ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A Survey on Compressing the Dependent Element of Multiset's Linear Form

Yogesh Shankar Landge[1], Dr. K.P.Adhiya[2], Prof. Dr.Girish K. Patnaik[3]

Student, Department of Computer Engineering, SSBT's COET, Bambhori, Jalgaon (M.S.),India[1]

Associate Professor, Department of Computer Engineering, SSBT's COET, Bambhori, Jalgaon (M.S.), India[2]

Professor and Head, Department of Computer Engineering, SSBT's COET, Bambhori, Jalgaon (M.S.), India[3]

ABSTRACT: A multiset isan unordered collection of mathematical objects with repetitions allowed. Many authors showed that there is only one way to represent the multiset, i.e. the representation of multiset as sequences. But no one showed that the multiset can also be represented in linear form. By performing operations like (union, intersection, sum And difference) on multisets having linear form, the dependent elements of multiset is obtained. And can be compressed and decompressed in the similar manner as the independent elements of the multiset, because both the elements are related to each other. A Lossless compression and decompression algorithms for multisets; taking advantage of the multiset's representation structure is described. The encoding technique that transforms the dependent elements of multiset into an order-invariant tree representation, and derive an arithmetic code that optimally compresses the tree. Thesugessted lossless compression algorithm achieves different arithmetic codes for dependent elements and independent elements of multiset and generating their lengths respectively on the basis on certain operations performed onmultisets. With the help of this scheme both the elements independents and dependent of multiset can be compressed and decompressed simultaneously.

KEYWORDS: Arithmetic coding; Dependent multiset; Tree data structures; Data compression.

I. INTRODUCTION

The theory of sets is indispensable to the world of mathematics. But in set theory where repetitions of objects are not allowed it often become difficult to complex systems. If one considers those complex systems where repetitionsof objects become certainly inevitable, the set theoretical concepts fails and thus one need more sophisticated tools to handle such situations. This leads to the initiation of multiset (M-set) theoryas a generalization of set theory. A multiset is a set that each item in the set has a multiplicity which specifies how many times the item repeats.

The motivation behind the present study is the knowledge of the representation of multisets in different forms and the certain operations performed on multisets [12] and [13]. With the help of the two concepts, the new idea can be generated, i.e. the detailed knowledge could be obtainedin the field of the dependent elements of multiset.

The contribution of the present study is to extend the concept of multiset theory suggested by Steinrueckenin [11].In [11], the author does not give any method to generate a mutiset having the dependent elements. The main contribution of the present study is to how to obtain the dependent elements of multiset. After that, applying a lossless compression algorithm on the dependent elements of multiset to compress them into a tree based arithmetic codes and vice-versa.

The organization of the paper is as follows: TheSection 2 describes Literature Survey based on sets and multiset's theory used in different areas. The solution suggested on compressing the dependent element of multisets based on operations performed on multiset's using lossless compression algorithm is described in the Section 3 and the Section 4 gives the conclusion implicating benefits of suggested work.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

II. LITERATURE SURVEY

Monro, in [1], proposedthe concept of multiset thatall duplicates of an object in a multisetwere indistinguishable. The object of amultiset was the distinguishable or distinct elements of the multiset. There was a problem of recognizing the objects and the elements of the multisets. The author provided the solution for the problem. The distinction was made between the terms object and element enriched the multiset language.Monro [1] appliedequivalence relation approach to explicate the concept of multisets. In [1], a multiset [a, a, b] was regarded as being really of the form [a, a', b] where a anda' were different objects of the same sort and b is of a different sort. Notationally, elements of distinct sorts was to be denoted by distinct letters and elements of the same sort be denoted by the same letter with a sequence of dashes distinguishing different elements of that sort.In [1], a multiset A is formally defined as a pair <A, z> where A is a set and z is an equivalence relation on A. The set A is called the field of the multiset. Elements of A in the same equivalence class is said to be of the same sort and elements in different equivalence classes is said to be of different sorts. For example, amultiset [a2, b, c3, d] is represented as [a, a', b, c, c', c'', d] where a, a' are of the same sort and c, c', c'' are also of the same sort, while b and d are of two different sorts. In other words, various equivalence classes determine the sorts. The disadvantage of the concept is that it does not consider theconsistency criterion for the existence of mono, epi, and iso-morphisms in multiset**.**

