JAISCR, 2012, Vol. 2, No. 3, pp. 247 – 258

GEODESIC DISTANCES FOR CLUSTERING LINKED TEXT DATA

Selma Tekir[1], Florian Mansmann[2] and Daniel Keim[3]

[1] *Dept. of Computer Engineering*
Izmir Institute of Technology
Izmir, Turkey 35430
Email: selmatekir@iyte.edu.tr

[2] *Faculty of Computer Science, Box 78*
University of Konstanz
Konstanz, Germany 78457
Email: Florian.Mansmann@uni-konstanz.de

[3] *Faculty of Computer Science, Box 78*
University of Konstanz
Konstanz, Germany 78457
Email: Daniel.Keim@uni-konstanz.de

Abstract

The quality of a clustering not only depends on the chosen algorithm and its parameters, but also on the definition of the similarity of two respective objects in a dataset. Applications such as clustering of web documents is traditionally built either on textual similarity measures or on link information. Due to the incompatibility of these two information spaces, combining these two information sources in one distance measure is a challenging issue. In this paper, we thus propose a geodesic distance function that combines traditional similarity measures with link information. In particular, we test the effectiveness of geodesic distances as similarity measures under the space assumption of spherical geometry in a 0-sphere. Our proposed distance measure is thus a combination of the cosine distance of the term-document matrix and some curvature values in the geodesic distance formula. To estimate these curvature values, we calculate clustering coefficient values for every document from the link graph of the data set and increase their distinctiveness by means of a heuristic as these clustering coefficient values are rough estimates of the curvatures.

To evaluate our work, we perform clustering tests with the k-means algorithm on a subset of the English Wikipedia hyperlinked data set with both traditional cosine distance and our proposed geodesic distance. Additionally, taking inspiration from the unified view of the performance functions of k-means and k-harmonic means, min and harmonic average of the cosine and geodesic distances are taken in order to construct alternate distance forms. The effectiveness of our approach is measured by computing micro-precision values of the clusters based on the provided categorical information of each article.

1 Introduction

The principal aim of the information retrieval systems is to retrieve only the relevant documents from the document collection. In order to determine the relevance, similarity/distance measures are utilized on the document representations. The representations are in terms of some low-level features that are directly measurable from data. Document lengths, the frequency of words in documents are examples of such features.

In the classical IR implementation, feature vectors are formed for documents and distance between these vectors is calculated to relate them. The ordinary distance metric is position-independent in the sense that if two data points are shifted by the same amount in one coordinate the distance between them does not change. In other words, it does not take into account the topology of the document space and assumes that the space is flat (zero curvature). A curvature metric can therefore provide additional information about the data points and is dependent on the position in the space.

In order to address the document semantics in a better way; there is a need for a generalization that captures all the geometric structure of space including the notions of distance, angle, volume, and curvature [1]. The formulation of this generalized metric (metric tensor) varies according to peculiarities of the space. Thus, the distance computation is dependent on the inherent space. This paper proposes a geodesic distance metric that extends the classical distance computation with the measurements of curvatures so that the specificities of the document space can be reflected in a better way.

The geodesic distance metric provides a way of combining features, which can be applied to data sets that offer multiple feature spaces. For our experiments, the selected data set contains linked text documents on which link and text based features can be calculated. The text analysis is conducted using the *Vector Space Model* (VSM) [2] of information retrieval. The term weights are calculated based on term frequencies plus some normalization mechanisms such as *inverse document frequencies* (idf). The regular cosine distance that is applied to the tf-idf version of the term-document vectors is used as the basic similarity measure.

In the computation of the geodesic distances, the cosine distance is combined with the curvature measurements. The curvature values are based on the clustering coefficient values from the link graph given the fact that the clustering coefficient values are rough estimates of the curvatures [3].

The importance of the geodesic distance metric lies in the fact that it utilizes a mathematical cost function for combining links with the text similarity measures. There exists link-based ranking approaches as well as retrieval models incorporating link evidence. However, there is a lack of optimal cost functions to combine cosine, link indegrees, PageRank, etc.

The experiments in this paper are conducted on the Wikipedia XML Corpus [4] English subset. Wikipedia seems a good selection because it is a known fact that in contrast to general web links, Wikipedia links are good indicators of relevance. In addition to this; in Wikipedia, outlinks and inlinks are similar in character and both contribute to the semantic analysis of the documents unlike the Web in which indegrees have a dominant role in determining the semantic relatedness [5]. Thus, the clustering coefficient computations that are based on the undirected link graph of the collection are plausible choices as link-based features for the fact that there is symmetry in the semantic nature of Wikipedia (if A is relevant to B then B is relevant to A, too).

For measuring the effectiveness of the proposed approach, we use data clustering algorithms. An overwhelming theme for different data clustering techniques/algorithms is to convert the objective into an optimization problem and propose an optimization (performance) function accordingly. The proposed optimization function is expected to measure the goodness of the data analysis objective at hand. Thus, dependable performance functions are of vital importance in the field.

We are given the Wikipedia categorical information as part of the data set. The most common text clustering algorithm *k-means* [6] is used for the tests. The rationale for selecting *k-means* is twofold. First, as we already know the number of categories to look for in the data set, we easily set the *k*, the main argument of the algorithm. Second, there exists an abstract framework for integrating multiple feature spaces for the k-means algorithm. The second property can be attributed to the simple but powerful nature of the k-means performance function. For practical reasons the algorithms are run on some subsets of the whole data collection.

