Simplification Modeling System by Projecting Paradigmatic to Syntagmatic Simplification for Explaining Artistic Texts

Ryotaro Kamimura
Tokai University
Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan
e-mail: ryotarokami@gmail.com

***Abstract*—The present paper proposes a small-sized neural network, called the *simplification modeling system*, designed to contain a massive amount of new information. This type of model is urgently needed because the complexity of neural networks and data sets has been rapidly increasing, making it increasingly difficult both to further augment model complexity and to interpret the underlying inference mechanisms. The proposed model is based on the concept of the *secondary modeling system*, which has been used to describe artistic texts, where new information is generated by observing texts from multiple viewpoints and forcing these viewpoints to compete with one another. This method is applied to the analysis of structural constraints in artistic texts that are imposed on natural language. The results obtained from French texts show that artistic texts tend to increase diversity at the word level while simultaneously increasing regularity at the character level. In this way, artistic texts actively control both diversity and regularity, producing a form of tension or contradiction that leads to the creation of new information.**

Keywords-simplification modeling system; compact neural networks; multi-level information processing; paradigmatic simplification; syntagmatic simplification; artistic text analysis.

I. Introduction

The present paper proposes a hypothesis of simplification in neural networks, based on our previous work [2]. Specifically, we assume that a network attempts to simplify its configuration and learning processes as much as possible. Traditionally, one common approach to dealing with increasingly complicated phenomena and data has been to increase the number of components in a model. As the number of components grows and network complexity increases, it becomes possible, in principle, to solve almost any complicated problems. However, as the complexity of neural networks as well as data sets continues to increase, it becomes increasingly difficult to regulate them to obtain suitable solutions. Moreover, a serious challenge arises in interpreting the final internal representations.

This situation suggests that it is not feasible to increase the complexity of network configurations indefinitely, and that it is time to reconsider the complexity problem from a different viewpoint. We therefore assume the necessity of observing a network and its components from multiple viewpoints in order to store and process massive amounts of information. Based on this idea, we propose a small-sized or compact model that aims to store and process as much information as possible. A system with such a simplification bias can be called a *simplification modeling system*, following the concept of compact models for artistic texts [1]. This system contains only a limited amount of information from a single viewpoint or level; however, when it is observed from multiple viewpoints, the total amount of information created through these viewpoints can become much larger. From any single viewpoint, the information to be stored is naturally restricted, but by introducing multiple viewpoints, it becomes possible to store substantially more information. If many different types of viewpoints or levels can be incorporated into a network, the information storage capacity can increase in proportion to the number of levels or viewpoints, while keeping the actual size of the network as small as possible.

In this context, we treat a neural network as a simplification modeling system. Within this system, we assume two types of simplification, namely, *paradigmatic* and *syntagmatic* simplification. In paradigmatic simplification, all components—for example, the network itself, layers, neurons, and connection weights—are considered, and each component attempts to simplify its own configuration as much as possible. Ideally, these components are assumed to be simplified even without explicit input information. This implies that paradigmatic simplification presupposes a form of self-organization, which is induced solely by the simplification force.

In contrast, syntagmatic simplification is applied to learning procedures, in which components resulting from paradigmatic simplification are selected, combined, and sequentially arranged. This process can be referred to as "projection." In actual learning, paradigmatic simplification is projected onto syntagmatic simplification. This projection is the result of severe competition among paradigmatic components, each attempting to be projected into the syntagmatic level.

While paradigmatic simplification tends to self-organize independently of input and output information, syntagmatic simplification seeks to use as little information as possible from both inputs and outputs. Furthermore, learning procedures within syntagmatic simplification must compete with one another in order to adapt to actual circumstances.

The paper is organized as follows. In Section Ⅱ, we try to explain briefly the studies on information-theoretic methods and related works on artistic texts. In Section Ⅲ, computational procedures for simplification are described. Section IV presents the experimental results on poetic texts, highlighting the roles of diversity and regularity. We conclude the paper in Section V.

