

RINGKASAN

REDUKSI DATA DENGAN METODE *SQUASHING*, Eto Wuryanto, Dyah Herawatie dan Rimuljo Hendradi, 2004, 41 halaman.

Pembuatan sub-sampel dengan menggunakan metode sampling acak dan metode sampling acak stratifikasi sebenarnya mudah dilakukan tetapi mempunyai kelemahan. Sub-sampel yang dihasilkan metode sampling acak mempunyai varians yang besar jika sampel induknya heterogen. Sedangkan pada metode sampling acak stratifikasi juga mempunyai kelemahan jika *strata* tidak didefinisikan secara jelas. Selanjutnya ternyata ada metode baru yaitu metode *Squashing*, permasalahannya adalah bagaimana cara mendapatkan sub-sampel dari metode ini (data *Squashing*)? atau dengan kata lain bagaimana mereduksi data berukuran besar dengan menggunakan metode *Squashing*? dan selanjutnya bagaimana menerapkan metode *Squashing* tersebut untuk mereduksi data yang berdistribusi Pareto tergeneralisir dengan bantuan program S-Plus?

Penelitian ini bertujuan untuk mereduksi data berukuran besar dengan metode *squashing* dan menerapkannya untuk mereduksi data yang berdistribusi Pareto tergeneralisir dengan bantuan program S-Plus. Sehingga dapat memberikan solusi alternatif dalam mereduksi data, khususnya untuk data berukuran besar yang berdistribusi Pareto tergeneralisir.

Reduksi data dengan menggunakan metode *squashing* dilakukan dengan cara. Pertama, melakukan pengelompokkan terhadap data induk ke dalam beberapa partisi. Kedua, menentukan pembobot dan *pseudo point* pada setiap kelompok atau partisi. Ketiga, menentukan nilai data *squashing* yang merupakan perkalian antara pembobot dan *pseudo point*, kemudian menghitung jumlahan data induk dan data hasil *squashing*. Dalam penelitian ini, data yang dipakai adalah data bangkitan yang diperoleh dengan cara membangkitkan bilangan acak berdistribusi $U(0,1)$ yang kemudian ditransformasi ke dalam distribusi Pareto tergeneralisir.

Algoritma reduksi data dengan menggunakan metode *squashing* pada data berukuran besar yang berdistribusi Pareto tergeneralisir diperoleh dengan cara sebagai

berikut : Pertama, dilakukan pengelompokan data ke dalam beberapa kelompok dengan anggota yang sama atau berbeda. Kedua, untuk per kelompok dihitung jumlah anggotanya yang berfungsi sebagai nilai pembobot dan nilai *pseudo point*-nya sama dengan rata-rata atau varians per kelompok.. Ketiga, menentukan perkalian antara jumlah anggota per kelompok dengan rata-rata atau variansnya untuk mendapatkan data *squashing*. Hasil reduksi data dengan menggunakan metode *squashing* diindikasikan dengan adanya persamaan antara jumlahan data hasil *squashing* dan data induknya.

Implementasi algoritma metode dilakukan dengan menggunakan program S-Plus. Sedangkan data yang digunakan pada penelitian ini adalah data bangkitan berdistribusi pareto tergeneralisir dengan $\alpha = 0.5$, $\beta = 0.25$, dan $n = 100.000$. Penerapan program terhadap data bangkitan dengan 3 jenis pengelompokkan menghasilkan semakin banyak jumlah kelompoknya maka momen sub sampel dari hasil *squashing* mendekati data induknya.

Adapun saran yang ingin disampaikan adalah : Pertama, penerapan metode *squashing* masih menggunakan data bangkitan, diharapkan dapat dicoba pada data sekunder. Kedua, program-program yang terdapat pada penelitian ini tidak hanya dapat diterapkan pada data yang berdistribusi Pareto tergeneralisir tetapi juga memungkinkan untuk data yang berdistribusi lainnya. Ketiga, untuk penelitian lebih lanjut dapat dilakukannya antara lain : penerapan metode *squashing* untuk variabel lebih dari satu, pengkajian terhadap indikator apa yang dapat dipakai untuk menyatakan bahwa sampel induk dapat diwakili oleh sub-sampel yang dihasilkan oleh metode *squashing* misalnya membandingkan suatu estimator suatu parameter yang dihasilkan oleh sampel induk dengan yang dihasilkan oleh sub-sampelnya.

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Airlangga, Nomor Kontrak : 890/JO3.2/PG/2004, Tanggal : 12 Juli 2004.

SUMMARY

REDUCTION DATA USING SQUASHING METHODS, Eto Wuryanto, Dyah Herawatie dan Rimuljo Hendradi, 2004, 41 pages.

A sub sample of random sampling methods and stratified random sampling methods is actually simple but they have any disadvantages. The sub sample of random sampling methods has a large variance if the main sample is heterogeneous. And the one of stratified random sampling methods also possess a weakness if the strata are not well-defined. In the present, there is a new sampling method that is called Squashing methods. The problem is what steps can be used to obtain a sub sample of Squashing methods? Or how can we reduce a large data to a considerably smaller one by using Squashing methods? And next, may we applicate the S-Plus program of Squashing methods to reduce a data from generalized Pareto distribution?

The goal of this research is to reduce a large data by using Squashing methods and to applicate the methods that has been translated in S-plus program to a data from generalized Pareto distribution. So the research can give an alternative solution in data reduction, especially for a large data from generalized Pareto distribution.

The stages of data reduction using Squashing methods : The first, cluster a main data into any partitions. The second, compute a weighted value and an appropriate pseudo point for each partition. The third, calculate multiplication between the weighted value and the pseudo point to get data squashing so we are able to know that the sum of main data and the one of squashing data have a same value. In this research, we utilize a generated data that is produced by transforming a random number of $U(0,1)$ distribution to a generalized Pareto distribution number.

The algorithm of data reduction using squashing methods for a large data from generalized Pareto distribution is : The first, make any group with equal or unequal elements. The second, each group determine a number of elements as weighted value and its momen (mean or variance) value as pseudo point. The third, compute squashing data that is equal to a number of elements and its pseudo point is

a mean or variance value for each group. Finnacle, both the the sum of main data and the one of squashing data have a same value.

Implementation of the squashing's algorithm is done by writing it to S-Plus program. The data that is taken in this research is generated data from generalized Pareto distribution with $\alpha = 0.5$, $\beta = 0.25$, and $n = 100.000$. Application of the program to the data for three kinds of group yield increasing of number of group cause the momen's value of sub sample from squashing methods approach the one of main data.

Any advice can be present as the following : The first, we hope you can find a real data and applicate squashing methods on it. The second, the S-plus program in this research not only can be used on data from generalized Pareto distribution but may for a data from another distribution. The third, this research can be continued by doing any work as the below : you are able to use squashing methods for multivariable data, finding what indicator may be utilised to know that main data can be represented by sub sample from squashing methods, for exemple : compare a estimator of parameter that is calculated by using the main data with the one of sub sample from squashing methods.

Mathematics Department, Faculty of Mathematics and Sciences, Airlangga University, Contract's Number : 890/JO3.2/PG/2004, Date : 12 Juli 2004.