

Evolving Lattices for Analyzing Behavioral Dynamics of Characters in Literary Text

Eugene S. Kitamura¹, Yukio-Pegio Gunji

¹Department of Earth and Planetary Sciences, Graduate School of Science, Kobe University, Rokkodaicho 1-1, Nada, Kobe, 657-8501, Japan
kitamura@stu.kobe-u.ac.jp

Abstract: This paper is about an application of rough set derived lattices in order to analyze the dynamics of literary text. Due to the double approximation nature of rough set theory, a pseudo-closure obtained from two different equivalence relations allows us to form arbitrary lattices. Moreover, such double approximations with different equivalence relations permit us to obtain lattice fixed points based on two interpretations. The two interpretations used for literary text analysis are subjects and their attributes. The attributes chosen for this application are verbs. The progression of a story is defined by the sequence of verbs (or event occurrences). By fixing a window size and sliding the window down the story steps, we obtain a lattice representing the relationship between subjects and their attributes within that window frame. The resulting lattice provides information such as complementarity (lattice complement existence rate) and distributivity (lattice complement possession rate). These measurements depend on the overlap and the lack of overlap among the attributes of characters. As the story develops and new character and attributes are provided as the source of lattices, one can observe its evolution. In fact, a dramatic change in the behavior dynamics in a scene is reflected in the particular shifts in the character-attribute relationship. This method lets us quantify the developments of character behavioral dynamics in a story.

Keywords: Literary text analysis, rough set theory, lattice theory, double indiscernibility, character dynamics, knowledge management



About the Author

Eugene S. Kitamura

holds a B.Sc. in Physics from James Madison University, U.S.A.. He is currently a doctoral student in the Earth and Planetary Science Department at Kobe University, Japan.

1. Introduction

Mathematical characterization or analysis of literary text and theater was active around the 1970's, Poetica matematica by Solomon Marcus (1970) having a considerable influence. Many mathematical methods were applied to text, such as finite-state automata for plot structure analysis (Kahn, 1973), game theory for capturing character relations (Steriadi-Bogdan, 1977; Dinu, 1977), combinatorics for dynamics of characters in drama (Mihnea, 1977), etc. Although successful in quantifying literary characteristics, these methods typically provided information per theatrical scene segment. Additionally, the analysis may depend on subjective categorization of situations or character intents/motives based on semantic interpretations.

Here we present a method that measures the dynamics of a character's actions with respect to the other characters within a scene as the plot unfolds, based on objective information in the drama. Our method relies on subject-action relation of characters in a literary text. These two groups or equivalence classes, the characters and their actions, are used to construct Boolean as well as non-Boolean lattices that correspond to the occurrences of events in the story. Complementarity and distributivity were used to numerically measure these adaptable lattices.

The lattices use fixed points drawn from the information of two equivalence classes as their elements (Gunji and Haruna, 2010). Rough set theory (Pawlak, 1981; Pawlak, 1982) is used to process the relation between the two equivalence classes and lattices are constructed based on the inclusion relation of the fixed point elements (Birkhoff, 1967; Davey, 2002). This allows us to obtain arbitrary lattices, where the relationship between the two equivalence classes may be represented quantitatively. We chose subjects (characters) and their verb attributes as the two equivalence classes and applied this method to analyzing literary text.

2. Methods and Procedures

2.1. Rough Set Theory to Lattice Theory

Rough set is an approximation method to estimate a target by its upper approximation $R^+(X)=\{x \in U|[x]_R \cap X \neq \phi\}$ and lower approximation $R_-(X)=\{x \in U|[x]_R \subseteq X\}$, where X is the target set, U is a universal set, x is an element in U , and $[x]_R=\{y \in U|xRy\}$ is an equivalence class with R being the equivalence relation between the elements x and y . The elements in a set are assumed to be indistinguishable.

A Galois connection is known to construct a complete lattice in a theory of partially ordered sets. We get a Galois connection when a pair of maps (F, G) with $F:P \rightarrow Q$ and $G:Q \rightarrow P$ is formed, where P and Q are two partially ordered sets. Thus, $F(x) \leq y \Leftrightarrow x \leq G(y)$ for any $x \in P$ and $y \in Q$. Then a closure operator $C:=FG:P \rightarrow P$ can be constructed from a Galois connection such that, for any $x, y \in P$, (i) $x \leq C(x)$; (ii) $x \leq y \Rightarrow C(x) \leq C(y)$; (iii) $CC(x)=C(x)$. This results in a complete lattice $L_C=\{x \in P|C(x)=x\}$.