The rest of the paper is organized as follows: In Section 2, we provide some related work to define the context and give an overview of the state-of-the-art. In Section 3, we discuss the geometric meaning of the geodesic distance in comparison to the existing similarity/dissimilarity measures and give the calculation scheme. In Section 4, we provide detailed information about the experimental setup-data set, algorithms, parameters, evaluation metrics and depict the experimental results. Finally, we conclude this paper and raise some issues for future work in Section 5.

2 Related Work

Ma et al [7] claim that they are the first researchers that use geodesic distance in text mining related research areas. Their work deals with the query-based sentence retrieval and compares geodesic with cosine distance in this context. The method constructs a graph of all sentences including the candidate (query) ones. In order to define the local neighborhood, a threshold variable epsilon is introduced. If the distance between two sentences are below the threshold value then a direct link is established between them. The geodesic distances are computed over the links by utilizing shortest-path algorithms on the sentence graph. Resulting rankings and correct ratio plots for given queries according to both geodesic distance and cosine are provided. The results show that for the particular values of the parameter epsilon the correct ratio values of the geodesic distance are superior to cosine's, for some other range it degenerates the cosine angle distance.

In hyperbolic IR [8], which is non-Euclidean, a geometric meaning is introduced to the positions in space. The query vector is assumed to be at the center of the hyperbolic sphere and the other documents are evaluated according to their hyperbolic distances to the query vector at the center of the sphere. In short, if a non-Euclidean aspect is to be introduced to a metric space model, the points should be specialized. Another important point is that change of hyperbolic distance according to the radius of the hyperbolic sphere as a parameter introduces equivalent ranking as traditional similarity measures plus weighting schemes.

Xiao et al [9] associate with the geodesic a cost based on length and sectional curvature. The sectional curvature is determined by the degree to which the geodesic bends away from the Euclidean chord. Hence for a geodesic in space, the sectional curvature can be estimated easily if the Euclidean and geodesic distances are known. Put it another way if the Euclidean distance and sectional curvature values are known, geodesic distances can be easily computed. Lou [3] states that *clustering coefficient values* are rough estimates of the sectional curvatures. Getting the sectional curvature values from the link graph, taking cosine from the term-document matrix, the geodesic distance computation can be easily adapted to the text documents with links.

In PageRank [10], global link structure of the document set is utilized to calculate the ranks of the documents. The *authority* concept is introduced in HITS [11] to determine the importance of the documents. An outlink from a source document to a target one means that the source gives some authority to the target. Additionally, it is also critical from whom you get authority. Therefore, there are a set of hub documents from which having inlinks is more valuable. This hub-authority pattern is the key idea and applied in a local context after filtering out documents by text-based queries.

Language models provide mechanisms to utilize link evidences along with the text content scores. The experimental results of Kamps and Koolen [5] show that local degree priors are better than the global degree priors and weighted local/global priors are even more helpful. Thus, the proposed approach is plausible as it presents a compromise between global and local by evaluating local connectivity on the global link graph.

Strehl et al [12] provide a framework for evaluating the impact of similarity measures on clustering web pages. In this work, the fundamental similarity measures are discussed along with their geometric interpretation. The clustering algorithms that are better suited to term-document matrix based text data are determined and the existing similarity/distance measures' performance with these algorithms are stated.

Oikonomakou and Vazirgiannis [13] review web document clustering approaches. They classify the existing algorithms according to characteristics or features that are used. The processed features in the web context are text and/or link-based features. Thus; text-based, link-based, and hybrid approaches exist for web document clustering. This work proposes a hybrid approach for this purpose.

In their work which combines the link-based measures with the content-based classifiers, Calado et al [14] state that the effectiveness of the combination approach may depend on the importance given to each of the sources of evidence to be combined. More weight should be given to those that provide more reliable information. They recognize finding the ideal weights for each of the evidences to be combined as the fundamental problem. They set their objective as pursuing methods to automatically determine such weights and alternative ways to combine link-based and content-based evidences.

Yang [15] points out that there is no unanimity in the research findings related to link analysis and/or fusion methods. Some claim that combining results of various retrieval methods is beneficial to retrieval performance, others' results state that fusion in general seemed to decrease retrieval performance. The main question according to Yang is finding out the reason of the general failure of the fusion may be due to the characteristics of test collections, failings of link analysis, inadequacies of fusion formula, or combinations of all or any of the above. He believes the future fusion efforts should focus on discovering the fusion formula that can best realize the fusion potential of combining diverse retrieval methods.

In this work, we attempt to use geodesic distances to better address the semantics of linked text documents. In other words, geodesic distance formula is proposed as a way of combining text-based and linked-based features. This paper is an extended version of the SSCI CIDM 2011 paper [16]. Our extensions include the evaluation of k-harmonic means algorithm to test the effect of initialization in the precision results of cosine and geodesic in k-means algorithm. Moreover, an alternate distance out of cosine and geodesic is calculated by taking the minimum and harmonic average of the given distances.

3 Geodesic Distance

The distinguishing property of the proposed geodesic distance is that local curvature values are considered in the calculation of the distance between the objects. The intuition behind the approach comes from the Riemannian geometry where a local curvature of uniform sign across the manifold implies strong global properties. Thus, we take into consideration the sign of the curvature in the algorithm and are in the pursuit of such global behaviors.

This intuition in mind, we come up with a calculation scheme for geodesic distance. The proposed scheme is based on the relationship between Euclidean and geodesic distances on the unit circle as shown in Figure 1. In order to introduce the geodesic similarity measure for the linked text documents, the formula is formed using the relationship between Euclidean and geodesic distances on the unit circle as shown in Figure 1. The line length between two points on a unit circle represents the Euclidean distance while the arc length between those points represents the geodesic one.