 ISBN: 978-1-68558-399-6

II. Related Work

An artistic text can be regarded as a compact model for storing and communicating valuable information across time, place, and audience. For this reason, numerous studies have attempted to clarify how multi-level information is produced in artistic texts. For example, the mathematician Kolmogorov analyzed Russian poetry using an extended information-theoretic framework [1]. He distinguished between two types of information. The first, denoted by h_1, represents the amount of information through which meaning can be transmitted per unit length, while the second, h_2, describes the number of different expressions available for transmitting information. According to this view, artistic effects arise only when the second type of flexible information exceeds the structural constraints imposed on the artistic text. On the other hand, Roman Jakobson analyzed artistic effects through the concept of the poetic function, emphasizing not only meaning but also artistic and verbal form [3][4]. The poetic function highlights formal constraints within a text as a central mechanism for producing new information. In particular, Jakobson argued that the poetic function projects the principle of equivalence from the axis of selection onto the axis of combination.

The notions of poetic function, flexible information, and the axes of selection and combination are closely related to our formulation. In our framework, the axis of selection corresponds to paradigmatic simplification, whereas the axis of combination corresponds to syntagmatic simplification. Moreover, this framework of flexible information in poetic texts is closely connected to our hypothesis that components and learning procedures should be examined from multiple viewpoints. However, conventional information-theoretic approaches in neural networks are not well suited to this type of multi-level information processing.

Information-theoretic methods have attracted attention since the early stages of neural network research. In particular, mutual information–based learning methods have been actively studied [5], as exemplified by Linsker's principle of maximum information preservation [6]. Closely related to the present study, information bottleneck methods aim to distinguish mutual information associated with inputs and outputs, in a manner similar to our forward and backward simplification [7]. However, these approaches can be regarded as methods that analyze information flow from a single viewpoint. In particular, they focus on extracting information from input patterns as faithfully as possible.

In contrast, the model proposed in this paper examines information flow in neural networks from multiple viewpoints. In some components, information should be minimized, whereas in others it should be maximized, depending on the role of each component. More specifically, in paradigmatic simplification, components attempt to maximize or minimize information depending on the situation, with information being generated without reliance on external inputs. In syntagmatic simplification, learning procedures aim to increase information related to inputs or outputs, while components obtained through paradigmatic simplification are projected onto the syntagmatic level. Our simplification principle is assumed to operate across all these situations, such that components and learning procedures are used both to increase and to decrease information content in order to achieve effective simplification.

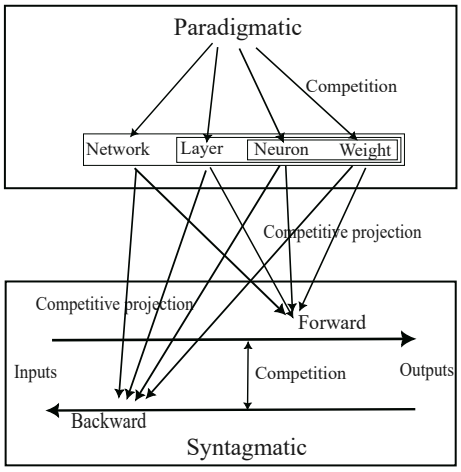

Figure 1. Concept of paradigmatic and syntagmatic simplification.

III. Theory and Computational Methods

A. Paradigmatic and Syntagmatic Simplification

The upper illustration in Figure 1 presents the concept of paradigmatic simplification. In paradigmatic simplification, all components, including the network itself, attempt to simplify their configurations as much as possible by controlling potentiality or information. Some components tend to increase information, whereas others attempt to reduce it. These components then compete with one another to be selected and projected into syntagmatic simplification.

In syntagmatic simplification, forward learning proceeds from input to output, while backward learning proceeds from output to input, using the minimum possible input and output information, as illustrated in Figure 2. Forward simplification redistributes information from lower layers to upper layers, as shown in the upper part of Figure 2, whereas backward simplification redistributes information from the output level to lower levels, as shown in the lower part of the same figure. These learning procedures then compete with one another during training, as depicted in the central part of Figure 2.

B. Network Level

At the network level, the strength of all connection weights should be as small as possible. The individual potentiality of a connection weight from the jth neuron to the kth neuron is defined as

$$u_{jk} = |w_{jk}|. \tag{1}$$

As noted above, all final quantities defined in this section can be obtained by averaging over all layers, including the input and output layers. For simplicity, it is assumed that all connection weights have nonzero strength.