In terms of rough set, a Galois connection is formed with $R^+:\mathcal{P}(U) \rightarrow \mathcal{P}(U)$ and $R_-:\mathcal{P}(U) \rightarrow \mathcal{P}(U)$, where $\mathcal{P}(U)$ is the power set of U . For any $X, Y \subseteq U$, $R^+(X) \leq Y \Leftrightarrow X \leq R_-(Y)$. When $C=R \cdot R^+$ is defined as a closure operator, this also results in a complete lattice $L_C=\{X \in U|C(X)=X\}$.

If, on the other hand, two unrelated binary relations R and S are applied to a universal set U , a pair of operations S^* and R^* (or S^* and R^*) do not form a Galois connection. When there is an operator $T=R \cdot S^*$, T is not a closure operator but is called a pseudo-closure operator, since it satisfies only (ii) $X \subseteq Y \Rightarrow T(X) \subseteq T(Y)$ and (iii) $TT(X)=T(X)$ for $X, Y \subseteq U$. Unlike the previous closure operators, the lattice constructed with fixed points satisfying $L_T=\{X \in U|T(X)=X\}$ with pseudo-closure is not a set lattice. This means that any arbitrary lattice can be constructed with L_T by choosing the right equivalence relations S and R .

2.2. Rough Set Derived Lattices

In a lattice $\langle L_T, \subseteq \rangle$ with $L_T=\{X \subseteq U|T(X)=X\}$ and $T=R \cdot S^*$, an element of L_T is a subset of the universal set, and order is defined by

inclusion \subseteq . In general, $L_T \neq \mathcal{P}(U)$, thus join and meet are defined by $X \wedge Y = T(X \cap Y)$, $X \vee Y = T(X \cup Y)$ for any $X, Y \in L_T$. Only if all subsets of U are collected, L_T forms a power set which is a set lattice, where join and meet are defined by union \cup and intersection \cap . For example, when a lattice is constructed of fixed points $R \cdot (R^+(X)) = X$ from one equivalence relation R , we only get a set lattice.

A set lattice has two important properties in lattice theory such as complementarity (for any $X \subseteq U$, there exists $Y \subseteq U$ such that $X \vee Y = U$, $X \wedge Y = \phi$) and distributivity ($A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$ for $A, B, C \subseteq U$). In a lattice L_T , U and ϕ are the greatest and the least elements. Lattices with both complemental and distributive property are called Boolean lattices.

When a lattice is constructed of fixed points $R \cdot (S^+(X)) = X$ with two equivalence relations R and S , the resulting lattice can also be a non-Boolean lattice as well as a Boolean lattice. This is a result of two equivalence classes fully or partially overlapping each other.

The difference between a lattice of $L_C = \{X \subseteq U | C(X) = X\}$ and $L_T = \{X \subseteq U | T(X) = X\}$ with $C = R \cdot R^+$, $T = R \cdot S^+$ is illustrated in Figure 1. In the case of L_C , when we denote $W = \{[a]_R, [b]_R, [d]_R\} = \{\{a\}, \{b, c\}, \{d, e\}\}$, $L_C = \mathcal{P}(W)$, where all possible combinations of equivalence class of R are obtained (Figure 1 (a)). By contrast, in L_T , although elements of L_T are possible unions of equivalence class R , some elements are missing (Figure 1 (b)). For example, let $X = \{a, b, c\}$ in interpretation R . Then $R \cdot S^+(\{a, b, c\}) = R \cdot ([a]_S \cup [c]_S) = R \cdot (\{a, b, c, d\}) = [a]_R \cup [b]_R = \{a, b, c\}$. Thus $\{a, b, c\}$ is an element of L_T . On the other hand, let $X = \{a, b, e\}$ in interpretation R . Since $R \cdot S^+(\{a, d, e\}) = R \cdot ([a]_S \cup [c]_S \cup [e]_S) = R \cdot (U) = U \neq \{a, d, e\}$, hence $\{a, d, e\}$ is missing in L_T . Due to the loss of information, an obtained lattice can be constructed as a non-Boolean lattice. The lattice construction procedure is explained in more detail in the following section.

