A COMPUTATIONAL FRAMEWORK FOR STUDYING THE EVOLUTION OF CULTURAL TRAITS AMONG CHATTING AGENTS USING A GENERATIVE MODEL

Reiji Suzuki[*], Soichiro Hirata and Takaya Arita

[*]Corresponding Author: reiji@nagoya-u.jp
Nagoya University, Japan

Constructive approaches have played a significant role in the understanding of cultural evolution and meme transmission, showing social relationships among hosts can affect the diversification of memes (Kobayashi et al. 2013, Weng et al. 2021). However, it is still unclear how novel memes emerge from complex communications in human language in the real world, such as chatting in a face-to-face situation and text-based communication in SNSs, and propagate in the population while a bigdata analysis contributed to clarifying the open-ended evolutionary dynamics of Web services such as social tags (Ikegami et al. 2019).

We propose a computational framework for studying the cultural evolution of memes among chatting agents using a generative model (Fig. 1 (i)). Individuals are represented as agents and arranged in a two-dimensional social space. The space abstracts their social relationships (i.e., the closer distance between agents represents more intimate relationships among them), which is inspired by the social particle swarm model (Nishimoto et al. 2013). When another neighboring agent is in its interaction radius, a communicative interaction occurs between them ((i)-(a)). Each agent has genetic and cultural traits, each composed of a list of words. The former represents non-evolvable and intrinsic preferences to the topic of a sentence to utter (e.g., gossip, (Mesoudi et al. 2006)). The latter represents the cultural ones, which evolve in the model. An agent utters a sentence produced by a pre-trained model of Japanese sentence generation (rinna/japanese-gpt2-small) based on GPT-2 (Radford et al. 2019). This model predicts and generates Japanese words that follow a given phrase. We generate an utterance of an agent by using a list of words in its genetic traits followed by its cultural traits as a given phrase, which is regenerated with a certain probability at each time step. We assume that agents who share more words among their utterances are
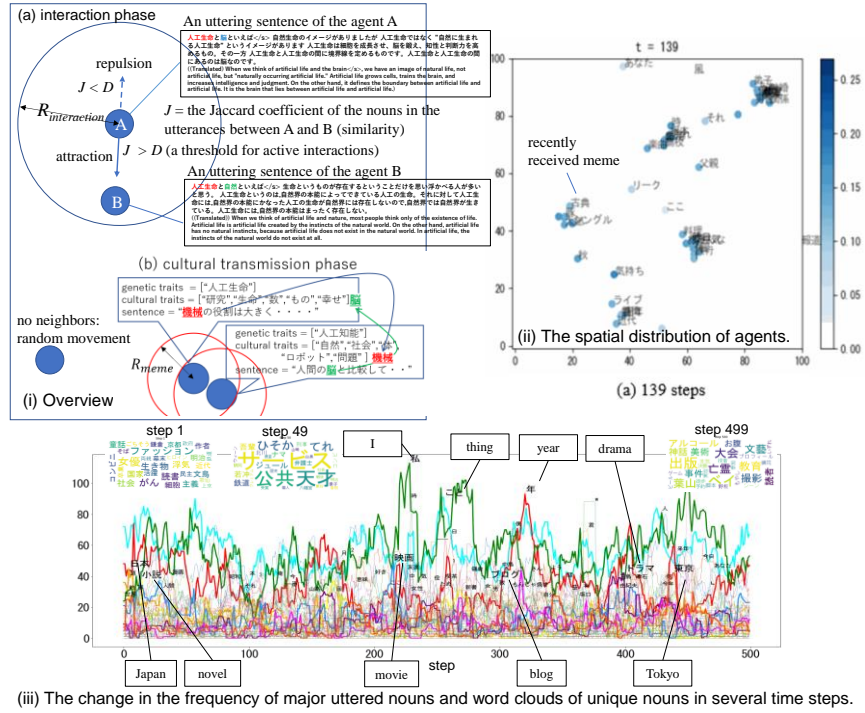
Fig. 1: (i) Overview of the model and (ii, iii) results.

speaking about the topics more attractive to both. We measure the Jaccard index between the sets of the agents' uttering nouns. If the index is above the threshold value, the interparticle attraction occurs, and the repulsion occurs otherwise, of which strength is inversely proportional to the distance between the agents. We also introduce cultural transmission of traits into the interaction process ((i)-(b)). Each agent randomly extracts a noun from the sentence uttered by each closely neighboring agent with a certain probability and adds it to its cultural traits (and removes the oldest one if the number of traits exceeds the limit).

The preliminary experiments with agents, each of whose genetic traits had one of the names of famous novelists mainly in Heian - Reiwa periods in Japan, showed that agents were actively moving and exchanging novel traits with neighboring agents and the emergence and collapse of chatting clusters happened repeatedly (ii). In addition, the agents initially tended to utter words related to Japanese literature, and then these were taken over by more various words, and further trends of Japanese literature emerged (iii). These indicate that the proposed framework enables us to discuss the cultural evolution of novel traits in communicating agents in language.

References

Ikegami, T., Hashimoto, Y. and Oka, M. (2019). Open-ended evolution and a mechanism of novelties in web services. *Artificial Life*, 25(2), 168-177.

Kobayashi, K., Suzuki, R. and Arita, T. (2012). Emergence of autocatalytic reaction in a meme propagation model based on particle motion. *Artificial Life and Robotics*, 17, 1, 92-96.

Mesoudi, A, Whiten, A. and Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, 97(3), 405-423.

Nishimoto, K., Suzuki, R. and Arita, T. (2013). Social particle swarm: explosive particle dynamics on cooperative/defective forces. In *Proceedings of IEEE ALIFE 2013 (IEEE SSCI 2013)* (pp. 134-139).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.

Weng, L. and Flammini, A., Vespignani, A. and Menczer, F. (2021). Blockchain and multi-agent system for meme discovery and prediction in social network. *Knowledge-Based Systems*, 229, 107368.