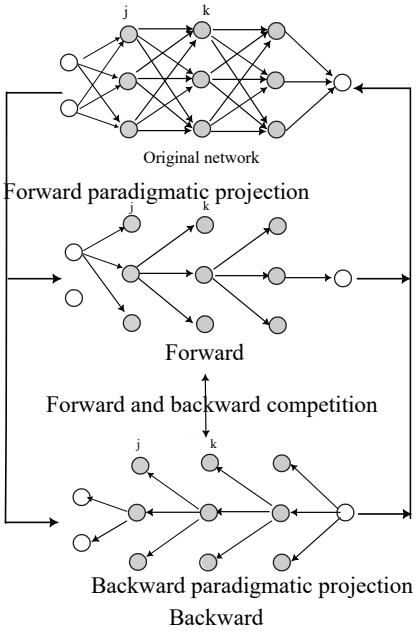

Figure 2. Original network configurations, final configurations obtained by forward and backward simplification, and their competition.

The primary objective of simplification is to reduce the total cost potentiality of the network. The *cost potentiality* is defined as the sum of individual potentialities:

$$C = \frac{1}{n_j n_k} \sum_{jk} u_{jk}, \tag{2}$$

where n_j and n_k denote the numbers of neurons in the corresponding layers.

Simplification aims to reduce this cost potentiality, which can be achieved in various ways.

C. Layer Level

At the layer level, simplification is applied to neuron activation patterns. In this paper, the strength of a neuron is represented by collectively considering all incoming connection weights. This type of potentiality is referred to as *collective simplification*. Collective simplification aims to simplify a group of connections from one layer to the subsequent layer. The potentiality of the kth neuron is defined as the sum of the absolute values of its incoming weights:

$$u_k = \sum_j u_{jk}. \tag{3}$$

The normalized potentiality of the kth neuron is then given by

$$h_k = \frac{u_k}{\max_{k'} u_{k'}}. \tag{4}$$

The collective potentiality is defined as

$$H = \frac{1}{n_k} \sum_k h_k. \tag{5}$$

D. Neuron Level

At the neuron level, all incoming connection weights are treated individually. This type of potentiality is referred to as *individual simplification*. Individual potentiality is defined for each connection weight. By normalizing the absolute weight by its maximum value within the neuron, the relative individual potentiality is defined as

$$g_{jk} = \frac{u_{jk}}{\max_{j'} u_{j'k}}. \tag{6}$$

Summing over all layers yields the individual potentiality:

$$G = \frac{1}{n_j n_k} \sum_{jk} \frac{u_{jk}}{\max_{j'} u_{j'k}}. \tag{7}$$

E. Unified Level for Contradiction Resolution

As discussed in the related work section, competition is applied both within each level and across different levels. We therefore define the relation, or contradiction, between collective and individual potentialities as

$$D = H - G. \tag{8}$$

As this contradiction increases, collective potentiality at the layer level increases while individual potentiality at the neuron level decreases, indicating a growing discrepancy between neurons and their connection weights.

This contradiction can be extended to the entire network by incorporating the cost potentiality. The objective is to increase the contradiction D while minimizing the cost, which leads to the contradiction ratio

$$R = \frac{D}{C}. \tag{9}$$

An increase in this ratio indicates a stronger contradiction between collective and individual potentiality accompanied by a reduction in total cost. Thus, all levels maintain simplification, while the contradiction between layer and neuron levels generates the complexity necessary for learning.

IV. Results and Discussion

A. Experimental Outline

In this study, we analyzed artistic texts as an application of the simplification modeling system. As discussed in the introductory section, artistic texts have long been regarded as typical examples of secondary modeling systems, in which additional structural constraints are imposed on the natural language of the primary modeling system. Artistic texts, particularly verse, are constrained by rhythmic and metrical forms, which are strictly imposed on natural language. Nevertheless, even under such strong constraints, artistic texts are expected to store and communicate a large amount of information across time and audience. We assume that these structural constraints, such as metrical and rhythmic forms, function to compact representations by distributing information across multiple levels and components. Secondary modeling systems were originally developed to incorporate cultural and social influences; however, in the present study, our simplification

model intentionally excludes external influences as much as possible. Through this exclusion, the system is expected to accommodate a wide variety of new information.

