Report on Faculty Research Grant Project: *Using Normalized Compression Distance to Study Software Evolutionary Stability*

Liguo Yu  
Department of Computer and Information Sciences  
Indiana University South Bend  
1700 Mishawaka Avenue  
South Bend, IN 46634  
574-520-5067  
ligy@iusb.edu

August 20, 2012

1. Introduction

In December 2011, I received the Faculty Research Grant entitled “Using Normalized Compression Distance to Study Software Evolutionary Stability.” The project started in January 2012 and ended in August 2012. Most of the research activities are performed in the summer of 2012. The project is successful with anticipated outcomes.

2. Research Activities

The followings summarize the activities supported by the grant.

- **Design information-level metrics.** Kolmogorov complexity and normalized compression distance are used to define several information-level metrics.
- **Build a tool to measure differences between software artifacts.** A MS Windows based C++ program is implemented to measure the compression distance between two software artifacts.
- **Evaluate the metrics.** The information-level metrics proposed in this study are evaluated with open-source software artifacts.
- **Analyze the data.** The results are summarized and analyzed using statistical methods and visualization tools.
- **Publicize the result.** The research is summarized in two manuscripts.

3. Outcomes

With the support of this grant, four kinds of outcomes are produced from this research.

- **Information-level evolutionary metrics**, which are presented in [1].
- **Normalized compression distance tool**, which are made publicly available [2].
- **Evolution pattern of open-source projects**, which are summarized in [1] and [3].
- **Publications indirectly supported by the grant** [4] [5] [6] [7].
4. Acknowledgement

I am thankful to the grant and everyone related to the approval and management of this grant. I would also like to thank my colleague Prof. Jim Wolfer for his idea of applying normalized compression distance in software engineering field. Thanks are also given to Prof. Rathinasamy B. Lenin at University of Central Arkansas and Dr. Kai Chen of Google Incorporation for their support.

References

2. Sourceforge, Normalized Compression Distance, https://sourceforge.net/projects/normalizedcompr/