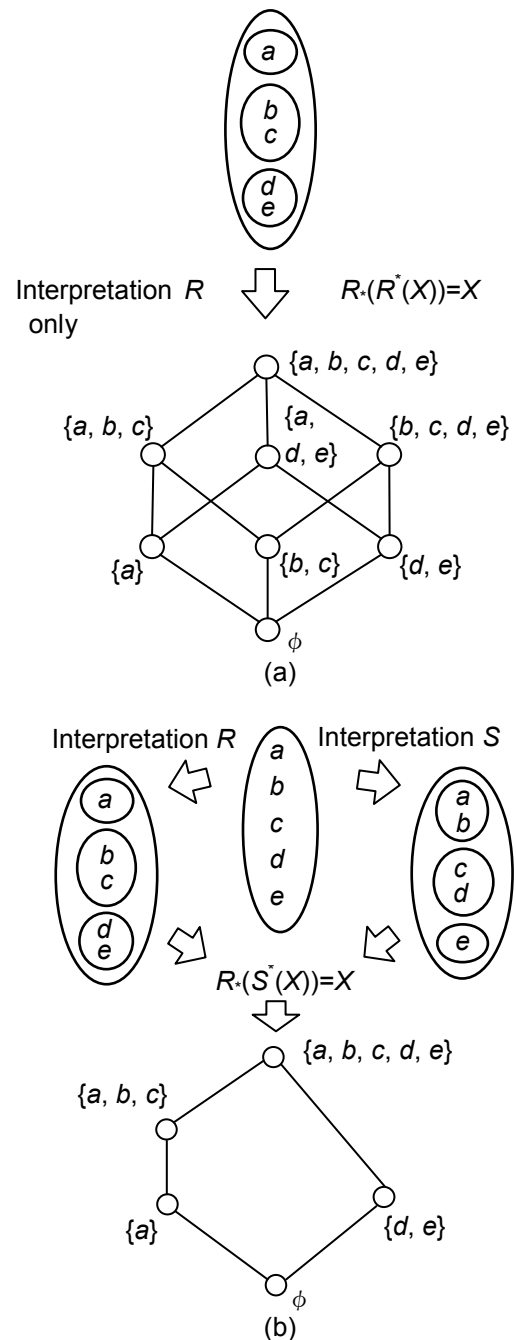


Figure 1: (a) A lattice created from one equivalence relation. It results only in Boolean lattices. (b) A lattice created from two equivalence relations. It results in Boolean as well as non-Boolean lattices.

2.3. Example for Constructing Rough Set Derived Lattice

To construct a rough set derived lattice from two equivalence relations, we need two criteria or interpretations, R and S . Each fixed point X

considered is an equivalence class, for example $\{a\}$ or $\{b, c\}$ or $\{d, e\}$ in Figure 1 (b) for the interpretation R . Treating each equivalence class as a unit, we consider its power set: ϕ , $\{a\}$, $\{b, c\}$, $\{d, e\}$, $\{a, b, c\}$, $\{b, c, d, e\}$, $\{a, d, e\}$, and $U=\{a, b, c, d, e\}$. Each equivalence class and its power set composition are used as an X , and operators S^* and R^* are applied in this order. When applying S^* to the X of the interpretation group R , one must take the upper approximation of X in terms of interpretation S . For example, take $\{a\}$ of interpretation R and apply S^* . We get $S^*(X)=\{a, b\}$. Next, apply R^* to $S^*(X)$. $R^*(\{a, b\})$ is $\{a\}$ in interpretation R , since only the equivalence class $\{a\}$ is included completely within the elements $\{a, b\}$. We started with $X=\{a\}$ and we get $R^*S^*(X)=\{a\}$. Therefore, the equivalence class $\{a\}$ is considered a fixed point. As another example, if we take $\{b, c\}$ in terms of interpretation R and follow the same procedure, we get $R^*S^*(X)=\{a, b, c\}$. Therefore, the equivalence class $\{b, c\}$ is not a fixed point. Repeat this process for all element sets in the power set, from ϕ to U . The empty set ϕ and the universal set U will always be a fixed point. Then collect the fixed points and use them as elements to build a lattice based on their inclusion relationship, with the universal set at the top and the empty set ϕ at the bottom.