For the experiments, we attempted to distinguish between verse and prose texts extracted from works by the French poet Baudelaire. We extracted 100 texts from *Le Spleen de Paris* and *Les Fleurs du Mal* and classified them as prose or verse. The aim of this experiment was to investigate how verse imposes formal constraints on primary natural language in order to store and transmit information. This task was selected to demonstrate the applicability of the simplification modeling system, as the present method does not directly address semantic features of natural language. Because verse emphasizes formal constraints, it is more suitable for analysis using the proposed framework.

We employed networks with ten hidden layers, as shown in Figure 3. Seven input variables were used: type/token ratio, Shannon entropy (word entropy), conditional entropy, average PMI, Kolmogorov bits per character, character bigram entropy, and approximate entropy (syllabic). These information-theoretic indices measure lexical diversity and the degree of formal constraints imposed on natural language. The original multilayer neural network was compressed layer by layer into the simplest network without hidden layers. The weights of the compressed network were then compared with the correlation coefficients between inputs and targets computed from the original data set.

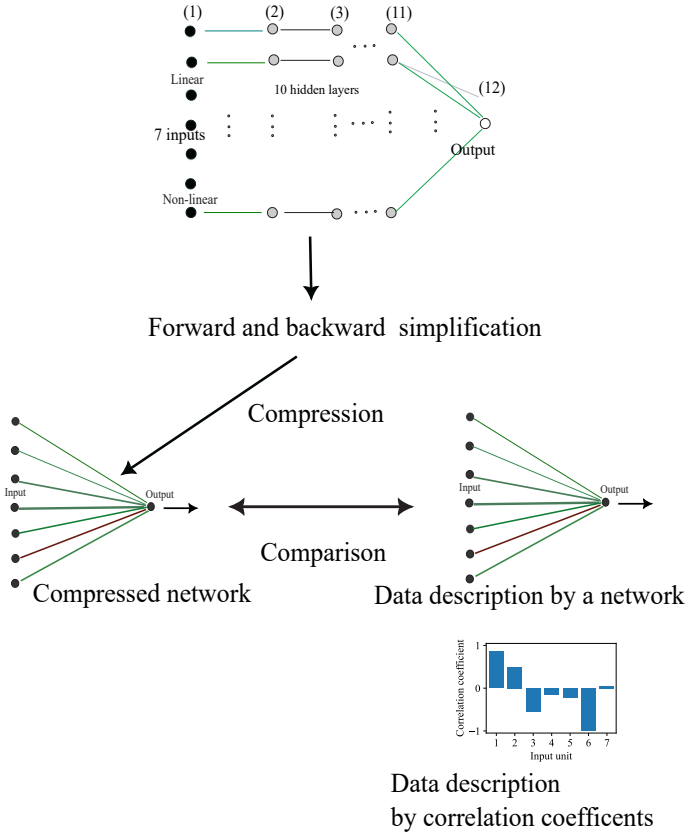

Figure 3. An original network architecture with ten hidden layers (upper), compressed network (middle), and supposed network based on correlation coefficients between inputs and target (lower).

B. Numerical Summary

We first summarize the experimental results using several numerical indicators. Table I presents the results obtained by five methods, including logistic regression for comparison. Backward learning produced the most favorable potentiality values (shown in bold) for all metrics except generalization performance. However, generalization accuracy was the lowest (0.576), indicating that favorable potentiality values alone cannot be fully exploited when backward learning is applied in isolation. When forward and backward learning were combined, the highest generalization accuracy (0.712) was achieved, with most potentiality values appropriately adjusted relative to those obtained by backward learning alone.

The conventional method without potentiality control naturally produced nearly identical values for all potentiality measures. In contrast, all potentiality-based methods effectively increased or decreased the corresponding quantities. In particular, the forward potentiality method increased collective forward potentiality (0.734), reduced individual forward potentiality (0.504), and increased the ratio of potentiality difference to cost (1.159). Logistic regression yielded the second-lowest generalization performance (0.624).

