has proposed a method for discovering mise-en-scène in movies by analyzing movie scripts, by concentrating on the mise-en-scène as the metaphorical symbol. The proposed method is significant, since the existing studies on computationally analyzing mise-en-scène have only focused on physical expressions rather than their meanings. However, this study also have various limitations. First, since we have not conducted

<sup>1</sup> <https://github.com/0-JounLee/CharNetBuilder>

evaluation for the proposed method yet, this approach is still remaining as a theoretical methodology. Also, narrative multimedia (including movies) employ various techniques spanning one or more physical expressions (i.e., visual, audible, and verbal expressions are compositely used). Thus, to discover the mise-en-scène in higher accuracy, analyzing and synchronizing multi-modal data is necessary, as with our previous study [9].

## Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICT Consilience Creative program (IITP-2019-2011-1-00783) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

## References

1. Deldjoo, Y., Elahi, M., Cremonesi, P., Garzotto, F., Piazzolla, P.: Recommending movies based on mise-en-scene design. In: Kaye, J., Druin, A., Lampe, C., Morris, D., Hourcade, J.P. (eds.) Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2016), pp. 1540–1547. ACM Press, San Jose, CA, USA (May 2016). <https://doi.org/10.1145/2851581.2892551>
2. Deleyto, C.: Focalisation in film narrative. *Atlantis* **13**(1/2), 159–177 (Nov 1991)
3. Jung, J.E., Lee, O.J., You, E.S., Nam, M.H.: A computational model of transmedia ecosystem for story-based contents. *Multimedia Tools and Applications* **76**(8), 10371–10388 (Apr 2017). <https://doi.org/10.1007/s11042-016-3626-5>
4. Kvisgaard, A., Klem, S.O., Nielsen, T.L., Rafferty, E.I., Nilsson, N.C., Hoeg, E.R., Nordahl, R.: Frames to zones: Applying mise-en-scène techniques in cinematic virtual reality. In: Proceedings of the 5th IEEE Workshop on Everyday Virtual Reality (WEVR 2019). IEEE, Osaka, Japan (Mar 2019). <https://doi.org/10.1109/wevr.2019.8809592>
5. Lee, O.J.: Learning Distributed Representations of Character Networks for Computational Narrative Analytics. Ph.D. thesis, Chung-Ang University, Seoul, Republic of Korea (Aug 2019)
6. Lee, O.J., Jo, N., Jung, J.J.: Measuring character-based story similarity by analyzing movie scripts. In: Jorge, A.M., Campos, R., Jatowt, A., Nunes, S. (eds.) Proceedings of the 1st Workshop on Narrative Extraction From Text (Text2Story 2018) co-located with the 40th European Conference on Information Retrieval (ECIR 2018). CEUR Workshop Proceedings, vol. 2077, pp. 41–45. CEUR-WS.org, Grenoble, France (Mar 2018)
7. Lee, O.J., Jung, J.J.: Explainable movie recommendation systems by using story-based similarity. In: Said, A., Komatsu, T. (eds.) Joint Proceedings of the ACM IUI 2018 Workshops co-located with the 23rd ACM Conference on Intelligent User Interfaces (ACM IUI 2018). CEUR Workshop Proceedings, vol. 2068. CEUR-WS.org, Tokyo, Japan (Mar 2018)
8. Lee, O.J., Jung, J.J.: Character network embedding-based plot structure discovery in narrative multimedia. In: Akerkar, R., Jung, J.J. (eds.) Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics (WIMS 2019), pp. 15:1–15:9. ACM, Seoul, Republic of Korea (Jun 2019). <https://doi.org/10.1145/3326467.3326485>
9. Lee, O.J., Jung, J.J.: Integrating character networks for extracting narratives from multimodal data. *Information Processing and Management* **56**(5), 1894–1923 (Sep 2019). <https://doi.org/10.1016/j.ipm.2019.02.005>
10. Lee, O.J., Jung, J.J.: Modeling affective character network for story analytics. *Future Generation Computer Systems* **92**, 458–478 (Mar 2019). <https://doi.org/10.1016/j.future.2018.01.030>

11. Martin, A.: *Mise en Scène and Film Style: From Classical Hollywood to New Media Art*. Palgrave Close Readings in Film and Television, Palgrave Macmillan UK (2014). <https://doi.org/10.1057/9781137269959>
12. McKee, R.: *Story: Substance, Structure, Style and the Principles of Screenwriting*. Harper-Collins, New York, NY, USA (Nov 1997)
13. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Burges, C.J.C., Bottou, L., Ghahramani, Z., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems 26: Proceedings of 27th Annual Conference on Neural Information Processing Systems (NIPS 2013)*. pp. 3111–3119. Curran Associates, Inc., Lake Tahoe, Nevada, US (Dec 2013)
14. Miller, G.A.: WordNet: a lexical database for english. *Communications of the ACM* **38**(11), 39–41 (Nov 1995). <https://doi.org/10.1145/219717.219748>
15. Tran, Q.D., Hwang, D., Lee, O.J., Jung, J.E.: Exploiting character networks for movie summarization. *Multimedia Tools and Applications* **76**(8), 10357–10369 (Apr 2017). <https://doi.org/10.1007/s11042-016-3633-6>
16. Truby, J.: *The anatomy of story: 22 steps to becoming a master storyteller*. Farrar, Straus and Giroux (Oct 2008)