2.4. Relation Table for Organizing the Equivalence Relation

The relationship between the two equivalence relations R and S are organized in a relation table shown in Table 1 (a) and (b). Such tables are helpful in identifying fixed points for rough set derived lattices. The example shown here uses the two equivalence relations R and S in Fig. 1 (b). Table 1 (a) shows the relationship between the elements of the equivalence relations R and S . The 1's denote the presence of a relationship and the 0's denote the lack of a relationship. Table 1 (b) shows a simplified version from Table 1 (a) by uniting repetitive rows and columns. Here, the simplification happens to result in a grouping of the equivalence classes.

Table 1: (a) A relation table showing the relations between R and S according to the elements. (b) A simplified version of (a). Unifying repetitive rows or columns gives the same results.

(a)

		S				
		a	b	c	d	e
R	a	1	1	0	0	0
	b	1	1	1	1	0
	c	1	1	1	1	0
	d	0	0	1	1	1
	e	0	0	1	1	1

(b)

		S		
		{a, b}	{c, d}	{e}
R	{a}	1	0	0
	{b, c}	1	1	0
	{d, e}	0	1	1

2.5. Application of Rough Set Derived Lattices onto Text Analysis

We used the modern translation of *Romeo and Juliet* by Shakespeare (Yates-Glandorf (Ed.), (1985)), Act 1, Scene 1. The subjects used were the drama characters. The attributes used were the verbs explicitly written as a narration and the lines by each character (the act of *speak*). Other actions implied in the conversations but not explicitly stated as a narration were not used to minimize subjective interpretations of the text. This reduced the attributes down to *enter*, *speak*, *strike down their swords*, *fight*, *stay*, *pause*, and *exit*. The ending of the action *fight* was chosen to be during the act of *speak* by the prince (Table 2 (a)). The attribute *pause* by Romeo in the modern translation acknowledges the verbal zero-signs (Teodorescu-Brinzeu, 1984). Additionally, the attribute "no reference" was used to label subjects that were present on stage but were not assigned actions within the considered window size explained below. To simplify the information processing, repeated conversations between the same characters

(for example, conversations between Sampson and Gregory in the opening and between Benvolio and Romeo toward the ending of Scene 1) were abbreviated. This reduced the number of actions in the story (performed lines and narrations) from about 110 lines to about 80 lines.

For our analysis we took a span of 16 action lines or steps for the source window and incremented this fixed window size downward step by step. The window span of 16 steps was chosen to roughly include more than three subjects and three attributes in order to obtain an informative lattice. The results obtained from the first 15 steps and the last 15 steps will not be a candidate for consideration due to the incremental increase and decrease in the window size (from size 1 to 15 or 15 to 1) from step 1 to 15 and step 62 to 77 in Figure 2.

The types of resulting lattices were quantitatively measured by their complementarity and distributivity. Complementarity is the complement existence

rate. It is the ratio between the number of elements with complements and the number of all elements in a lattice. If there are elements without complements, the complementarity will be less than 1. Distributivity is the complement possession rate. It is the ratio between the total number of complements and the number of elements with complements. If there are elements with multiple complements such as N_5 or M_3 , the distributivity will be greater than 1.

3. Results

We obtained a time evolution of lattices as the story progressed (Figure 2). The complementarity and distributivity were plotted as the number of story steps advanced from 1 to 77. Along with the complementarity and distributivity, the number of characters and attributes (verbs) in the lattices were also plotted, as well as the appearance duration of each character.