These results indicate that each potentiality-based method was effective in optimizing its target quantities; however, the joint use of forward and backward methods provided the most appropriate balance of potentialities for improving generalization.

TABLE I. SUMMARY OF RESULTS, BASED ON MAXIMUM TESTING ACCURACY.

Meth	Type	Step	Coll	Indi	C-I	Cost	Ratio	Accu
Conv	F	899	0.688	0.518	0.170	0.209	0.814	0.680
	B	899	0.700	0.556	0.144	0.209	0.689	0.680
For	F	965	0.734	0.504	0.230	0.198	1.159	0.680
	B	965	0.690	0.551	0.139	0.198	0.700	0.680
Back	F	703	0.643	**0.460**	0.183	**0.090**	2.029	0.576
	B	703	**0.876**	0.470	**0.406**	**0.090**	**4.516**	0.576
FB	F	1126	0.632	0.470	0.162	0.134	1.210	**0.712**
	B	1126	0.852	0.476	0.376	0.134	2.807	**0.712**
LR								0.624

Ratio: (C-I)/Cost; LR: logistic regression; F: forward; B: backward

C. Collective and Individual Potentiality

We compared the results obtained using the conventional method with those obtained using the forward and backward methods, because the differences between these approaches can be explained clearly. Figure 4(a) and (b) show the values obtained using forward and backward potentiality computation under the conventional method. Although slight differences can be observed, nearly identical values are naturally obtained by the conventional method. Figure 4(c) and (d) show the values computed using the forward and backward approaches when both methods were applied sequentially. Specifically, the forward method was applied for the first 500 learning steps, after which the backward method was applied for the remaining steps. The backward method in Figure 4(d) produced clearer

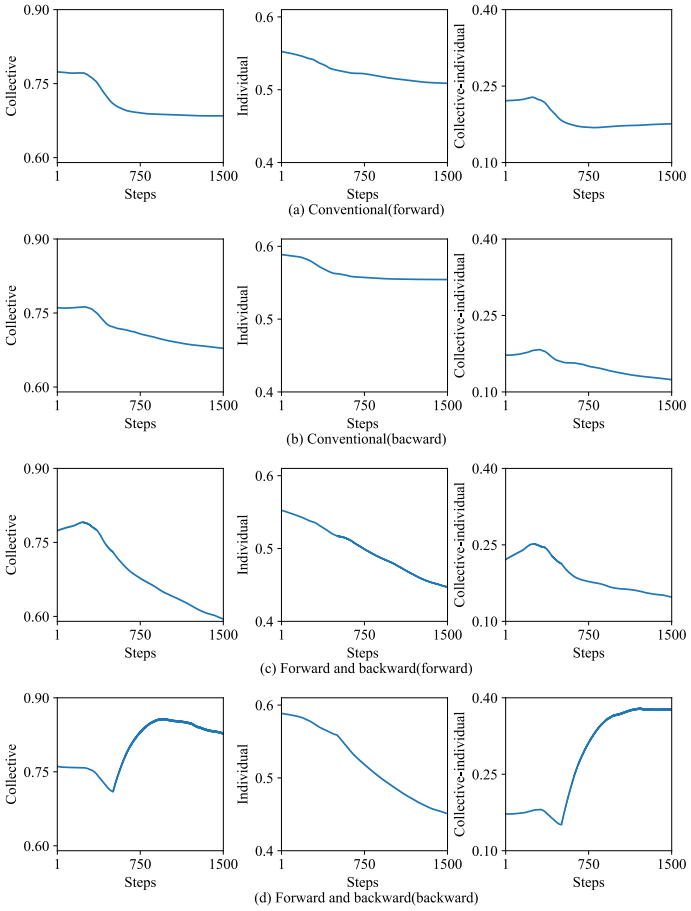

Figure 4. Collective (left), individual (middle) potentiality, and difference or contradiction between collective and individual potentiality (right) as a function of the number of learning steps (epochs) by the conventional method with the forward computation (a) and the backward computation (b), and by the forward and backward simplification with forward potentiality values (c) and backward potentiality values (d).

effects, where collective potentiality increased substantially, individual potentiality decreased sharply, and the difference between them increased significantly. In contrast, the forward method in Figure 4(c) exhibited limited effectiveness, with decreases in collective potentiality, individual potentiality, and their difference.

