Prediction of Evaluation Value from Product Reviews and Explanations to Healthcare Products Using Network

Ayumu Sakaguchi, Shoji Nohara, and Ryosuke Saga

Abstract—**Given that reviews are difficult to score, predicting evaluation values automatically can be useful. The evaluation value depends on users who buy certain products with known information that meet expectations. Numerous studies have already predicted the evaluation value from reviews. However, many of these studies only paid attention to user reviews, and the opinions of product providers were not considered. In this paper, we will comment about the keywords of reviews and explanations, create a co-occurrence network showing keywords, and perform multiple regression analysis using the features of the network. As a result, we can acquire results of high accuracy from multiple regression analysis to predict the value of products based on the reviews and explanations to the service.**

Index Terms—**Service Science, Evaluation Value Prediction, Co-occurrence Network, Multiple Regression Analysis**

I. INTRODUCTION

An increasing number of services are offered on the Internet, and the evaluation of a user is critical to another user. Nowadays, many comments and evaluations are found on E-commence, blogs, or SNS, and the number of users increase or decrease because of these comments. In particular, comments of famous people or strong users of information are easily spread worldwide. In other words, the evaluations of these people have a great influence on the propagation and penetration of the service. Thus, knowing and managing the evaluation of the service of such users is crucial.

The evaluation of service exists in many kinds of forms. On E-commerce sites, such as eBay or Amazon.com, some texts or pictures are carried with the evaluation, including a score. On TripAdvisor, the evaluations are carried without a score. Evaluations are not usually scored on individual sites, such as a blog or SNS, but they appear as comments on text form.

Numerous studies have predicted the evaluation of service from text. These studies considered many reviews as training data and predicted the evaluation value by mining the text of reviews. In terms of predicting evaluation value, many studies used Support Vector Machines or Bayesian network [1,2]. Recently, examples of using sentiment analysis have been reported [3].

In the field of service science, evaluation of service is related to quality of service. The SERVEQUAL model proposed by Perasuraman et al. expresses service quality based on the relation between service provider and service consumer [4]. According to the SERVEQUAL model, a gap exists between service provider and service consumer. Quality of service, that is, evaluation of service is believed to increase by filling this gap.

In this paper, we predict the evaluation value based on the gap between service provider and service consumer. Saga's study expresses a gap between service provider and service consumer [5] and attempts to manage the state of service via network visualization. In this paper, based on Saga's study, we predict evaluation value from the network, which shows the gap between two sides.

This paper is organized as follows. Section 2 explains the related works. Section 3 explains a specific approach to show the gap between two sides. Section 4 describes the results of predicting evaluation value. Section 5 presents the discussion. Section 6 concludes this paper.

II. RELATED WORK

Lee used support vector machines to the corporate credit rating problem [1]. Harvey et al. used a Bayesian latent variable model for rating prediction [2]. Ganu et al. predicted evaluation value by predicting sentiment from contents of review or meta data and creating a regression model [3]. Tang et al. proposed a neural network method for review rating prediction [6]. This approach considers both the text and the author of the text. Long et al. proposed a novel approach to accurately predict feature ratings of products [7]. Their approach selects user reviews that extensively discuss specific features of products, using information distance of reviews on the features. McAuley et al. considered how users respond to new products [8]. Thus, this approach considers review text, so this study aimed to combine latent rating dimensions with latent review topics. Qu et al. considered unigram and n-gram to predict a user's numeric rating [9]. However, unigram and n-gram are known for certain problems. Unigram cannot capture important expressions that are essential for prediction models of rating. Meanwhile, n-gram of words rarely occurs in the training set and fails to yield robust predictors. This approach is proposed to overcome the limitations of these two models. Ghose et al. explored multiple aspects of review text, various measures of readability, and extent of spelling errors for identifying important text-based features to predict the

This research was supported by KAKENHI 25240049 and 25420448.
Ayumu Sakaguchi is with the department of Knowledge and Information Systems, School of Sustainable System Sciences Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka, Japan (e-mail: sxa01106@edu.osakafu-u.ac.jp).
Shoji Nohara and Ryosuke Saga are with the department of the Computer Science and Intelligent Systems, Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka, Japan (sv107029@edu.osakafu-u.ac.jp, saga@cs.osakafu-u.ac.jp).