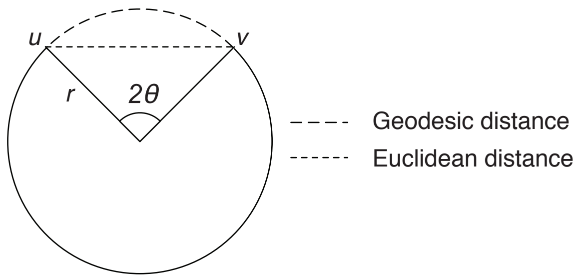

Figure 1. Euclidean and geodesic distances on a circle.

The line length is computed using the respective triangle and can be stated as follows:

$$d_E(u,v) = 2r\sin\theta \qquad (1)$$

The arc length is given by the following formula:

$$d_g(u,v) = 2r\theta \qquad (2)$$

The sine in the Euclidean distance formula can be approximated using the Maclaurin series:

$$d_E(u,v) = 2r(\theta - \frac{1}{6}\theta^3 + \ldots) \qquad (3)$$

Substituting for θ obtained from the geodesic distance, we have

$$d_E(u,v) = d_g(u,v) - \frac{d_g^3(u,v)}{24r^2} \qquad (4)$$

Finally, radius is represented in terms of the curvature of the circle as follows;

$$r = \frac{1}{\kappa} \quad (5)$$

and the resulting equation is solved for the geodesic distance.

$$d_g(u,v)^3 - 24\frac{1}{\kappa^2}d_g(u,v) + 24\frac{1}{\kappa^2}d_E(u,v) = 0 \quad (6)$$

In this equation, the parameters are dependent on κ and $d_E(u,v)$ values respectively. Thus, geodesic distances can be calculated in terms of κ curvature values and $d_E(u,v)$ Euclidean distances.

As we work with linked text documents in our context, we compute the clustering coefficient values from the link graph to substitute for κ curvature value and replace $d_E(u,v)$ Euclidean distance by the cosine text similarity measure.

The Maclaurin series expansion in equation 3 can be extended by one more term such as the following:

$$d_E(u,v) = 2r(\theta - \frac{1}{6}\theta^3 + \frac{1}{120}\theta^5 \ldots) \quad (7)$$

Again by substituting for θ obtained from the geodesic distance, the equation becomes:

$$d_E(u,v) = d_g(u,v) - \frac{d_g^3(u,v)}{24r^2} + \frac{d_g^5(u,v)}{1920r^4} \quad (8)$$

Lastly, radius is replaced by the curvature equivalent and the following quintic equation is obtained:

$$d_g^5(u,v) - \frac{80}{\kappa^2}d_g^3(u,v) + \frac{1920}{\kappa^4}d_g(u,v) - \frac{1920}{\kappa^4}d_E(u,v)$$

$$=0 \quad (9)$$

The clustering coefficient is defined as $C_i = 2n/(k_i(k_i-1))$, where n denotes the number of direct links connecting the k_i nearest neighbors of node i [17]. It is the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them. The coefficient represents the local connectivity of a document by giving a measure of the degree of interconnectedness in the neighborhood of a node. A node whose neighbors are all connected to each other has $C = 1$, whereas a node with no links between its neighbors has $C = 0$.

The clustering coefficient provides an approximation of the scalar curvature in the sense that $C = 0$ implies that the scalar curvature at that vertex is negative, while $C = 1$ means that the scalar curvature is positive, with $C = 1/2$ the borderline case of vanishing scalar curvature. In our case, each document is thus assigned one clustering coefficient value.

The geodesic distance equation given in equation 6 is a special cubic equation in which the coefficient of the squared term is zero. In this cubic form there are two complex roots and one real root and the average of the roots of the equation is zero. For solving the equation Cardano's method for cubics is utilized [18]. In our experiments, the real root is used as the geodesic distance. The quintic geodesic distance equation, on the other hand, can be solved numerically by Newton's method.

Geodesic distance can be seen as a weighted distortion measure in the clustering context in that the weights are taken from the link graph and the cosine distance is used as a base distortion measure. Distortion measures are used to evaluate the results of the clustering along with the ground truth categorization. The weighted distortion measures are defined as

$$D^{\alpha}(x,x') = \sum_{l=1}^{m} \alpha_l D_l(F_l, F_l') \quad (10)$$

where D is the distortion measure, α the weight, x the cluster set, x' the category set, and F_l and F_l' lth feature vectors in the cluster and category sets respectively. The vector of weights is called feature weighting.

Document-cluster set membership and document-category set membership matrices are shown in Table 1. At the last row, the corresponding feature vectors are depicted.

Distortion measures should be utilized in a way that within category distances must be smaller. In the k-means algorithm, it means that document to cluster centroid distances must be smaller that is

Table 1. The Feature Weighting for Cluster and Category Sets Respectively

	Cluster Set					Category Set			
	x_1	x_2	...	x_m		x'_1	x'_2	...	x'_m
d_1	w_{11}	w_{12}	...	w_{1m}	d_1	w_{11}	w_{12}	...	w_{1m}
d_2	w_{21}	w_{22}	...	w_{2m}	d_2	w_{21}	w_{22}	...	w_{2m}
:	:	:	...	:	:	:	:	...	:
d_n	w_{n1}	w_{n2}	...	w_{nm}	d_n	w_{n1}	w_{n2}	...	w_{nm}
	F_1	F_2	...	F_m		F'_1	F'_2	...	F'_m

Table 2. Link Related Statistics of the Wikipedia XML Corpus

	min	max	mean	median	stdev
indegree	0	74950	20.9016	4	289.0161
outdegree	0	5176	20.9016	12	37.3416
ccoef	0	1	0.2493	0.2	0.1875

k disjoint clusters $\pi_1^\dagger, \pi_2^\dagger, ..., \pi_k^\dagger$ are generated in the manner to minimize the objective function which is given below:

$$\{\pi_u^\dagger\}_{u=1}^k = \underset{\{\pi_u\}_{u=1}^k}{\operatorname{argmin}}\left(\sum_{u=1}^{k} \sum_{x \in \pi_u} D^\alpha(x, c_u)\right) \tag{11}$$

c_u thereby denotes the cluster centroids vector.