These results indicate that when forward and backward learning are applied jointly, backward learning dominates the competition and effectively optimizes the network for improved generalization. The losing process can be interpreted as contributing additional complexity necessary for learning, as summarized in Table I.

D. Cost and Generalization

Figure 5 shows the evolution of cost (left), the ratio of the difference between collective and individual potentiality to cost (middle), and generalization accuracy (right). Figure 5(a) and (b) present the results obtained using forward and backward potentiality values under the conventional method. Although some differences can be observed, the overall trends for forward and backward computations are similar. In particular, the cost increases rapidly, while the corresponding ratio decreases. In contrast, when both forward and backward learning are applied, as shown in Figure 5(c) and (d), clearer characteristics emerge. Under backward computation, cost decreases substantially, while the ratio of the difference between collective and individual potentiality to cost increases sharply toward the end of learning. As a consequence, generalization accuracy improves, particularly in the later stages of training, although some fluctuations are observed.

These results demonstrate that when both forward and backward methods are used, backward learning dominates the competition and significantly increases the potentiality difference per unit cost. In terms of information in the ordinary sense, this implies that information efficiency per cost is substantially improved, leading to enhanced generalization performance.

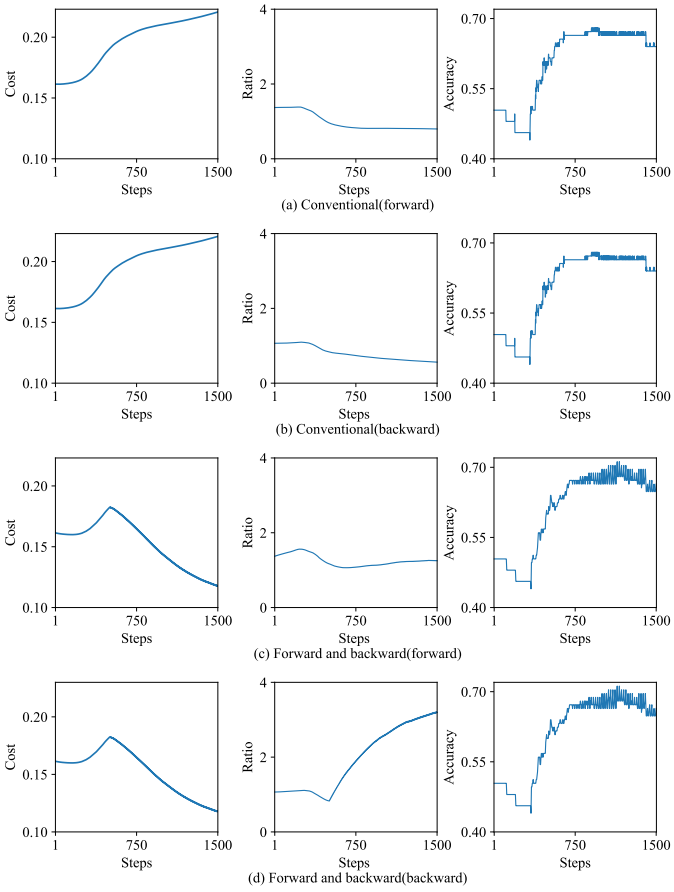

Figure 5. Cost (left), ratio of contradiction or difference potentiality to cost (middle), and generalization accuracy (right) as a function of the number of learning steps by the conventional method with the forward computation (a) and backward computation (b), and by the forward and backward simplification with the forward computation (c) and backward computation (d).

E. Interpreting Compressed Weights

Figure 6(a) shows the correlation coefficients between the weights of compressed networks and the original correlation coefficients computed from the original data set using the conventional method. The correlation coefficient initially drops

to a low negative value, then rapidly increases to a high level close to 0.9. In the final stage, this correlation decreases slightly. Figure 6(b) shows the correlation coefficients obtained using both forward and backward methods. Once the correlation increases, it remains high until the end of learning.

These results indicate that the forward and backward methods attempt to produce final weight configurations that closely approximate the correlation structure of the original data set. In other words, forward and backward simplification tends to extract the simplest forms consistent with the correlation coefficients inherent in the data.