Singh and Isah, in [2], proposed a note on category of multisets (Mul) that overcomes the problem suggested by the author Monro in [1]. Monro [1] was a prototype work on Multiset, Where the consistency criterion for the existence of mono, epi, and iso-morphisms in Multiset was not considered. In [2], the authors provided the consistency criterion for the existence of mono, epi, and iso-morphisms by using mapping concept between multisets. The authors give the criterion that objects of the same sort cannot be mapped to objects of different sorts. (i) f: X →Y is a monomorphism in Multisetiff f: X→ Y is one-to-one, sort-preserving. For example, f: [a, a’, a’’]→[b, b’, b’’] can be a monomorphism. (ii) f: X →Y is an epimorphism in Muliff f: X→Y is onto, sort-preserving. For example, f: [a, a’, a’’] → [b, b’] can be an epimorphism. (iii) f: X →Y is an isomorphism in Muliff f: X→ Yis a bijection, sort-preserving. The advantage of the work is that it makes easier to identify the type of morphism between multisets.

Reznik, in [3], proposed the concept ofcoding of sets of words. The classic problem of source coding is to encode a given sequence of words (or a message) w =w1, w2. . . with the goal of minimizing the number of bits needed for such encoding. Most commonly, it is further assumed that words must be decoded in the same order as they appear, and that the resultof decoding must be unique and matching the original. This setting, coupled with the assumption about stochastic nature of the source, has lead to many fundamental results and techniques, including Shannon’s source coding theorem, Huffman codes, and others. Nevertheless, in practice, a slightly different problem may be encounter: the message may be given by a set of words {w1. . . wm}, where their order is not important. This happens, for example,when we formulate a request to a search engine by providing a list of keywords. Such keywords can be communicated in any order, without affecting the meaning of the message. Given this flexibility, one may expect a code constructed for a set {w1. . . wm} to consume about log2m! Less bits than a code constructed for a particular sequence wi1. . . wim. However, the construction of codes for unordered sets does not appear to be entirely trivial: the problem was that most existing source coding tools assume sequential processing. Here the author offered one possible practical solution for this coding problem. Overall, theanalysis shows that the proposed scheme delivers close to the predicted optimal performance in the memoryless model, and that the difference (redundancy rate) becomes progressively small, as the length of words n increases.

Larssonand Moffat, in [4], proposedthe concept ofoffline Dictionary-Based Compression. In most implementations of dictionary-based compression, the encoder operates onlineincrementally inferring its dictionary of available phrases from previous parts of the message, and adjusting its dictionary after the transmission of each phrase. Doing so allows the dictionary to be transmitted implicitly, since the decoder simultaneously makes similar adjustments to its dictionary. An alternative approach is to use the full message (or a large block of it) to infer a complete dictionary in advance, and include an explicit representation of the dictionary as part of the compressed message. Intuitively, the advantage of this offlineapproach is that with the benefit of having access to the entire message, it should be possible to optimize the choice of phrases so as to maximize compression performance. Indeed, the authors demonstrated that very good compression can be attained by an offline method without compromising the fast decoding that is a distinguishing characteristic of dictionary-based techniques. The compression technique obtains excellent compression, yet is also capable of high decoding rates. Moreover, the advantage of the approach is that the amount of memory used by the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

decoder can be reduced at the cost of slower decoding. The principal drawback of the approach is that encoding is relatively expensive.

Ghilezan, in [5], proposed binary relations and algebras on multisets. Contrary to the notion of a set or a tuple, a multiset is an unordered collection of elements which do not need to be different. As multisets are already widely used in combinatorics and computer science, in [8], the author applied algebras on multisettheory. The author considers generalizations of known results that hold for equivalence and order relations on sets and get several properties that are specific to multisets. Furthermore, the author exemplify the novelty that brings this concept by showing that multisets are suitable to represent partial orders.The advantage of the concept is that it proves two algebras on multisets cannot be isomorphic even if their root algebras are isomorphic.