4 Experimental Evaluation

The experiments are conducted on the Wikipedia XML Corpus [4], which is composed of hyperlinked Wikipedia articles. Two category files are also included in the data set. One contains id_document, id_category pairs and the other lists the category names for the defined category ids. In total, there are 659,388 documents in 72 English portal categories in the collection. In Table 2, link related statistics (indegree, outdegree, and clustering coefficient) of the data set are provided:

Link related statistics say that the indegrees follow a power-law distribution and the clustering coefficient values have a tendency to be smaller than 0.5, which means that the curvature is mainly negative in the inherent document space.

We randomly selected 10 categories to test our approach. In Table 3 you find the selected category names along with the corresponding document counts.

The clustering coefficient values are calculated based on the global link graph rather than the link graph for the selected categories because the clustering coefficient values begin to converge when the node count increases. In fact, you cannot get clustering coefficient values other than NaN using the category link graphs as in-category links are quite rare. As for the text part, each document is considered as a multi-dimensional vector and bag-of-words approach with tf-idf is utilized to form final document vectors. In the experimental scenario, as we deal with high dimensional data, clustering algorithms that have to face the curse of dimensionality would not fit the scheme. Thus, the popular k-means algorithm has been chosen. k-means is a good choice because we use two feature sets, namely a) curvature values and b) term-document vectors and there exists an abstract framework for integrating multiple feature spaces in the k-means algorithm [19].

Table 3. Selected Categories with Size

Category Name	Size
Bangladesh	393
Colombia	304
Finland	1887
Hong Kong	11056
Morocco	230
Netherlands	1350
New Zealand	2393
Romania	1340
Uganda	232
Venezuela	569
Total	19754

The feature combination is done using a mathematical function to compute geodesic distances by exploiting both textual information and the link topology. In our approach, we calculate the clustering coefficient values using the whole adjacency matrix of the data set and save these values along with the belonging document ids. As the clustering coefficient values are rough estimates of the curvatures, simple heuristics are applied to them in order to increase their distinctiveness. Algorithm 1 details the heuristic we use to generate the curvature values.

Algorithm 1 generateCurvature algorithm.

$ccoef$: clustering coefficient value
if $ccoef \rangle 0.5$ **then** ▷ Curvature is positive?
    $ccoef$=$ccoef$-0.5
else
    $ccoef$=$ccoef$+1
end if

The clustering coefficient values of the documents in the collection are mainly negative with a mean of 0.2. This means that the documents reside in a hyperbolic space. In the heuristic defined in the *generateCurvature* algorithm it is assumed that for negative curvature values which are close to 0.5 (zero curvature), distortion should be greater than for values that are far from 0.5. Thus, 1 is added to the negative curvature values in order to arrange the distortion accordingly. This is consistent with the graph of the distortion of embedding the Internet as a function of the curvature of the embedding space given by [20] in Figure 2. When it comes to positive curvatures, their ordering is preserved and their effect on the centroid curvature is weakened by making a subtraction.

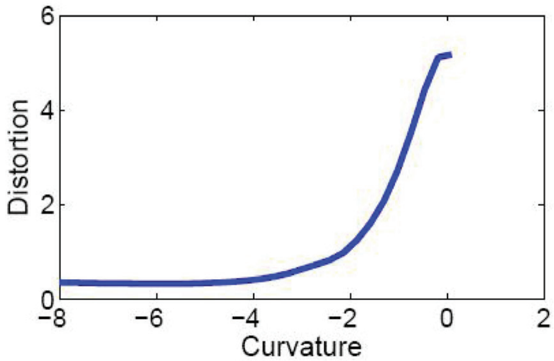

Figure 2. The distortion of embedding the internet in dimension two as a function of the curvature of the embedding space [20].

After generating the curvature values we need to compute the average of them in order to represent the curvature of the centroids in the k-means clustering algorithm. The centroid curvature is calculated by taking the average of the individual curvature values belonging to the documents that are classified around the same cluster centroid. In this computation, we disregard the NaN values, which are quite rare.

In the experiments, k-means with cosine similarity measure is compared against the k-means with geodesic similarity measure. In order to have a fair comparison we fix the initial cluster assignments. In the cosine case, we run k-means with no initial cluster assignments since the code randomly determines the initial centroids. In the geodesic case, we use the same initial centroids from the cosine case in order to see the effect precisely. In short, in every run the cosine and geodesic share the same initial cluster assignments. However, the initial cluster assignments differ among different runs. We run the experiments 10 times.

We set the number of clusters parameter k as twice the number of categories in order to see the effect more clearly. In the same way, Strehl et al. [12] choose clusters that are twice the number of categories and explain that this setting provides the more natural number of clusters as indicated by preliminary runs and visualization.