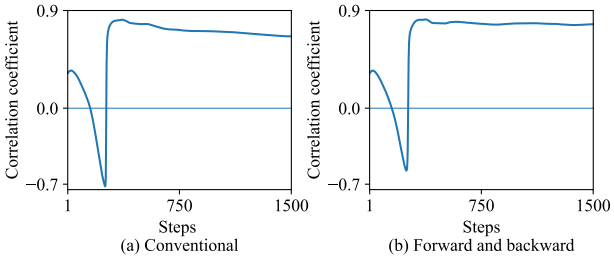

Figure 6. Correlation coefficients between the original correlation coefficients computed from the data sets and the connection weights of the compressed networks obtained using the conventional method (a) and the forward and backward method (b).

Figure 7(a) shows the actual and normalized correlation coefficients computed from the original data set. The sixth input (approximate entropy) and the first input (type/token ratio) exhibit higher absolute strengths. This indicates that lexical diversity, as measured by the type/token ratio, is high, while syllabic regularity is strongly imposed, since lower entropy corresponds to higher regularity. Thus, the correlation coefficients suggest a combination of high average diversity and strong regularity.

Figure 7(b) shows the regression coefficients obtained using logistic regression. Input No. 3 (conditional word entropy) has a negative coefficient, indicating higher regularity associated with lower entropy. Inputs No. 4 (average PMI) and No. 6 (approximate entropy for syllabic forms) show relatively large absolute values, reflecting local and global connectivity among words. Lower average PMI indicates weaker local word associations, while lower approximate entropy indicates stronger overall regularity. Although these results appear partially contradictory, the dominance of conditional word entropy suggests that logistic regression primarily captures global regularity rather than character-level or local constraints.

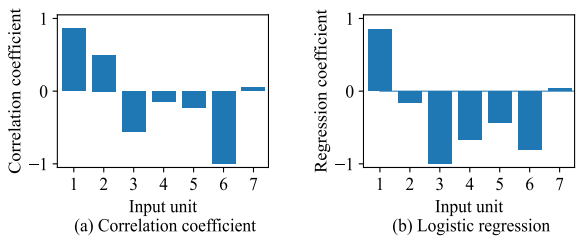

Figure 7. Correlation coefficients between inputs and targets (a) and regression coefficients (b).

Figure 8 shows the signed potentialities or normalized connection weights obtained using the conventional method and the forward and backward methods. Under the conventional method, shown in Figure 8(a), input No. 1 (type/token ratio) has the largest weight, while input No. 5 (character regularity) also exhibits substantial strength due to lower entropy. However, the dominance of input No. 1 indicates that lexical diversity plays a more important role than character-level regularity. In contrast, under forward and backward simplification with the best generalization performance (1126 learning steps in Table I), input No. 1 becomes slightly weaker, while input No. 5 (character bigram entropy) becomes stronger and negative, indicating increased predictability at the character level. Input No. 6 (approximate syllabic entropy) also shows increased negative strength, although it remains weaker than character bigram entropy. These results suggest that verse weakens lexical diversity in order to emphasize character-level regularity. In this sense, competition between diversity and regularity favors regularity in verse texts. Compared with logistic regression, which emphasizes word-level and syllabic regularity, the forward and backward simplification places greater emphasis on character-level regularity, while still incorporating syllabic constraints.

Considering the highest generalization performance achieved by forward and backward simplification, verse texts appear to balance diversity and regularity by shifting the level of emphasis. Specifically, diversity is expressed at the word level, whereas regularity is enforced at the lower character level.