ISBN: 978-988-14047-3-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

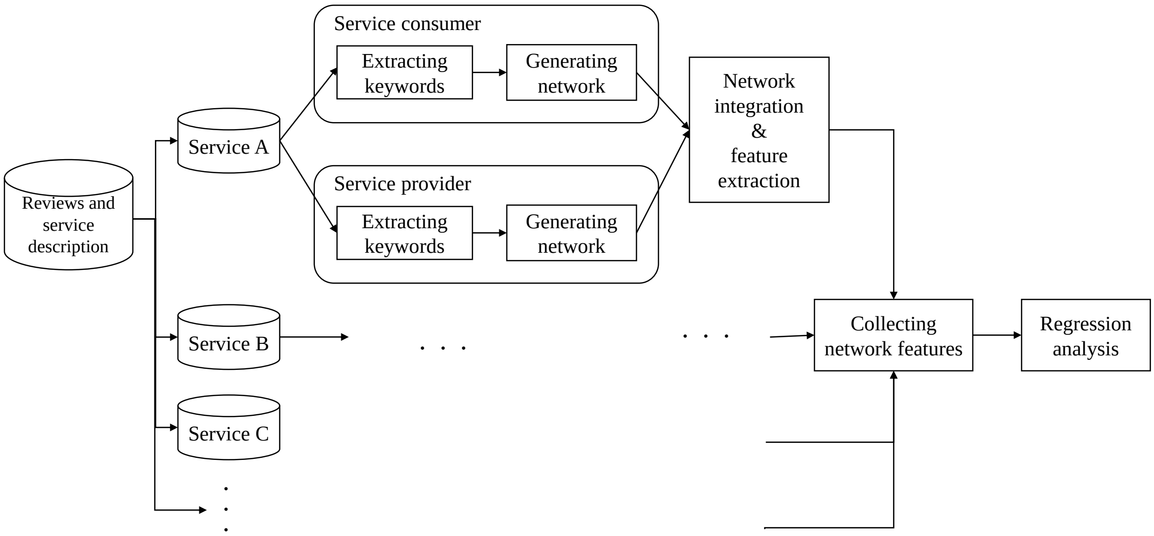

Fig. 1. Process to Predict Evaluation Value

helpfulness and economic effect of product reviews [10]. In addition, this study also examined multiple reviewer-level features such as average usefulness of past reviews and self-disclosed identity measures of reviews that are displayed next to a review. This study accurately predicted the influence of reviews on sales and perceived usefulness using random forest. Linshin used topic model to predict evaluation value of Yelp [11]. He proposed an approximation of a modified latent Dirichret allocation (LDA) in which term distributions of topics are conditional on star ratings. Albornoz et al. believed that evaluation value depends on users' feelings about product features, and they predicted evaluation value based on a user's opinion about different product features [12]. The salience and values of different product features to measure a user's opinion have been used to contrast a Vector of Feature Intensities representing the review and applied for machine learning model, which classifies reviews to different evaluation categories. Gapta et al. used regression and a classification modeling problem, as well as explored several combinations of syntactic and semantic features to predict ratings [13].

All related studies have not considered the gap between service provider and service consumer to predict evaluation value. Therefore, this paper will predict evaluation value using the gap between two sides.

III. Approach to Generate the Network and Extraction of Feature Quantity

Fig. 1 shows the process used in this research. We assume that this process uses an Amazon.com dataset. To generate a network expressing gaps between service consumer and service provider, this process generates two networks from text data for each service. For service consumer, the network is generated from review text implying the value of use for service consumers. Meanwhile, the network for service provider is generated from service description. In Amazon.com, this service description is expressed as a product explanation consisting of "about product" and "product description." To generate networks, the process extracts keywords regarded as nodes and relationships regarded as edges. After generating networks, the process integrates two networks into one network and expresses the gap on the network. From the network, the process calculates features, such as diameters and cluster coefficient, and collects the network data. Finally, the process analyzes the collected network features via multi-regression analysis and confirms our hypothesis that evaluation value can be predicted from the features of network with gap information.