The clustering results are evaluated using the metrics rand index [21] and adjusted rand index (AR) that are pair-counting based as well as mutual information [22] [12] and normalized mutual information (NMI), which are information-theoretic measures. In particular, the k-means clustering results are evaluated according to the normalized versions of van Dongen, mutual information and rand index criteria which are stated as the right measures for the algorithm by Wu et al [23]. In the computation of these specified clustering metrics we need the category labels vector and the cluster labels vector as input. As we set the number of clusters for the k-means algorithm to twice the number of categories, while category labels vary between 0 and n, cluster labels have range 0 to 2n. In other words we end up with a contingency table which has n rows (categories) and 2*n columns (clusters). The approach to be taken at this stage to calculate the evaluation metrics for the clustering is complicated. The difficulty lies in determining the criterion to select the n clusters out of 2*n. If you ignore this varying range problem and calculate the metrics accordingly, the clustering quality suffers. If you take the columns (clusters) that have the highest intersection with the categories, it is not fair because in one case the second largest group can be very close in size to the first one whereas in others the gap can be quite big.

The intersection among the selected categories (documents that belong to more than one category) form a small set thus the effect on the clustering can be ignored.

In the evaluation part, we calculate the precision numbers in order to measure the overlap between a given clustering and the ground truth classification. In our case the ground truth classification is given as Wikipedia categories. We comparatively analyze the clustering results for the k-means with cosine and k-means with geodesic with the real categories. The precision computations are done based on the methods provided by [19]. Their work establishes the framework for integrating multiple feature spaces in the k-means clustering algorithm. Thus, valid comparisons between single feature spaces and multiple feature spaces in the k-means case can be best accomplished using the framework's defined precision metrics rather than the traditional clustering metrics for the k-means namely NMI and AR. In our experiments, we also calculated NMI and AR values. The results verify that the order of the NMI and AR values in the cosine and geodesic cases is in accordance with the order of the defined precision metric values in both cases for every run.

To meaningfully define precision, we convert the clusterings into classification using the following simple rule: identify each cluster with the class that has the largest overlap with the cluster, and assign every element in that cluster to the found class. The rule allows multiple clusters to be assigned to a single class, but never assigns a single cluster to multiple classes.

Suppose there are c classes $\{\omega_i\}_{i=1}^{c} = 1$ in the ground truth classification of n objects. Precision is defined using the following equations where a_i denotes the number of data objects that are correctly assigned to the class ω_i, b_i the documents that are incorrectly assigned to the class ω_i, and c_i denotes the documents that are incorrectly rejected from the class ω_i.

$$p_i = \frac{a_i}{a_i + b_i} \quad and \quad r_i = \frac{a_i}{a_i + c_i}, \quad 1 \leq i \leq n \tag{12}$$

The precision is defined per class. In order to capture the performance averages across classes micro-precision (micro-p) values are calculated as follows:

$$micro - p = \frac{1}{n} \sum_{i=1}^{c} a_i \tag{13}$$

The experimental results (micro-precision values) are shown in Table 4. The first column lists the values belonging to k-means with cosine, the second column k-means with geodesic, the third column k-means with a geodesic derivative, the fourth, min of cosine-geodesic pair, and finally the last one harmonic mean of cosine-geodesic pair respectively. The difference between the two geodesic approaches is in the calculation of the average centroid curvature values. The former one sums the curvature values without paying attention to the signs of the curvature. In the latter one the summation operation takes into account the signs that is the positive ones are added to the sum whereas the negative values are subtracted from it.

k-means' performance function aims at minimizing the total within-cluster variance by the way of minimizing the total mean squared distance for each point and the closest centroid. The closest centroid assignment of a point implies that the algorithm implicitly assigns every point to exactly one cluster, imposing a hard membership for points. k-harmonic means [24], on the other hand, uses the distances to all centroids in order to assign weights

Table 4. The micro-p values with the clusterings with k-means cosine, geodesic, geodesic derivative, first twos' min and harmonic means respectively. The corresponding mean and standard deviation values are added as the last two rows.

run ♯	cosine	geodesic	geodesic-derivative	min	harmonic
1	0,727448	0,737775	0,743191	0,732712	0,737268
2	0,72355	0,733117	0,738585	0,709122	0,729473
3	0,718285	0,724208	0,724866	0,721120	0,721930
4	0,738585	0,74552	0,740812	0,746178	0,746229
5	0,662752	0,676774	0,684823	0,670902	0,673534
6	0,700618	0,706135	0,696973	0,707958	0,713779
7	0,702288	0,705528	0,703756	0,701023	0,698289
8	0,678394	0,683912	0,685633	0,679660	0,684722
9	0,728612	0,728663	0,720462	0,730687	0,728106
10	0,724309	0,708565	0,690088	0,715298	0,706439
mean	0,7105	0,7150	0,7129	0,7115	0,7140
stdv	0,0243	0,0228	0,0234	0,0234	0,0233

to the points and before the final convergence phase there's no assignment to any particular clusters. Therefore, k-harmonic means utilizes soft membership and has the capability of moving points to other cluster centers in the case of high locally dense data points and centers [25].

Inherently, k-means has sensitivity to initialization and k-harmonic means is said to be essentially insensitive to initialization due to the above mentioned capability over k-means. Therefore, it's a good starting point to investigate the effect of initialization on the k-means algorithm in pursuing the factors related to the performance of geodesic over cosine in the experiments. Both cosine and geodesic approaches were run on a k-harmonic means implementation, but no distinguishing difference was observed. Then the effect of initialization can be disregarded in comparing the effectiveness of cosine and geodesic similarity measures in k-means clustering applications.

k-means' performance function given in equation 11 can be rewritten as:

$$\{\pi_u^\dagger\}_{u=1}^k = \sum_{i=1}^N \min(|x_i - c_u|^2 \,|u = 1,...,k) \qquad (14)$$

This new representation is the result of a unified view of the k-means and k-harmonic means' performance functions [24]. The part that comes right after the min, represents the distance function. The min assigns the documents to clusters according to minimum distances. In the k-harmonic case, the harmonic averages (HA) of the distances from each data point to the centers are computed as components to the relative performance function.