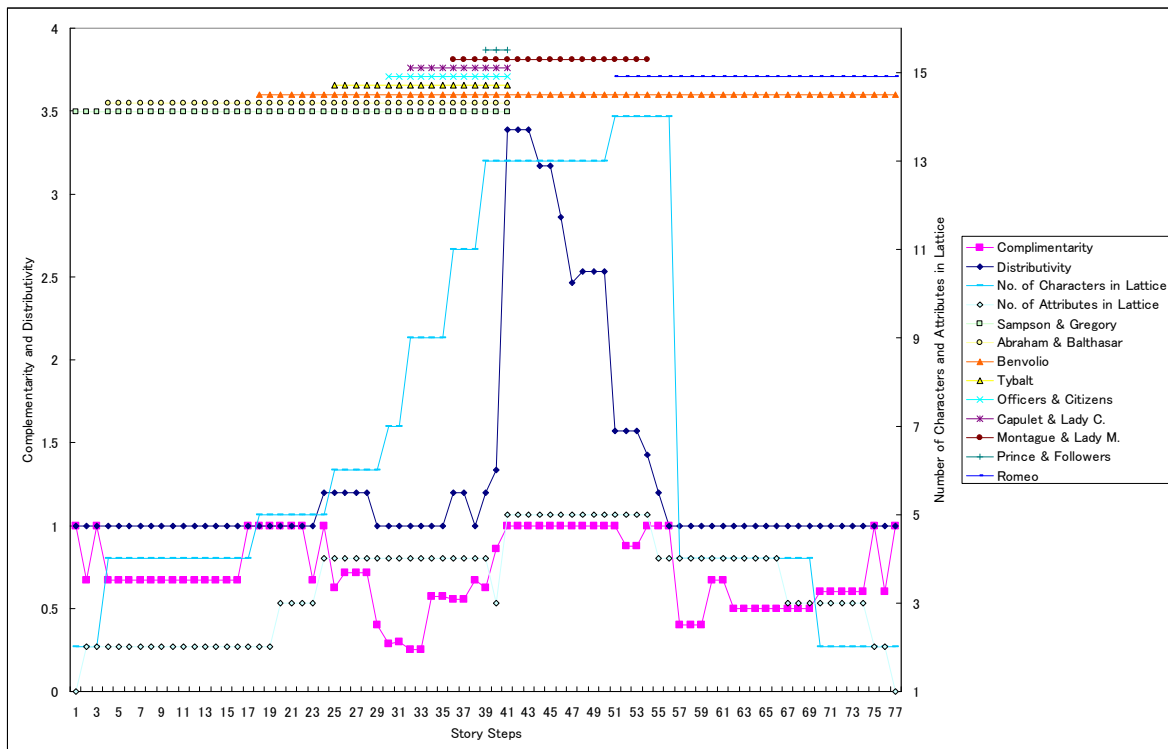


Figure 2: A plot of complementarity, distributivity, the number of characters and attributes in lattice, and the appearance duration of characters on stage.

The most notable feature of the graph is the sharp jump of the distributivity that coincides with the “All leave except Montague, Lady

Montague, and Benvolio” at story step 41. The distributivity increases from 1.33 to 3.38. Another notable feature is the recurring minor

dips in the complementarity. Most of these dips correspond to the entry of characters on stage (story step 25, 30, 32, 36, 39).

From step 17 to 22 and at step 56 the value of C and D are both 1.0. This is the case when the lattices are Boolean lattices.

The number of characters and attributes in lattices drags longer than the actual presence of characters on stage. This is due to the length of window size (16), where a character or its attribute would not exit the lattice unless the window passes over their attributes. Therefore the presence of characters and attributes in lattices would be prolonged compared to their actual appearances.

4. Discussion

There is a great contrast in the distributivity between step 40 and 41 as mentioned earlier. Until Prince Escalus arrives at step 39, into a scene of a gradually escalated fight, the attributes used to construct the lattices were *enter*, *speak*, and *fight*. After the prince 'spoke' his words, two more attributes are added at step 41, *leave* (or *exit*) and *stay*. The resulting lattices drastically change, raising the distributivity value. The change is due to the shift in common and non-common attribute patterns of the characters. Until step 40, the attribute overlap and differences tend to be clustered (Table 2 (a)). However, the attributes *leave* and *stay* induces cavities in the clusters, disrupting its uniform structure (Table 2 (b)). Granted, one may argue that the rise in distributivity is due to the mutually exclusive attributes of *leave* and *stay*, which is a result of faithfully extracting the literal information written in the script. We have confirmed that even if we exclude the attribute *stay*, which is seemingly redundant since Montague, Lady Montague, and Benvolio are already on stage and continue to remain there, the distributivity still resulted to be the highest value in the graph (D=2.60).

Even though lattices are constructed based on information of subjects and their attributes, generally there is not necessarily a direct correspondence between the resulting lattices and the number of characters and attributes. It is the similarity and dissimilarity of relationships among the attributes of characters that determine the resulting lattices.