Isah and Tella, in [6], proposed the concept on categories of multiset (Mul) and topological spaces (Top).The authorsillustrated the concept of an isomorphism in categorical context and shows that a bimorphism is not necessarily an isomorphism in the categories multisetand topological spaces. Multisets (considered as objects) and multiset functions (considered as morphisms) together determine the category of multisets, denoted Mul. For example, f: [a, b, c] → [d, e, e'], defined by f(a) = d, f(b) = e', f(c) = e, is an msetmorphism which is a monomorphism and an epimorphism, and therefore a bimorphism, but not an isomorphism: since f(b) and f(c) are of the same sort but b and c are of different sorts. The multisetmorphism f: [a, a', b] → [c, c', d], defined by f(a) = c', f(b) = d, f(a') = c, is an isomorphism. Moreover, the morphism g: [a, b] → [c, d], defined by g(a) = c, g(b) = d, is also an isomorphism.The advantage of the concept is that the author showed bothMul and Top are not balancedthus; a bimorphism in Mul, Top is not necessarily an isomorphism. And, whereas Set, Abelian group, Ring and Group are balanced since every bimorphism is an isomorphism.

Girish and John, in [7], proposed the idea on multiset topologies.General topology is defined as a set of sets but multiset topology is defined as a set of multisets. The concept of topological structures and their generalizations is one of the most powerful notions inbranches of science such as chemistry, physics and information systems. In most applications the topologyis used to handle the qualitative information. In any information system, some situationsmay occur, where the respective counts of objects in the universe of discourse are not single. In suchsituations wheredealing with collections of information in which duplicates are allowed. In suchcases multisets play an important role in processing the information. The information system dealing withmultisets is said to be an information multisystem. Thus, information multisystems are more compact whencompared to the original information system. In fact, topological structures on multisets are generalizedmethods for measuring the similarity and dissimilarity between the objects in multisets as universes. Thetheoretical study of general topology on general sets in the context of multisets can be a very useful theoryfor analyzing an information multisystem. Most of the theoretical concepts of multisets are originated from combinatorics. Combinatorial topology is the branch of topology that deals with the properties of geometric figures by considering the figures as being composed of elementary geometric figures. The combinatorial method is used not only to construct complicated figures from simple ones but also to deduce the properties of the complicated from the simple. In combinatorial topology it is remarkable that the only machinery to make deductions is the elementary process of counting. In such situations, elements can occur more than once. The theory of M-topology may be useful for studying combinatorial topology with collections of elements with duplicates.

Tella and Daniel, in [8], proposed the concept of the symmetric groups under multiset perspective. Research on the multiset theory has not yet gained ground and is still in its infant stages. The research carried out so far shows a strong analogy in the behavior of multisets and sets and it is possible to extend some of the main notions and results of sets to that of multisets for instance, the theoretical aspects of multisets by extending the notions of relations, functions, composition and partition has been explored. The advantage of the concept is that it defines a symmetric multigroup and derives the analogous Cayley's theorem.

Tella and Daniel, in [9], proposed a study of group theory in the context of multiset theory. Unlike set, a multiset is an unordered collection of elements where elements can occur more than once .the problem was that the study of classical group theory in the context of sets was not being presented, because sets do not allow elements to occur more than once. In [9], the author presented the study of classical group theory in the context of multisets because multisets allow repetition of elements. But the disadvantage of the presented study is that it does not address the concepts of Lagrange's theorem, homomorphism of groups and symmetric groups in the context of multigroup theory.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Girish and John, in [10], proposed the concept of Rough Multisets and Information Multisystems. The most essential property of multisets is the multiplicity of the elements that allows distinguishing it from a set and considering it as a qualitatively new mathematical concept. In any information system, some situations may occur, where the respective counts of objects in the universe of discourse are not single. There are many situations where dealing with collections of information in which duplicates are significant. In such cases multisets play an important role in processing the information. The information system dealing with multisets is said to be an information multisystem. Thus, information multisystems are more compact when compared to the original information system. The authors have given a new dimension to Pawlak's rough set theory, replacing its universe by multisets. This is called a rough multiset and is a useful structure in modelling information multisystem. The process involved in the intermediate stages of reactions in

Chemical systems are a typical example of a situation which gives rise to multiset relations.Information multisystem is represented using rough multisets and is more convenient thanordinary rough sets. The author proposed the idea with Yager's theory of multisets. After presenting the theoretical study, the concepts mset relations, equivalence mset relations, partitions, and knowledge mset base have been established. Finally the concepts of rough multisets and related properties with the help of lower mset approximation and upper mset approximations have been introduced. The advantage of the proposed concept that rough multisets are important frameworks for certain types of information multisystems.