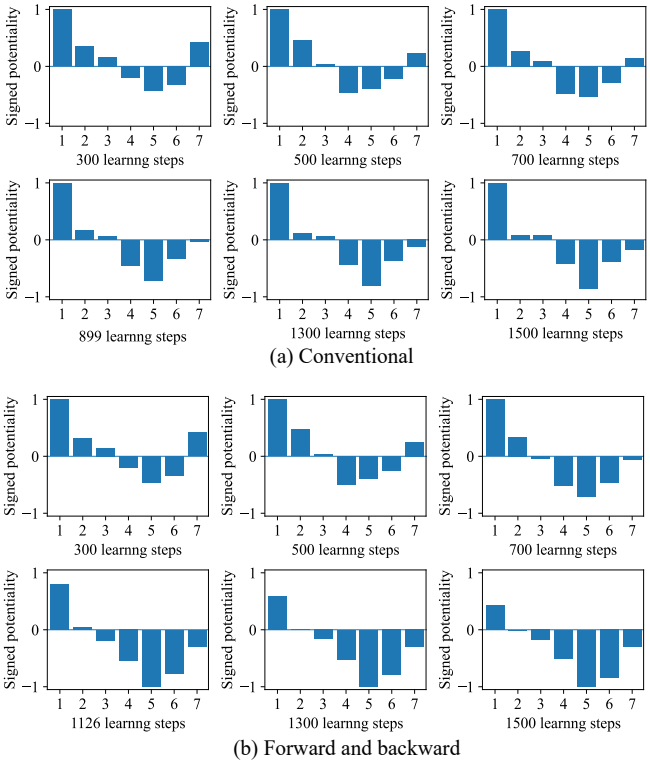

Figure 8. Weights of compressed networks obtained using the conventional method (a) and the forward and backward method (b).

V. CONCLUSION AND FUTURE WORKS

The present study demonstrates the necessity of economical and compact neural networks for complex data sets. In the proposed compact neural network framework, components and learning procedures arising from paradigmatic simplification are projected onto syntagmatic or combinative simplification. This formulation enables multi-level information processing by allowing different components to maximize or minimize information (potentiality maximization and minimization).

We applied this multi-level information processing framework to artistic (poetic) texts, which are regarded as secondary modeling systems formed by imposing strong structural constraints on natural language, while still attempting to store and communicate large amounts of information within compact forms. Our results show that verse texts can be identified by increased regularity at the character level, while simultaneously maintaining lexical diversity at the word level. These findings demonstrate the successful extraction of multi-level processing characteristics in secondary modeling systems and highlight the relevance of artistic texts as models for compact and efficient information processing.

Nevertheless, several challenges remain before the proposed multi-level neural network framework can be considered a general solution for compact information processing. First, the number of levels should be increased. For experimental simplicity, we restricted the framework to two levels—paradigmatic and syntagmatic simplification—but introducing additional levels may enable richer information representation and transmission. Second, more complex experimental settings are required. In this study, we limited the task to distinguishing verse and prose texts by a single French poet and restricted the number of input variables to seven. Future work should extend the analysis to larger and more diverse corpora of artistic texts.

Although the scope of the present study is limited, the results demonstrate the effectiveness of multi-level information processing in increasing information capacity even in compact networks. We regard this work as an initial step toward truly compact and economical information processing systems that may serve as alternatives to massively large models.

REFERENCES

[1] J. M. Lotman, *The Structure of the Artistic Text*, trans. by G. Lenhoff and R. Vroon, ser. Michigan Slavic Contributions 7. Ann Arbor: University of Michigan, 1977, ISBN: 978-0930042158.

[2] R. Kamimura, "Self-competitive simplification: Competition between forward and backward simplification in multi-layered neural networks," in *Proceeding of COGNITIVE2026*, 2026.

[3] R. Jakobson, "Closing statement: Linguistics and poetics," in *Style in Language*, T. A. Sebeok, Ed., Cambridge, MA: The M.I.T. Press, 1960, pp. 350–377.

[4] R. Goodrich, "On poetic function: Jakobson's revised'prague'thesis," *Literature & Aesthetics*, vol. 7, 1997.

[5] A. Chariton, N. Passalis, N. Pleros, and A. Tefas, "Mutual information-based neural network distillation for improving photonic neural network training," *Neural Processing Letters*, vol. 55, no. 7, pp. 8589–8604, 2023.

[6] R. Linsker, "Self-organization in a perceptual network," *Computer*, vol. 21, no. 3, pp. 105–117, 1988.

[7] N. Tishby and N. Zaslavsky, "Deep learning and the information bottleneck principle," in *2015 IEEE Information Theory Workshop (ITW)*, IEEE, 2015, pp. 1–5.