A. Dataset: Reviews and Service Description

In this analysis, we utilize the dataset provided by SNAP page [14-16]. We target the Health and Personal Care categories because it is one of the most representative services, that is, the services in this category have IHIP features; in particular, the influence of I (intangibility), H (heterogeneity), and I (inseparability) appear in the service [17]. In addition, we limit the service by having more than 100 reviews. If the number of reviews is small, the strengths of co-occurrence are not discriminated. Therefore, we use the service with more than 100 reviews and acquire a total of 310 services and 51,573 reviews. We show the number of each evaluation value at Fig. 2.

B. Keyword Extraction with Preprocessing

Keyword extraction is a basic process in text mining. Before keyword extraction, morphological analysis is conducted for text data. In this process, we can identify part of speech and filter the unnecessary Pos set. We adopt words that belong to a verb, noun, adjective, adverb, and interjection. Subsequently, we remove stop words that have no meaning and performed stemming. Stemming refers to words that mean the same but look different, such as being in plural form or past tense. For example, "go" and "goes" are integrated with to "go," and they are regarded as the same words. Therefore, any words are seen at a glance, and they become one word. After preprocessing, we extract keywords/. To extract keywords, we use the TF-IDF algorithm as follows,

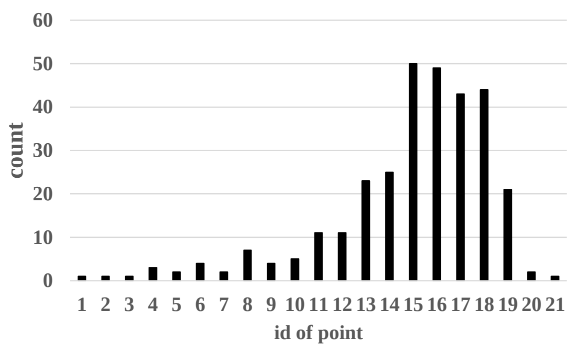

Fig. 2. Number of each evaluation value

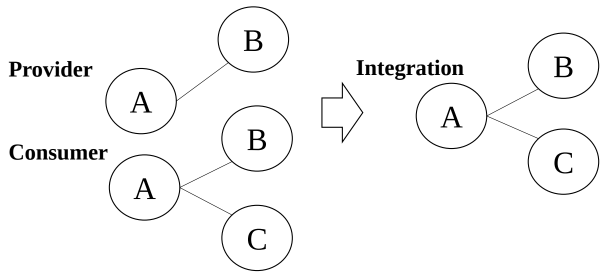

Fig. 3. Integration Two Networks

$$TFIDF_{ij} = TF_{ij} * IDF_i \quad (1)$$

$$TF_{ij} = \frac{v_{ij}}{n_j} \quad (2)$$

$$IDF_i = \log\frac{N}{df_i} + 1 \quad (3)$$

where v_i is the frequency of the appearance of word i in document j ($j \in \boldsymbol{D}$, $\boldsymbol{D}$: document set), n_i is the sum of words in document j, N is the number of documents in a document set, and df_i is the number of documents word i appears in. Note that the document set of service provider consists of "about product" and "product description," so N for service provider is always 2. For a service consumer, N is the number of reviews (≥100).

Finally, we extract keywords for document set using Eq. (4).

$$TFIDF_i = \sum_{j \in D} TFIDF_{ij} \quad (4)$$

Using Eq. (4) and ranking the studies in order based on the value of TFIDF$_i$, we can extract important keywords for document sets, that is, service provider and service consumer. In this study, the top 100 words (including same ranked words) are regarded as keywords and utilized as nodes in the network.

C. Generating Network

We use Jaccard coefficient to extract edges. Jaccard coefficient of keyword s and keyword t is expressed as

$$Jaccard = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

where A is document set including keyword s, B is document set including keyword t, and $|\,|$ is count of elements in the document set. In the case of service provider, we draw the edge with Jaccard coefficient more than 0.5. In the case of service consumer, we draw with Jaccard coefficient more than 0.3.

