

Which One Is Better, Simple or Complex Metrics?

Fangjun Wu^{1,2,3}

¹School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China

²Jiangxi Key Laboratory of Data and Knowledge Engineering, Jiangxi University of Finance and Economics, Nanchang, China

³High Level Engineering Research Center of Electronic-Commerce, Jiangxi Provincial Colleges and Universities, Nanchang, China

Email: wufangjun@jxufe.edu.cn

Received October 2015

Abstract

At the early stage of software lifecycle, the complexity measurement of UML class diagrams plays an important role in software development, testing and maintenance, and provides guidance for developing high quality software. In order to study which one is better, simple or complex metrics, this paper analyzes and compares four typical metrics of UML class diagrams from experimental software engineering view points. Understandability, analyzability and maintainability were classified and predicted for 27 class diagrams related to a banking system by means of algorithm C5.0 within the famous toolkit Weka. Results suggest that the performance of simple metrics is not inferior to that of complex metrics, in some cases even better than that of complex metrics.

Keywords

Software Measurement, UML Class Diagram, Empirical Validation, Classification

1. Introduction

During the past years, more and more complexity measurement of UML class diagrams have been developed in literatures, which play an important role in software development, testing and maintenance, and provide guidance for developing high quality software. Among these complexity measurements of UML class diagrams, some only focus on counting respective numbers of attributes, methods and relationships among classes [1], thus they are simple; the others are based on entropy-distance [2]-[4], they are relatively complicated. Later researchers carried out some empirical validation works to declaim their advantages [2] [5]-[8]. Despite that simple and complex metrics have their own advantages and disadvantages respectively; it is difficult for users to choose a suitable one in practice.

In order to help user to determine which one is better, simple or complex metrics, this paper analyzes and compares four typical metrics for UML class diagrams from experimental software engineering viewpoints. Understandability, analyzability and maintainability were classified and predicted for 27 class diagrams related to a banking system [1] by means of algorithm C5.0 within the framework of the Weak [9] toolkit.

The remainder of this paper is organized as follows: following the introduction, Section 2 overviews related complexity measurement of UML class diagrams and typical empirical validation works. Understandability, analyzability and maintainability were classified and predicted based on four typical UML class diagram metrics in Section 3. Finally, conclusions are drawn in Section 4.

2. Measuring Complexities of UML Class Diagrams

Till now, there are lots of complexity measurements of UML class diagrams. They can be divided into two groups, namely, simple and complex. One simple metric is Genero's metrics; three complex metrics are Zhou's metric, Yi's metrics, and Wu's metrics respectively.

2.1. Genero's Metrics [1]

Genero believed attributes, methods and relationships among classes all have impact on complexities of UML class diagrams, hence Genero's metrics focus on counting numbers of attributes, methods and relationships among classes and depth of class tree. Genero's metrics are respective numbers of classes, attributes, methods, associations relationships, aggregation relationships, generalization relationships, dependency relationships, classes depended by other classes, classes depend on other classes (abbreviated as *NC*, *NA*, *NM*, *NAssoc*, *NAGg*, *NGen*, *NDep*, *NDepOut* and *NDepIn*, respectively), respective maximum longest path from a class to its root and from a class to its leaves (abbreviated as *MaxDIT* and *MaxHagg*, respectively), respective hierarchy numbers of generalization and aggregation (abbreviated as *NGen* and *NGenH*, respectively).

2.2. Zhou's Metric [2]

Unlike Genero, Zhou claimed that attributes and methods in classes have little impact on complexities of UML class diagrams, but believes that relationships among classes are the key factor. Moreover, he supposed that relationships among classes are random. Firstly, UML class diagrams are transferred into class dependent graphs by Baudry B.'s transformation rules [10]. Secondly, different weight values are specified to various kinds of relationships manually. Finally, complexities of UML class diagrams are measured based on entropy-distance.