Taking inspiration from this rewritten form of the performance function, the min and HA can be evaluated as operators that are applied to the succeeding distance functions. In the context of this paper, these operators can be moved inside and be applied directly to the distance function part as well. As we have cosine and geodesic distances in the experimental setting, by calculating the minimum and harmonic averages of the two distances, an alternate distance form can be generated to be useful. In Table 4, the calculations for both of these variations are listed as the fourth and fifth columns respectively.

In order to compare the effects of the different distance measures, we perform the nonparametric Friedman's test. The test is conducted on three different triples: "cosine-geodesic-geodesic derivative", "cosine-geodesic-min", and "cosine-geodesic-harmonic". The rows are the different runs. The resulting p values of the Friedman's test for the triples are as follows: 0.0608, 0.0450, and 0.0273.

The Friedman's test evaluates the hypothesis that the column effects are all the same against the alternative that they are not all the same. The first result says that the three distance measures are not the same within the 90 % confidence interval. The other two prove that these variations introduce statistically important effects to the already computed values within the 95 % confidence interval. In other words, the methods affect the clustering effectiveness.

When we have a look at the micro-p values given in Table 4, we see that the worst performance for the geodesic cases is in the last run. In order to find out the reason behind that, we expanded Maclaurin series approximation in the equation 3 by one more term ending up with the equation 7. We used Newton's solver to numerically estimate the root of the quintic equation given in equation 9. The clustering results we get show that the quintic geodesic equation improves the results in favor of some specific categories whereas it works against the remaining ones resulting in almost the same micro-p value we have. When we analyze the contingency tables for cosine and geodesic in every run, we also realize that the geodesic runs' improvements are the results of great performances on those specific categories. Thus, the geodesic approach's effectiveness must have some relation with some category-specific attribute. However, we have not clarified it yet.

According to the mean and standard deviation of the different distance measurements provided in Table 4, the geodesic better expresses the inherent clustering structure of the data (due to higher mean) and at the same time it is more robust as it has less variance.

5 Conclusion

In this work, we propose a novel distance measure for clustering hypertext documents, which is based on both textual information and the link topology of the hypertext document collection. It is useful to highlight the basic components of our approach:

- The basic assumption is that clustering coefficient values that indicate the local connectivity structure of the documents can be used as curvatures. This assumption is based on the fact that clustering coefficients are rough estimates of curvatures [3]. They are computed on the global link graph of the data set.
- The notion of geodesic distances in curved spaces is used to define a mathematical function to do feature combination. For this purpose the geodesic distance calculation scheme on the 0-sphere is utilized.
- Text features are combined with the generated curvature values in order to improve the clustering results in the k-means case. This means integrating multiple feature spaces in the k-means algorithm. One needs a framework that covers the comparative analysis of multiple feature spaces over single feature spaces in the k-means algorithm to test the effectiveness of the candidate functions for feature combination. The abstract framework provided for the feature weighting in the k-means algorithm [19] defines the context and the evaluation methodology for the work.

The experiments are conducted on the Wikipedia XML Corpus English subset [4]. The evaluation metrics are based on the ground truth classification provided as the Wikipedia categorical information. The results show that the curvature values calculated based on the link graph of the data set can be used to fine-tune the similarity values so that the objective function for the clustering can be minimized. Furthermore, the k-means algorithm has proven to be suitable for the proposed geodesic method because of the centroid concept. Using centroid curvature value rather than the individual document clustering coefficient values to fine-tune the cosine is more reasonable as centroid curvature value is a better indicator of locality. Thus, the geodesic approach can be transferred to contexts where there is a multiple feature space, in which one feature can represent curvature, and there is a cumulative calculation potential for this feature.

5.1 Future Work

The experimental results show that in some runs the geodesic approaches perform better than cosine whereas in some others they are slightly worse. The next task to be done is to discover the factors related to the success or failure of the geodesic method.

In pursuing the factors related to the performance of geodesic over cosine in the experiments, the effect of initialization on the k-means algorithm has been investigated. As previously denoted, cosine and geodesic approaches' outcomes do not have any relationship with the choice of initialization in the k-means algorithm. K-harmonic means implementations were run to analyze the initialization sensitivity and no important improvements were observed in both similarity measures and one's performance in comparison to the other's.

The results motivate alternative computation schemes for geodesic distances. We use geodesic distance computation formula for 0-sphere (circle) in this work. Alternatively, great-circle distances (1-sphere) can be utilized as geodesics. On the other hand; rather than assuming that the space is spherical, taking into consideration the fact that the clustering coefficient average for the data collection coincides with a negative curvature value, the underlying space can be assumed as hyperbolic. Hyperbolic distance calculation schemes in accordance with the given parameters can be devised.

We believe that the heuristics that are applied to the clustering coefficient values in order to generate curvatures can be systematically studied and improved. Furthermore, other linked data sets can be used to further evaluate the effectiveness of the geodesic distance measure.

References

[1] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern Geometry - Methods and Applications: Part I: The Geometry of Surfaces, Transformation Groups, and Fields (Graduate Texts in Mathematics)*. Springer, 1991.

[2] G. Salton, *Automatic text processing : the transformation, analysis, and retrieval of information by computer*. Addison-Wesley., 1989.

