IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Kolmogorov Complexity.

Ankur Singh Bist
[*1]Govind Ballabh Pant University of Agriculture And Technology
ankur1990bist@gmail.com

Abstract

This paper describes various application issues of kolmogorov complexity . Information assurance , network management , active network are those areas where kolmogorov complexity is applied . Our main focus is to show it's importance in various domain including the domain of computer virus detection .

Keywords- Kolmogorov Complexity, Information Assurance .

Introduction

The Kolmogorov complexity of a string x is the length of the
smallest program that outputs x, relative to some model of
computation [1] . That is,
Cf (x) = min p {|p| : f (p) = x}
for some computer f .
Informally, C(x) measures the information content, degree of
redundancy, degree of structure, of x.
Cf (x) depends on both f and x.

Invariance theorem[1]:----

There exists a universal description method ψ_0, such that:

$$C_{\psi_0}(x) \leq C_{\psi}(x) + c$$

for some constant c that depends on ψ and ψ_0 (but not on

Another important factor is[1]:----
For all universal description methods f , g:

$|Cf(x) - Cg(x)| \leq c$
for some constant c that depends only on f and g.

Conditional kolmogorov complexity[1] :----
The conditional Kolmogorov complexity of a string x, relative to a string y and a model of computation f , is:
Cf (x|y) = min{|p| : Cf (p, y) = x}

$$C_f(x) = C_f(x|\epsilon)$$

Information Assurance And Active Network Management by kolmogorov complexity

Amit B. Kulkarni and Stephen F. Bush proposed an approach using kolmogorov complexity and complexity theory to design self-managed networks and vulnerability analysis methods that draw on basic properties of information to identify, analyze, and correct faults including security vulnerabilities in an information system.

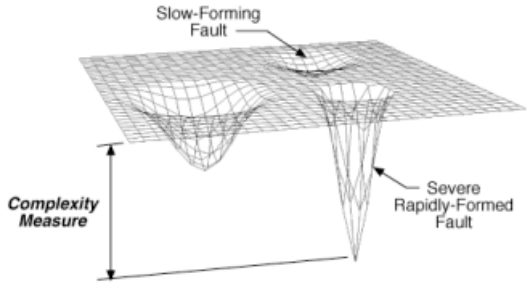


Fig 1. Complexity based analysis of network health [2]

Finally authors showed complexity model and experimental validation to show the results. Thus they explained the first step in applying algorithmic information theory and techniques to network management to develop the next generation of networks that will contain self-diagnosing, self-managing, and self-healing nature .
Scott Evans, Stephen F. Bush, and John Hershey proposed a approach to look towards information assurance with the help of kolmogorov complexity and minimum message length criteria.

Theorems of conservation[3]:------

Theorem 1: Bound on Conditional Complexity

$$K_{\varphi}(y \mid x) \leq K_{\varphi}(y)$$

Theorem 2: Bound on Complexity Increase Due to Computational Operation

$$K_\varphi(y \mid x, p) \leq K_\varphi(x) + L(p)$$

Since kolmogorov complexity is not computable, few applications exist for Kolmogorov Complexity. One developing application is a statistical technique with good links to information theory called Minimum Message Length coding(MML) . MML coding encodes information as a hypothesis that detects the presumptive distribution, from which data originated, appended with a string of data, coded in an optimal way. The length of an MML message is calculated as follows[3]:

#M = #H + #D,

#M = the message length

#H = the length of the specification of the hypothesis regarding the data,

#D = length of the data, encoded in an optimal manner given hypothesis

MML coding approaches the Kolmogorov Complexity or actual bound on the minimum length required for showing a string of data.

When MML approach is used to complexity measurement, apparent complexity, then complexity as observed by a particular attacker could be determined.

Computer Virus Detection By Kolmogorov Complexity

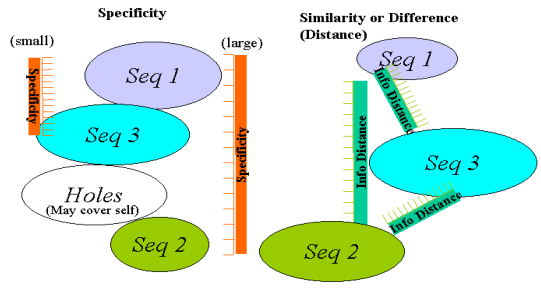
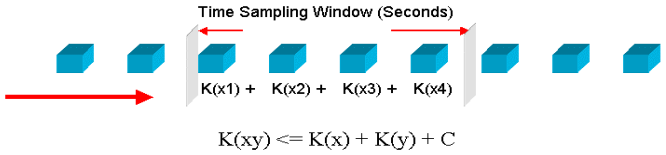


Fig1. Comparing specificity and complexity[4]



K(xy) <= K(x) + K(y) + C

Fig2. Differential complexity[4]

Sanjay Goel ,Stephen F. Bush explained complexity as a function to define immunology and epidemiology . For this they firstly explained the basic theme of kolmogorov complexity and then explained about network states in terms of virus infection as[4]:

1. Susceptible
2. Infected
3. Immune

Various simulation analysis is made to explain the concept of virus spread. To give broader explanation the concepts of random graphs is explained . They also focused on the literature on this concerned issue in lucid manner. Analysis using stochastic model is done including following parameters[4]:-------

1. Impact of large diameter ,low connectivity graph on stochastic model
2. Impact of medium diameter graph on stochastic model
3. Impact of low diameter , high connectivity graph on stochastic model

Detailed simulation results explaining specificity is shown and with this comparison of specificity and complexity is shown while explaining differential complexity.

Expected matching time that is shown in paper includes techniques[4]:-----

1. Contiguous matches
2. Contiguous sliding matches
3. Total matches
4. Total sliding matches
5. Entropy
6. Zip
7. Lz

Atlast they discuss about correlation among traditional signature matching techniques and correlation of complexity differential estimates versus signature matching .Thus kolmogorov complexity concept get used to classify information and gives a fitness criteria for classification.

Conclusion

In this paper we discussed various approaches regarding application of kolmogorov complexity and show the problem solving methods with the help of kolmogorov complexity. This review study gives a collective look on different scattered aspect under one thread and will be helpful for those working in this direction .

References

[1] Neil convey, " The issue of kolmogorov complexity."

[2] Amit B. Kulkarni and Stephen F. Bush ," Active Network Management and Kolmogorov Complexity" NY, USA

[3] Scott Evans, Stephen F. Bush, and John Hershey GE Corporate Research and Development ," Information Assurance through Kolmogorov Complexity."

[4] Sanjay Goel ,Stephen F. Bush ," Kolmogorov Complexity Estimates For Detection Of Viruses In Biologically Inspired Security Systems: A Comparison With Traditional Approach"