2.3. Yi's Metrics [3]

In comparison with Zhou's viewpoint, Yi implied that not only relationship among classes, but also attributes and methods all affect complexities of UML class diagrams, furthermore considers their public, private and protected properties. Yi's metrics are made up of three seed metrics, namely *EDCRC*, *EDCAC* and *EDCMC*. These three seed metrics make up *EDCC*. The above four metrics together are called Yi's metrics. Zhou's metric and *EDCRC* have the same frame work; the latter is a modified version of the former. Yi's metrics are also measured on entropy-distance.

Despite advantages of Zhou's metric and Yi's metrics, no consensus has yet been reached as for what weight values of relationships should be specified. Different views are held by different research. On one hand, some of them regard the weight value of association relationships is smaller than that of aggregation; on the other hand, others believe association and aggregation should have the same weight value.

2.4. Wu's Metrics [4]

To overcome above-mentioned shortcoming in Zhou's metric and Yi's metrics, Wu proposed a novel method to measure complexities of UML class diagrams (abbreviated as *UMLDMCN*) based on data mining and complex networks. Wu's metrics are also made up of three seed metrics, namely *EDCRC*, *EDCAC* and *EDCMC*. These three seed metrics make up *UMLDMCN*. The above four metrics together are called Wu's metrics. Wu's metrics and Yi's metrics have the same framework. *EDCAC* and *EDCMC* of Wu's metrics are exactly the same as those of Yi's metrics respectively. However, *EDCRC* of Wu's metrics is an improved version of that of Yi's metrics. The difference lies in weight values of relationships are automatically computed by virtue of Page Rank algorithm in the former.

2.5. Related Comparative Research of Typical UML Class Diagram Metrics

Several empirical validations were conducted in order to analyze and compare the above metrics systematically

and deeply.

Reference [1] performed some experiments and concluded that: there is statistically significant correlation between some Genero's metrics (namely *NC*, *NA*, *NM*, *NAgg*, *NGen*) and understandability; other Genero's metrics (namely *NC*, *NA*, *NM*, *NGen*) and analyzability; still other Genero's metrics (namely *NC*, *NA*, *NM*) and modifiability. NDep is the only one that has a lesser correlation.

Reference [5] compared Marchesi's, Genero's, In's, Rufai's and Zhou's metrics from different viewpoints, different types of relationships, different types of metric values, complexity, and theoretical & empirical validation. The results showed that the above metrics have their shortcomings while being effective or efficient for some special characteristics of systems.

Reference [6] validated Zhou's metric by using twenty-seven UML class diagrams related to bank information systems as material. The results showed that Zhou's metric is perfectly positively correlated with understandability, analyzability, and modifiability respectively.

Reference [7] compared Marchesi's, Genero's and Yi's metrics both theoretically and experimentally through Internet banking system from different viewpoints, different types of relationships, different types of metric values and complexity. The results showed that the above metrics have their shortcomings while being effective or efficient for some special characteristics of systems.

Reference [8] compared advantages and disadvantages of Genero's and Zhou's metrics through twenty-seven UML class diagrams related to bank information systems. Their understandability, analyzability and maintainability were classified and predicted by means of algorithm C5.0 in tool SPSS Clementine. Results showed that Genero's metrics have higher classification accuracy than that of Zhou's metric.

In short, existing empirical validations pay attention to particular metrics, not a kind of metrics. Once the particular metrics are changed to another one, people still don't know how to choose metrics. This paper groups numerous metrics into two kinds, namely simple and complex metrics. Once we suggest that simple metrics is better than complex metrics or vice versa, there is no confused to choose a particular metric.

3. Comparative Study on Classification and Prediction of Typical UML Class Diagram Metrics

3.1. Dataset

In order to better compare with Reference [6]-[8], this paper also selected twenty-seven UML class diagrams related to bank information systems as object. **Table 1** shows metric values computed by the four kinds of metrics, in which the values of understandability, analyzability and maintainability were determined manually by twenty-four students in third year of computer science in the department of computer science at the university of Castilla-La Mancha in Spain, and twenty-six students in the fourth year of computer science in Italy according to their own experiences. From **Table 1**, we can't determine which kind of metrics is better.

3.2. Classifier

This paper chose algorithm C5.0 within the framework of the toolkit Weka as the classifier. Furthermore, the default parameters of the J48 were adopted, namely -C 0.25-M 2.