[3] M. Lou, "Traffic pattern in negatively curved network," Ph.D. dissertation, University of Southern California, 2009.

[4] L. Denoyer and P. Gallinari, "The wikipedia xml corpus," *SIGIR Forum*, vol. 40, no. 1, pp. 64–69, 2006.

[5] J. Kamps and M. Koolen, "Is wikipedia link structure different?" pp. 232–241, 2009.

[6] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. L. Cam and J. Neyman, Eds., vol. 1. University of California Press, 1967, pp. 281–297.

[7] M. Hui-Fang, H. Qing, and S. Zhong-Zhi, "Geodesic distance based aproach for sentence similarity computation," in *Machine Learning and Cybernetics, 2008 International Conference on*, vol. 5, 2008, pp. 2551–2557.

[8] J. Goth and A. Skrop, "Varying retrieval categoricity using hyperbolic geometry," *Inf. Retr.*, vol. 8, no. 2, pp. 265–283, 2005.

[9] B. Xiao and E. Hancock, "Geometric characterisation of graphs," in *Image analysis and processing: ICIAP 2005, 13th international conference, Cagliari, Italy, September 6-8, 2005*. Springer, 2005, pp. 471–478.

[10] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford InfoLab, Technical Report, 1999, previous number = SIDL-WP-1999-0120.

[11] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[12] A. Strehl, E. Strehl, J. Ghosh, and R. Mooney, "Impact of similarity measures on web-page clustering," in *In Workshop on Artificial Intelligence for Web Search (AAAI 2000*, 2000, pp. 58–64.

[13] N. Oikonomakou and M. Vazirgiannis, "A review of web document clustering approaches," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Springer US, 2005, pp. 921–943.

[14] P. Calado, M. Cristo, E. Moura, N. Ziviani, B. Ribeiro-Neto, and M. A. Gonalves, "Combining link-based and content-based methods for web document classification," pp. 394–401, 2003.

[15] K. Yang, "Combining text- and link-based methods for web ir." in *Proceedings of the 10th Text Rerieval Conference (TREC-10)*. Washington, DC.: US Government Printing Office., 2001.

[16] S. Tekir, F. Mansmann, and D. Keim, "Geodesic distances for web document clustering," in *Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on*, april 2011, pp. 15 –21.

[17] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998, 10.1038/30918.

[18] J. White and D. Kalman, "Cardano: An adventure in algebra in 8 parts."

[19] D. Modha and W. Spangler, "Feature weighting in k-means clustering," *Machine Learning*, vol. 52, pp. 217–237(21), September 2003.

[20] E. Begelfor and M. Werman, "The world is not always flat or learning curved manifolds." School of Engineering and Computer Science, Hebrew University of Jerusalem., Tech. Rep., 2005.

[21] W. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.

[22] T. Cover and J. Thomas, *Elements of Information Theory 2nd Edition*. Wiley-Interscience, 2006.

[23] J. Wu, H. Xiong, and J. Chen, "Adapting the right measures for k-means clustering," pp. 877–886, 2009.

[24] B. Zhang, M. Hsu, and U. Dayal, "K-harmonic means - a data clustering algorithm," 1999.

[25] D. Turnbull, "K-means and k-harmonic means: A comparison of two unsupervised clustering algorithms." University of California San Diego Department of Computer Science and Engineering, Tech. Rep., 2002.

JAISCR, 2012, Vol. 2, No. 3, pp. 259 – 274

METAHEURISTIC OPTIMIZATION OF MARGINAL RISK CONSTRAINED LONG-SHORT PORTFOLIOS

G A Vijayalakshmi Pai[1] and Thierry Michel[2]

[1] *Department of Computer Applications PSG College of Technology, Coimbatore, India*

[2] *Tactical Asset Allocation and Overlay*
Lombard Odier Darier Hentsch Gestion
Paris, FRANCE

Abstract

The problem of portfolio optimization with its twin objectives of maximizing expected portfolio return and minimizing portfolio risk renders itself difficult for direct solving using traditional methods when constraints reflective of investor preferences, risk management and market conditions are imposed on the underlying mathematical model.

Marginal risk that represents the risk contributed by an asset to the total portfolio risk is an important criterion during portfolio selection and risk management. However, the inclusion of the constraint turns the problem model into a notorious non-convex quadratic constrained quadratic programming problem that seeks acceptable solutions using metaheuristic methods.

In this work, two metaheuristic methods, viz., Evolution Strategy with Hall of Fame and Differential Evolution (rand/1/bin) with Hall of Fame have been evolved to solve the complex problem and compare the quality of the solutions obtained. The experimental studies have been undertaken on the Bombay Stock Exchange (BSE200) data set for the period March 1999-March 2009. The efficiency of the portfolios obtained by the two metaheuristic methods have been analyzed using Data Envelopment Analysis.

1 Introduction

A financial *portfolio* is a basket of tradable assets such as bonds, stocks, securities etc. *Portfolio optimization* which deals with determining a combination of assets that best suits the investor's preferences, is a traditional problem in *quantitative finance*. Given the investor's attitude, tolerance to risk and his or her return expectations, the issue is to allocate *weights* on the different assets of the portfolio.

Modern Portfolio Theory [Elton et al., 2003] is based on a formal expression of the representative investor's preferences. Markowitz [1952] laid the framework for solving the portfolio selection problem using a mean-variance model, assuming the asset returns follow a Gaussian law. Hence the *expected return* of a portfolio is described using the mean returns of the assets and the *risk* of the portfolio is summarized by the variances and covariances of the assets' returns.