At step 41, although the attribute decreases one unit and increases two units and the number of characters rises steadily between story steps 18 to 39, the distributivity does not change in a like manner. It is the shift in the commonality and differences in the attributes that affected the outcome of lattices. The distributivity in Figure 2 may suggest that the author had prepared a dramatic scene by increasing the number of characters and when the potential for tension is at its highest, the attribute relations are shifted to release the tension. As the window of consideration shifts and old attributes exit the window, the distributivity value eventually returns to where it used to be.

Table 2: (a) A relation table at step 40. (b) A relation table at step 41. (Abbreviations are S=Sampson, G=Gregory, A=Abraham, BI=Balthasar, Bn= Benvolio, T= Tybalt, O=officers & citizens, C=Capulet, LC=Lady Capulet, M=Montague, LM=Lady Montague, Pr=Prince Escalus, Pf=Prince's followers.)

(a)

		Actions		
		<i>enter</i>	<i>speak</i>	<i>fight</i>
Subjects	S, G, A, BI	0	0	1
	Benvolio	0	1	1
	Tybalt	1	1	1
	O, C, LC, M, LM, Pr	1	1	0
	Pf	1	0	0

(b)

		Actions				
		<i>enter</i>	<i>speak</i>	<i>fight</i>	<i>leave</i>	<i>stay</i>
Subjects	S, G, A, BI	0	0	1	1	0
	Bn	0	1	1	0	1
	T	0	1	1	1	0
	O, C, LC, Pr	1	1	0	1	0
	M, LM	1	1	0	0	1
	Pf	1	0	0	1	0

As far as the minor dips in the complementarity values are concerned, most of them are due to the appearance of new characters. Newly appearing characters typically provide only the attribute *enter*. This attribute usually already existed in the relation table. Consequently a new character would be included in the attributes of another character. Therefore, the new character would not have a complement of its own, reducing the value of complementarity. Generally, complementarity decreases when there is a character whose attributes completely include the attributes of other characters (Table 2 (a)).

5. Conclusion

We have used a rough set derived lattice to represent the logical relationship between two equivalence classes. This tool was applied to literary text to capture the relations between the characters and their actions as the story progressed. The transition of the complementarity and distributivity of the resulting lattices reveals particular shifts in the relations of the attributes (actions) of the characters, also seen in the relation tables. Such shifts of characters' attributes may be suggestive of a writer's technique for assigning character attributes to manage the suspense levels and drama of the stories. This method can be useful in gauging the complexity in the subject-activity as a story progresses.

References

- Birkhoff, G. (1967). *Lattice Theory*. Coll. Publ., XXV, Providence, American Mathematical Society.
- Davey, B.A., and Priestley, H.A. (2002). *Introduction to Lattice and Order* (2nd ed.). Cambridge, Cambridge University Press.
- Dinu, M. (1977). How to Estimate the Weight of Stage Relations. *Poetics*, 6 (3-4), 209-227.
- Gunji, Y.-P., and Haruna, T. (2010). A Non-Boolean Lattice Derived by Double Indiscernibility. *Transactions on Rough Sets XII, LNCS*, 6190, 211-225.
- Kahn, E. (1973). Finite-State Models of Plot Complexity. *Poetics*, 3(1), 5-20.
- Marcus, S. (1970). *Poetica matematica*. Bucurcsti: Editura Academiei R.S.R.
- Mihnea, T. (1977). Combinatorics and dynamics of characters in drama. *Poetics*, 6 (3-4), 229-254.
- Pawlak, Z. (1981). Information systems-theoretical foundations, *Information Systems*, 6, 205-218.
- Pawlak, Z. (1982). Rough Sets. *International Journal of Computer and Information Science*, (11), 341-356.
- Steriadi-Bogdan, M. (1977). The Evolution of the Plot and Problems of Strategy in a Detective Play. *Poetics*, 6 (3-4), 375-382.
- Yates-Glandorf, J. B. (Ed.) (1985). *Shakespeare's Romeo and Juliet. In the Original and Modern English. A Parallel Text Edition*. Logan: Perfection Learning Corporation.
- Teodorescu-Brinzeu, P. (1984). The Verbal Zero-Sign in Theater. *Poetics* 13 (1-2), 47-56.