3.3. Evaluation Criteria

A large number of evaluation criteria have been used in literature, among which we chose correctly, TP Rate, FP Rate, Precision, Recall, F-Measure and AUC [11] in this paper.

3.4. Experimental Parameters

The ultimate goal of this paper is to compare the performance of simple and complex metrics, 6*3 sets of experiments were conducted. Classification and prediction performance of Genero's metrics, Zhou's metric, EDCC and UMLDMCN were compared. Classification and prediction performance of Genero's, Zhou's, Yi's and Wu's metrics were also compared.

3.5. Results and Discussion

This section provides a detailed report of our experimental results. This paper orders Genero's metrics, Zhou's

Table 1. Metric values.

No.	NC	NA	NM	NAssoc	NAgg	NDep	NGen	NAggH	NGenH	MaxHagg	MaxDIT	NAssocVC	NAggVC	NDepVC	Zhou's metric	Yi's metric			Wu's metric				Understandability	Analyzability	Maintainability	
																EDCRC	EDCAC	EDCMC	EDCC	EDCRC	EDCAC	EDCMC				UMLDMCN
1	2	4	8	1	0	0	0	0	0	0	0	0.5	0	0	0	0	1.23	1.23	0.61	0	1.23	1.23	0.61	1	1	1
2	3	6	12	1	1	0	0	1	0	1	0	0.33	0.33	0	0.67	0.67	1.19	1.12	0.91	0.69	1.19	1.12	0.92	2	2	2
3	4	9	15	1	2	0	0	1	0	2	0	0.25	0.50	0	0.94	0.94	1.17	1.18	1.06	0.92	1.17	1.18	1.05	2	2	2
4	3	7	12	3	0	0	0	0	0	0	0	1	0	0	1.39	1.39	1.19	1.19	1.29	1.39	1.19	1.19	1.29	2	2	2
5	5	14	21	1	3	0	0	2	0	2	0	0.2	0.6	0	0.99	0.99	1.11	1.27	1.09	1.04	1.11	1.27	1.11	2	2	2
6	3	6	12	2	0	0	0	0	0	0	0	0.66	0	0	0.69	0.69	1.19	1.19	0.94	0.69	1.19	1.19	0.94	2	2	2
7	4	8	12	3	0	1	0	0	0	0	0	0.75	0	0.25	1.15	1.15	1.16	1.18	1.16	1.04	1.16	1.18	1.11	2	3	3
8	6	10	14	2	2	0	2	1	1	2	1	0.33	0.33	0	1.21	1.21	1.39	1.89	1.42	1.25	1.39	1.89	1.44	3	3	3
9	3	9	12	1	0	1	0	0	0	0	0	0.33	0	0.33	0.38	0.38	1.14	1.19	0.77	0	1.14	1.19	0.58	2	2	2
10	7	14	20	2	3	0	2	1	1	2	1	0.28	0.42	0	1.27	1.27	1.37	1.75	1.42	1.36	1.37	1.75	1.46	3	3	3
11	9	18	26	2	3	0	4	1	2	3	1	0.22	0.33	0	1.17	1.17	1.34	1.62	1.32	1.23	1.34	1.62	1.35	3	3	3
12	7	18	37	3	3	0	2	1	1	3	1	0.42	0.42	0	1.55	1.55	1.16	1.16	1.36	1.52	1.16	1.16	1.34	3	3	3
13	8	22	35	3	2	1	2	1	1	2	1	0.37	0.25	0.12	1.41	1.41	1.22	1.32	1.34	1.49	1.22	1.32	1.38	3	3	3
14	5	9	26	0	0	0	4	0	1	0	2	0	0	0	0.69	0.69	1.24	1.27	0.97	0.69	1.24	1.27	0.97	2	2	2
15	8	12	30	0	0	0	10	0	1	0	3	0	0	0	1.30	1.30	1.43	1.47	1.38	0.69	1.43	1.47	1.07	2	3	3
16	11	17	38	0	0	0	18	0	1	0	4	0	0	0	1.25	1.25	1.26	1.28	0.66	0.69	1.26	1.28	0.98	4	4	4
17	20	42	76	10	6	2	10	2	3	2	2	0.5	0.3	0.1	1.75	1.75	1.29	1.54	1.60	0	1.29	1.54	0.71	6	6	6
18	23	41	88	10	6	2	16	2	3	4	3	0.43	0.23	0.06	1.80	1.80	1.28	1.47	1.62	1.79	1.28	1.47	1.58	6	6	6
19	21	45	94	6	6	1	20	2	2	4	4	0.28	0.28	0.04	1.95	1.95	1.28	1.47	1.66	1.90	1.28	1.47	1.64	6	5	6
20	29	56	98	12	7	3	24	3	4	4	4	0.41	0.24	0.1	1.89	1.89	1.28	1.47	1.63	2.00	1.28	1.47	1.69	6	6	7
21	9	28	47	1	5	0	2	2	1	4	1	0.11	0.55	0	1.28	1.28	1.15	1.23	1.23	1.24	1.15	1.23	1.21	3	3	3
22	18	30	65	3	5	0	19	1	2	3	4	0.16	0.27	0	1.61	1.61	1.47	1.35	1.53	1.62	1.47	1.35	1.52	5	5	5
23	26	44	79	11	6	0	21	2	5	4	3	0.42	0.23	0	1.78	1.78	1.38	1.57	1.64	1.67	1.38	1.57	1.57	6	6	6
24	17	32	69	1	5	0	19	1	1	2	5	0.05	0.19	0	1.73	1.73	1.34	1.29	1.40	0.92	1.34	1.29	1.12	5	5	5
25	23	50	73	9	7	2	11	3	4	4	1	0.4	0.30	0.08	1.98	1.98	1.24	1.55	1.69	2.14	1.24	1.55	1.76	5	6	5
26	22	42	84	14	4	4	16	2	3	2	3	0.63	0.18	0.18	2.02	2.02	1.26	1.56	1.71	2.27	1.26	1.56	1.84	6	5	6
27	14	34	77	4	9	0	7	2	2	3	4	0.28	0.04	0	2.03	2.03	1.47	1.26	1.70	2.04	1.47	1.26	1.70	4	5	5