D. Network Integration and Feature Extraction

Fig. 3 illustrates the integration of two networks of explanation and review of each product to one network. If overlap nodes exist, such as A, A becomes one, A has connections to B, and C is not changed. If there is an edge between overlap nodes, such as A and B, we regard them as the same edge, and edges are overlapped into one. We can acquire co-occurrence showing the gap between service provider and service consumer network by integrating networks.

Subsequently, we extract and calculate features with the generated network. We adopt the following as features: diameter, average of cluster coefficient, average of degree, average of betweenness, modularity, number of components, number of overlap nodes, number of overlap edges, number of reviews, number of nodes, and number of edges [18]. Table I shows examples of information of network features.

E. Multiple Regression Analysis

Multiple regression analysis is an analysis method for extracting strength of causality between a response variable and multiple explanatory variables [19]. This analysis is conducted based on a multiple regression model that consists of k (k=10) explanatory variables and one response variable.

The multiple regression model assumes that there is no influence from a potential variable and calculates the influence only from each observation data such as hardware and genre. The model equation of a service i is

TABLE I. INFORMATION OF NETWORK FEATURES

id	point	diameter	average of cluster coefficient	average of degree	average of betweenness	modularity	number of components	number of overlap node	number of overlap edge	number of reviews	number of nodes	number of edges
1	2.2	3	0.80	109.0	135.1	0.4	3	31	148	142	176	9594
2	2.7	3	0.65	73.6	64.9	0.3	5	7	15	323	112	4120
:	:	:	:	:	:	:	:	:	:	:	:	:
9	3.7	4.25	0.64	78.4	104.9	0.3	1.5	20.7	55.3	1215	128	5250
:	:	:	:	:	:	:	:	:	:	:	:	:
21	4.9	5	0.41	35.9	122.5	0.3	2	12	18	200	111	1992

TABLE II. RESULT OF MULTIPLE REGRESSION ANALYSIS

	Coefficient	Standard error	t-value	p-value
(Intercept)	$3.506*10^{-16}$	$7.315*10^{-2}$	0.000	1.000
Cluster	$-8.214*10^{-1}$	$1.491*10^{-1}$	-5.508	$7.710*10^{-5}$
Overlap node	$-8.676*10^{-1}$	$2.764*10^{-1}$	-3.139	$7.252*10^{-3}$
Number of reviews	$3.406*10^{-1}$	$1.089*10^{-1}$	3.126	$7.432*10^{-3}$
Diameter	$5.568*10^{-1}$	$1.183*10^{-1}$	4.708	$3.360*10^{-4}$
Modularity	$4.411*10^{-1}$	$1.328*10^{-1}$	3.322	$5.038*10^{-3}$
Overlap edge	1.043	$2.649*10^{-1}$	3.936	$1.493*10^{-3}$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i \quad (6)$$

where Y_i shows the response variable that is a variable explained by the explanatory variable, which indicates point in this paper; β_k indicates the partial regression coefficient that shows the degree of causality between response variable and explanatory variable; X_{ik} shows the normalized explanatory variable, which indicates a factor of causal relationship; and ε_i shows residual. This model is evaluated using R^2 called multiple correlation coefficients between Y and Xs, and the stepwise method based on Akaike's Information Criteria (AIC) is used to exclude the influence of the observation data with multicollinearity and choose the better combination of explanatory variables.

In this study, we use the mean of each network feature for each point as an explanatory variable, because the distribution of data is biased to a high rated score over 4.0 and the calculated model may be biased. Therefore, using the mean, which is one of the representative values showing data, the model can decrease the influence of bias.

IV. ANALYSIS RESULT

Table II shows the results of multiple regression analysis and features chosen by stepwise regression on Table I. R^2 was 0.9213 and adjusted R^2 was 0.8876. The p-value was 5.82×10^{-7}. Note that the significant level is less than 0.01, and the features not in Table I were removed because they were not considered useful.

The features regarded as useful were cluster, overlap node, number of reviews, diameter, modularity, and overlap node. The features were separated into positive and negative values.