Kuo, in [11], proposed the concept ofMultiset Permutations in Lexicographic Order. The author proposed a simple and flexible method for generating multisetpermutations in lexicographic order.Multiset permutation can be applied in combinations generation, since a combination of pitems out of q items set is a special case of multisetpermutations that contain q 1s and (q-p) 0s. For example, a combination of 5 items out of 10 items can be described as 1010011001 to stand the first item, the 4th item, the 5th item, 8^{th} item and the last item are picked up. Obviously, this is a permutation of a multiset$\{5 \cdot \mathbf{0}, 5 \cdot \mathbf{1}\}$. In other words, the advantage of the technique is that the multiset permutations can be generated directly without any help of remapping. The new method is conceptually easy to understand and implement and is well-suited to a wide variety of permutation problems.

Steinruecken, in [12], proposed the concept of compressing sets and multisets of sequences. The problem was that the information was being wasted for encoding multisets by simply storing its elements in some sequential order. In [15], the author gave the solution to the problem, by proposing a technique that transforms the multiset into an order-invariant tree representation, and derives an arithmetic code that optimally compresses the tree. The advantage of the technique is that the information is saved while encoding the multisets. And on the other hand, the disadvantage of the technique is that it does not compress the dependent elements of multisets.

The past woks on the concept of the multiset theory arediscussed here. And it concludes that the research on the multiset theory is still in at infant stage. The past works presented were covered different areas in multisets, but the idea on compressing the dependent elements of multiset was not explored there. In [15]Steinruecken was provided the concept of compressing the independent elements of multiset, where the author represented multiset as sequences. But the author did not provide the method to generate the dependent elements of multiset and compressing the dependent elements of multiset into tree-based arithmetic codes. The proposed scheme gives the method to generate the dependent multiset from independent multisets and compresses the dependents elements of multiset into arithmetic codes similar to the independent elements of mutiset.

III. PROPOSED SOLUTION

In tree based encoding and decoding, multisets can be encoded and decoded using various techniques. Tree based encoding needs to ensure that the compression should be lossless compression. The compression and decompression of independent multisets has been done with the help of encoding and vice-versa. But in the existing scheme[11], when the scenario ismultisets having the dependent elements, there is no way to obtain the dependent elements of multiset and after that, compress and decompress like the independent elements of multiset to convert into arithmetic codes.

In this scheme it is assure that encoding is done with lossless compression and it is done on dependent multisets. Different operations are performed on multisets inorder to generate dependent multisets on which encoding is done.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

The suggested scheme takes into considerations two things that are required to provide the solution to the problem.

1) The representation of the multisets as a linear form instead of sequences and
2) The operations on multisets

There are certain types of operations like union, intersection, sum and difference can be applied on multiset's, after performing the respective operations on multiset A and multiset B (assume), the resulting multiset (assume multiset c) is obtained.

a. Case If the union operation is performed on multiset A and multiset B, the resulting multiset C is obtained, where the multiset A and multiset B having the independent element while the multiset C contains the dependents elements.
b. Case If the intersection operation is performed on multiset A and multiset B, the resulting multiset C is obtained, where the multiset A and multiset B having the independent element while the multiset C contains the dependents elements.
c. Case If the sum operation is performed on multiset A and multiset B, the resulting multiset C is obtained, where the multiset A and multiset B having the independent element while the multiset C contains the dependents elements.
d. Case If the difference operation is performed on multiset A and multiset B, the resulting multiset C is obtained, where the multiset A and multiset B having the independent element while the multiset C contains the dependents elements.

The existing scheme only compresses and decompresses the independent elements of multiset (taking one multiset at a time) and generated the respective arithmetic codes for the multiset. It is known that there are two parts of multisets, one part is multiset having the independent elements and second part is multiset having the dependent element.