metric, Yi's metrics and Wu's metrics from simple to complex. **Table 2** describes their classification and prediction performance.

It can be seen from **Table 2** that all performance indicators (namely correctly, TP Rate, Precision, Recall, F-Measure and AUC) of Genero's metrics for classifying and predicting understandability are the best in those of Genero's metrics, Zhou's metric, Yi's metrics and Wu's metrics. **Table 2** shows that the main performance indicators (namely correctly, TP Rate, Precision, Recall and F-Measure) of Genero's metrics for classifying and

Table 2. Classification and prediction performance

Metrics	Understandability							Analyzability							Maintainability						
	Correctly	TP Rate	FP Rate	Precision	Recall	F-Measure	AUC	Correctly	TP Rate	FP Rate	Precision	Recall	F-Measure	AUC	Correctly	TP Rate	FP Rate	Precision	Recall	F-Measure	AUC
Genero	62.96%	0.63	0.10	0.57	0.63	0.60	0.80	66.67%	0.67	0.10	0.63	0.67	0.65	0.72	55.56%	0.56	0.12	0.52	0.56	0.54	0.68
Zhou	55.56%	0.56	0.15	0.45	0.56	0.49	0.74	51.85%	0.52	0.14	0.47	0.52	0.49	0.77	55.56%	0.56	0.12	0.49	0.56	0.52	0.78
Yi	51.85%	0.52	0.14	0.46	0.52	0.48	0.73	55.56%	0.56	0.13	0.52	0.56	0.53	0.76	62.96%	0.63	0.11	0.57	0.63	0.59	0.75
EDCC	59.26%	0.59	0.13	0.49	0.59	0.53	0.74	51.85%	0.52	0.14	0.47	0.52	0.49	0.76	55.56%	0.56	0.12	0.49	0.56	0.52	0.78
Wu	51.85%	0.52	0.17	0.43	0.52	0.47	0.66	33.33%	0.33	0.19	0.34	0.33	0.33	0.58	37.04%	0.37	0.19	0.33	0.37	0.34	0.60
UMLDMCN	59.26%	0.59	0.15	0.46	0.59	0.52	0.64	29.63%	0.30	0.22	0.25	0.30	0.27	0.64	29.63%	0.30	0.21	0.25	0.30	0.27	0.58

predicting analyzability are the best except AUC indicator in those of Genero's metrics, Zhou's metric, Yi's metrics and Wu's metrics. The above results indicate that the performance of simple metrics is better than that of complex metrics.