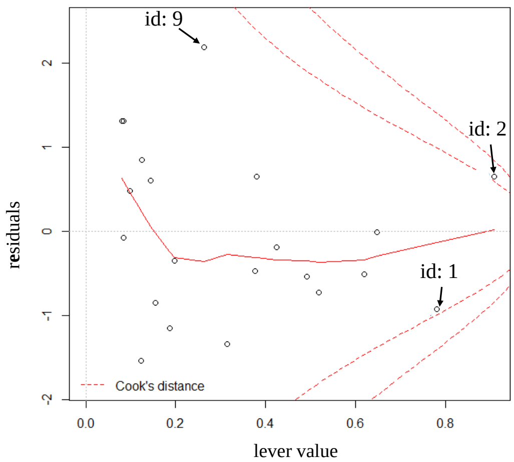

Fig. 4. Cook's distance

The number of reviews, diameter, modularity, and number of overlap edges have positive values. The number of overlap edges is the strongest among the positive ones, but diameter is most important because t-value is the largest. By contrast, the average of cluster coefficient and number of overlap nodes have a negative coefficient, and the former has the most negative value among them.

We show the Cook's distance at Fig. 4 to understand the data that specifically influence the model. At Fig. 4, the horizontal axis is the lever, the vertical axis is the residual, and the dotted line is Cook's distance. The lever value shows how the data apply to the model. If the lever value is high, applying for the model is good. If Cook's distance is more than 0.5, the data can exert an influence. If Cook's distance is more than 1, the data exerts an influence. Id of the number shows data id of feature average for each evaluation value. From this result, the data of id 1 and id 2, namely, features with evaluation value of 2.2 and 2.7 influence multiple regression analyses.

V. DISCUSSION

We can express evaluation value of service using the average of network feature for each evaluation value. Given that overlap node and overlap edge influence evaluation value, we can consider the gap between service provider and service consumer related to evaluation value. Therefore, we can predict evaluation value from the gap between two sides.

If a single topic contains numerous explanations or reviews, its evaluation value is high based on the result of overlap node and overlap edge. Moreover, if several reviews exist and are discussed from various positions, the evaluation value is high because of cluster average, number of reviews, diameter, and modularity. Although keywords exist, the evaluation value is low if different topics are discussed. Therefore, eliminating the gap is critical to obtain a high evaluation value.

However, the evaluation values of the top three values from the most fitted data of explanations and reviews were discussed about the same topic from various positions were 4.1, 4.1, and 3.6. These evaluation values were not high. The evaluation values were not high even though data were fitted to obtain high evaluation values because we considered using feature average for each evaluation value. We used average because the number of data differed for each evaluation value, and bias by the number of data must be eliminated. We could predict evaluation value using the feature average of the gap. However, the real data exhibited errors from the average, so the real data had residuals from the model.

Therefore, we need to increase the number of data with low evaluation value, collect sample data with the same number for each evaluation value, and analyze without average.

VI. Conclusion

In this paper, we predicted evaluation value using network showing the gap between explanation of service provider and service consumer. We used the average of network feature and conducted multiple regression analysis to predict and understand the influence on evaluation value. As a result, we constructed a model consisting of the average of cluster, the number of overlap nodes, number of reviews, diameter, modularity, and number of overlap edges.

Some of the data in this study influenced the results because we used the average of network feature. Thus, we failed to acquire an equation that can be applied for all data. Future studies require further data to improve the equation so that the influence on results is not dependent on a part of data, and the equation must be applicable for all data.

Appendix

We explain network features in detail [18].

- *Diameter*: Diameter is the length of the longest finite geodesic path anywhere in the network.
- *Degree*: Degree of a vertex in a graph is the number of edges connected to it.
- *Cluster Coefficient*: Cluster Coefficient represents the average probability that a pair of i 's friends of one another and is given by (7).

$$C_i = \frac{R_i}{k_i - 1}. \tag{7}$$

where R_i is the mean number of connections from a neighbor of i to other neighbors, and k_i is the degree of vertex i.

- *Betweenness*: The number passing through each vertex is simply proportional to the number of geodesic paths the vertex lines on. This number of geodesic paths is betweenness. We can express the betweenness of vertex i by (8).

$$x_i = \sum_{st} \frac{n_{st}^i}{g_{st}}. \tag{8}$$

where g_{st} is the total number of geodesic paths from s to t, and n_{st}^i is the number of geodesic paths from s to t that pass through i.