To obtain the second part of multiset that is multiset having the dependent element, firstlythe multiset is represented in the linear form. And secondly the certain operations like union, intersection, sum and difference are perform on multiset's having the independent elements, so the resulting multiset is a multiset having the dependents element. This is the process to get the second part of multiset (multiset having the dependents element).

Now, performing similar thelossless compression and decompression algorithms [11], on the dependent elements of multisets, the resulting arithmetic's codes are generated based on operationsperformed. With the help of arithmetic's codes, the difference between independent elements of multisets and dependentelement of multiset can be made.

There are certain advantages using the suggested scheme:

1. At the same time both the part of multiset (independent and dependent) can be compressed and decompressed.
2. It means that minimize transmission time on simultaneously executing both the part of multisets.
3. It minimizes storage space and
4. It minimizes transmission cost.

IV. CONCLUSION

In the present paper, the suggested scheme ensures that the multisets can be represented in various forms. The linear form is the one of them. The advantage of representing multiset in linear form is that the dependent multiset can be obtained by performing operations on multisets having linear forms.The second part of multiset (multiset having dependent element can be compressed in the form of arithmetic codes and vice versa.The sugessted scheme not only compresses the independent elements of multisets but also the dependents elements of multiset by using operations on multisets.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

REFERENCES

1. G. P. Monro, 'The concept of multiset', ZeitschriftfürmathematischeLogik und Grundlagen der mathematic [mathematical logic quarterly], Volume2, Issue 33, pp. 171-178, 1987.
2. [1]Dasharath Singh and [2]Ahmed Ibrahim Isah, 'A Note on Category of Multisets (MUL)', International Journal of Algebra, Volume 7, Issue 2, pp.73 – 78,2013.
3. Y. A. Reznik, 'Coding of sets of words', in Proc. IEEE Data Compress. Conf.,pp. 43–52, March, 2011.
4. Silvia Ghilezan, JovankaPantović, and [3]Gradimir Vojvodić, 'Binary Relationsand Algebras onMultisets', PUBLICATIONS DE L'INSTITUT MATHÉMATIQUENouvelle série, tome 95 (109), 111–117, 2014.
5. [1]Ahmed Ibrahim Isah and [2]Yohanna Tella, 'On Categories of Multisets (MUL) and Topological Spaces (TOP)', International Journal of Science and Technology Volume 2, Issue No. 6, June, 2013.
6. K. P. Girisha and Sunil Jacob Johnb, 'On Multiset Topologies', Theory and Applications of Mathematics & Computer Science, Volume 3, Issue 4, pp.37–52, 2012.
7. YohannaTella and Simon Daniel, 'Symmetric Groups under multiset perspective', IOSR Journal of Mathematics , Volume 7, Issue 5, PP 47-52, Jul. – August, 2013.
8. Y.Tella and S. Daniel, 'A Study of Group Theory in the Context of Multiset Theory', International Journal of Science and Technology Volume 2 No. 8, August, 2013.
9. K. P. Girish and Sunil Jacob John, 'Rough Multisets and Information Multisystems', Hindawi Publishing Corporation Advances in Decision Sciences Volume 17, issue12,pp 211-218,2011.
10. Ting Kuo, 'Multiset Permutations in Lexicographic Order', International Journal of Emerging Technology and Advanced Engineering ,Volume 4, Issue 1, January, 2014.
11. Christian Steinruecken, 'Compressing Sets and Multisets of Sequences', IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 61, Issue NO. 3 MARCH, 2015.
12. D. Singh, A. M. Ibrahim, T. Yohanna, and J. N. Singh, 'A systematization of fundamentals of multisets', LecturasMatematicas, Volume 29,pp-33-48, 2008.
13. N J Wildberger,' A new look at multisets', School of Mathematic,2003.

BIOGRAPHY

Yogesh Shankar Landgeis a Student in Department of Computer Engineering, SSBT's COET, Bambhori, Jalgaon(M.S.),India.

Dr.K.P. Adhiyais an Associate Professor in Department of Computer Engineering, SSBT's COET, Bambhori, Jalgaon(M.S.),India.

Prof. Dr.Girish K. Patnaik is a Professor and Head in Department of Computer Engineering, SSBT's COET, Bambhori, Jalgaon(M.S.),India.