It is obvious from **Table 2** that the performance of Genero's metrics for classifying and predicting maintainability is not the best one; however it is not the worst one. The above results indicate that the performance of simple metrics is not inferior to that of some complex metrics.

In a word, experimental results suggest that the performance of simple metrics is not inferior to that of complex metrics, in some cases even better than that of some complex metrics.

4. Conclusion

This paper empirically validated the ability of complexity measurement of UML class diagrams to classify and predicate understandability, analyzability and maintainability. Experimental results showed that the performance of simple metrics is not inferior to that of complex metrics, in some cases even better than that of complex metrics. This observation, as well as confirmed by the experiments reported in previous studies [8], can provide some practical guidance for users to select suitable complexity measurements of UML class diagrams.

Acknowledgements

This work has been partially supported by the Natural Science Foundation of China (Project No. 61163007, 61262010), Natural Science Foundation of Jiangxi (Project No. 20142BAB207010, 20114BAB211019, 20132BAB201036) and Scientific Research Foundation of Jiangxi Provincial Education Department (Project No. GJJ12731, GJJ13305).

References

- [1] Genero, M. (2002) Defining and Validating Metrics for Conceptual Models. Ph.D. Thesis, University of Castilla-La Mancha, Wapedia.
- [2] Zhou, Y.M. (2002) Research on Some Software Measurement Problems. Ph.D. Thesis, Southeast University, Nanjing.
- [3] Yi, T. (2006) Research on UML-Model-Oriented Dependence Analysis and Its Applications. Chinese Science and Technology Press, Hefei.
- [4] Wu, F.J. Measuring Complexities of UML Class Diagrams Based on Data Mining and Complex Networks. *Journal of Computational Information Systems*, Submit.
- [5] Yi, T., Wu, F.J. and Gan, C.Z. (2004) A Comparison of Metrics for UML Class Diagrams. *ACM SIGSOFT Software Engineering Notes*, **29**, 22-27. <http://dx.doi.org/10.1145/1022494.1022523>
- [6] Yi, T. and Wu, F.J. (2004) Empirical Analysis of Entropy Distance Metric for UML Class Diagrams. *ACM SIGSOFT Software Engineering Notes*, **29**, 11-16. <http://dx.doi.org/10.1145/1022494.1022524>
- [7] Yi, T. and Wu, F.J. (2008) A Longitudinal and Comparative Study of Complexity Metrics for UML Class Diagrams

-
- through Internet Banking. *Proceedings of Management Track within WiCOM: Engineering, Services and Knowledge Management (EMS'2008)*, Dalian, 12-14 October 2008, 1-4. <http://dx.doi.org/10.1109/wicom.2008.2947>
- [8] Yi, T. (2010) Comparison Research of Two Typical UML Class Diagram Metrics: Experimental Software Engineering. *Proceedings of the International Conference on Computer Application and System Modeling (ICCASM'2010)*, Taiyuan, 22-24 October 2010, 86-90.
- [9] Witten, I.H. and Frank, E. (2011) *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd Edition, Morgan Kaufmann, San Francisco.
- [10] Baudry, B., Traon, Y.L. and Sunyé, G. (2002) Testability Analysis of a UML Class Diagram. *Proceedings of the 8th International Symposium on Software Metrics*, Ottawa, 4-7 June 2002, 54-63. <http://dx.doi.org/10.1109/metric.2002.1011325>
- [11] Huang, J. and Ling, C.X. (2005) Using AUC and Accuracy in Evaluating Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, **17**, 299-310. <http://dx.doi.org/10.1109/TKDE.2005.50>