- *Modularity*: Modularity is a measure of the extent to which like is connected to like in a network. Modularity is given by (9).

$$Q = \sum_i (e_{ii} - a_i^2). \tag{9}$$

where e_{ii} is the rate of sum of edges for each node in community i to sum of edges, and a_i is the rate of sum of edges that connect from community i to sum of edges.

- *Component*: Component is a subset of vertices of a network such that there exists at least one path from each member of that subset to another member.

References

[1] Y. Lee. (2006, April, 18). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications*. [Online]. 33(1). pp. 67-74. Available: http://www.sciencedirect.com/science/article/pii/S0957417406001205

[2] M. Harvey, M. J. Carman, I. Ruthven and F. Crestani, “Bayesian Latent Variable Models for Collaborative Item Rating Prediction,” in *Proceedings of the 20th ACM international conference on Information and knowledge management*, October. 2011, pp. 699-708.

[3] G. Ganu, N. Elhadad and A. Marian, “Beyond the Stars: Improving Rating Predictions using Review Text Content,” *WebDB*, vol. 9, Jun. 2009, pp. 1-6.

[4] P. Ananthanarayanan, V. A. Zeithaml and L. L. Berry, “Reassessment of expectations as a comparison standard in measuring service quality: implications for further research,” *the Journal of Marketing*, 1994, pp. 111-124.

[5] R. Saga, “Text Comparison Visualization from Two Information Sources by Network Merging and Integration,” *Knowledge, Information and Creativity Support Systems.* Springer International Publishing, 2016, pp557-563.

[6] D. Tang, B. Qin, T. Liu and Y. Yang, “User Modeling with Neural Network for Review Rating Prediction,” in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, July. 2015, pp. 1340-1346.

[7] C. Long, J. Zhang, M. Huang, X. Zhu, M. Li and B. Ma, “Predicting Feature Rating through an Effective Review Selection Approach,” Knowledge and information systems, 2014, 38.2: pp. 419-446.

[8] J. McAuley and J. Leskovec, “Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text,” in *Proceedings of the 7th ACM conference on Recommender systems*, October. 2013, pp. 165-172.

[9] L. Qu, G. Ifrim and G. Weikum, “The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns,” in *Proceedings of the 23rd international Conference on Computational Linguistics, Association for Computational Linguistics*, August. 2010, pp. 913-921.

[10] A. Ghose, “Predicting the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics,” *IEEE Transactions on Knowledge and Data Engineering*, 2011, 23.10: pp. 1498-1512.

[11] J. Linshi. (2014). Personalizing Yelp Star Ratings: a Semantic Topic Modeling Approach. *Yelp Data Challenge*. [Online]. Available: https://www.yelp.com/dataset_challenge

[12] J. C. de Albornoz, L. Plaza, P. Gervas and A. Diaz. "A joint model of feature mining and sentiment analysis for product review rating," in *Proceedings of the European Conference on Information Retrieval*, April. 2011, pp. 55-66.

[13] N. Gupta, G. Di. Fabbrizio, and P. Haffner. "Capturing the stars: predicting ratings for service and product reviews," in *Proceedings of the NAACL HLT 2010 Workshop on Semantic Search*. Association for Computational Linguistics, 2010.

[14] SNAP: Web data: Amazon reviews, https://snap.stanford.edu/data/web-Amazon.html

[15] J. McAuley, C. Targett, J. Shi and A. van den Hengel, “Image-based recommendations on styles and substitutes,” in *Proceedings of SIGIR*, 2015, pp. 43-52.

[16] J. McAuley, R. Pandey and J. Leskovec, “Inferring networks of substitutable and complementary products,” in Proceedings of Knowledge Discovery and Data Minig, 2015, pp. 785-794.

[17] P. P. Maglio, C. A. Kieliszewski and J. C. Spohrer, *Handbook of Service Science*, Springer US, 2010.

[18] M. E. J. Newman, *Networks An Introduction*, Oxford university press, 2010.

[19] G. W. Snedecor and W. G. Cochran, *Statistical Methods*. Blackwell Publishing, 1989, pp. 333-366.

ISBN: 978-988-14047-3-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)