The theory asserts that the only risks priced in the market are those common to several assets, and not the specific risk inherent to each individual asset. From this point of view, it is thus optimal to diversify the investment over a wide spectrum of stocks in order to neutralize *idiosyncratic risks*. The rational investor in this framework should then add as many assets as possible in his or her portfolio.

Mathematically, the problem of portfolio optimization reduces to the optimization of a *bi-criterion objective function* that characterizes *minimizing risk* and *maximizing the expected portfolio return*. The solution to the problem lies in graphically obtaining what is termed an *efficient frontier* which is a risk-return tradeoff curve, giving the minimum level of risk to take for an expected return or, alternatively, the maximum return one can expect for a given level of risk.

The Markowitz model, for the solution of the portfolio selection problem assumed a perfect market disallowing short sales, ignoring transaction costs and taxes and permitting trading of securities as any fraction. Under these assumptions, the problem reduces to a quadratic programming problem that can be easily solved by classical methods.

However, in practice, the portfolio optimization problem model is beset with *constraints* reflecting investor preferences, market frictions and risk management that have rendered them difficult for direct solving using analytical methods. It is in such scenarios that metaheuristic strategies have been looked up to for their solution. Thus Chang *et al*., [2000], Streichert *et al*., [2003], Kendall and Su, [2005], Maringer [2005], Gilli *et al.,* [2008] Nikos *et al*., [2009], Pai and Michel [2009, 2011, 2012(a), 2012(b)] investigated multi-agent methods such as Evolutionary algorithms, Differential Evolution and Particle Swarm Optimization, and Fernando and Gomez [2007] investigated Hopfield Neural Networks for the solution of constrained portfolio optimization problem models.

In this work, we discuss a portfolio optimization problem that is governed by the *basic constraint*, *bounding constraint*, *short selling* besides *marginal risk control*.

A basic constraint ensures that the weights allotted to each asset in the portfolio lies between 0 and 1 with their sum adding to 1. In other words, the investor expects his or her capital to be fully invested. In practice, it is quite often the case that a portfolio manager is limited in his choices by his or her mandate, with a ceiling on either individual assets and/or types of assets, such as sectors, styles or geographic zone. In the former case the constraint is referred to as *bounding constraint* and in the latter case as *class constraint. Short selling* is indulged in when assets are borrowed from a third party with the understanding of returning them to the lender at a later date with the short seller hoping to profit from a decline in the price of the assets between the sale and the repurchase, for the short seller hopes to pay less to buy the assets back than he or she received while selling them. Thus short selling constraints allow weights to carry *negative values* and such assets are termed *short positions* in a portfolio. In contrast, *long positions* are assets which yield returns during their rise in prices and hence are indicated by positive weights.

A portfolio made up of only long positions is termed *long only portfolio* in portfolio management parlance. In contrast, it is also possible to operate under a constraint when the weights can vary between –c to +c for some constant c and their sum adds up to 1. A portfolio which is an assortment of both long positions and short positions is known as a *long-short portfolio*.

The *marginal risk contribution* is a measure of risk to monitor the risk contribution of a portfolio. Mathematically, Grinold and Kahn (1999) defined the marginal contribution of risk of a given asset as the partial derivative of the standard deviation of the portfolio return with respect to the position of that asset.

The inclusion of the marginal risk constraint turns the portfolio optimization problem model into a quadratic constrained non convex quadratic programming optimization problem rendering it difficult for direct solving by classical methods.

Zhu *et al.,* (2010) attempted solution of a marginal risk constrained portfolio optimization model by proposing an efficient branch and bound method which identified and exploited the special properties of the problem model. However, their investigations remarked that '…global optimality of the solution generated by the branch-and-bound method is guaranteed by the fact that the relaxed subproblem over a subrectangle can approximate the original problem over the same subrectangle with any accuracy provided that the subrectangle is sufficiently small…' Pai and Michel (2011) investigated a metaheuristic solution to a marginal risk constrained portfolio optimization problem model, however, the objective of the model was to maximize the Sharpe ratio of the portfolio and dealt with global asset allocation with their own specific constraints.

In this work, we discuss the application of two metaheuristic methods belonging to two different genres of evolution computation, viz., Evolution Strategy with Hall of Fame and Differential

Evolution with Hall of Fame, for the solution of the marginal risk constrained portfolio optimization problem model also governed by the basic, bounding and short selling constraints. The metaheuristic methods employ penalty functions and weight standardization procedures to efficiently tackle the constraints. In the absence of existing work discussing a metaheuristic solution of the problem model, the two methods served to compare the solutions obtained and their performance analyses.

The experimental studies have been undertaken on the Bombay Stock Exchange (BSE200) data set for the period March 1999-March 2009. The efficiency of the portfolios obtained by the two metaheuristic methods have been analyzed using Data Envelopment Analysis.

Section 2 details the mathematical formulation of the marginal risk constrained portfolio optimization problem. Section 3 details the two metaheuristic methods of Evolution Strategy with Hall of Fame and Differential Evolution with Hall of Fame. Section 4 explains the weight standardization procedures for constraint handling. Section 5 describes the application of the metaheuristic strategies for tackling the portfolio optimization problem. Section 6 details the experimental studies and the performance analysis tackling the constraints. Section 7 presents the conclusions of the study.

2 Mathematical formulation of the marginal risk constrained portfolio optimization problem model

The mathematical formulation of the marginal risk constrained portfolio optimization problem model is as shown below.

If N is the number of assets in the universe, μ_i the expected return of the asset *i* and σ_{ij} the covariance between the returns of assets *i* and *j* and *Wi* are the weights to be invested in asset *i* thenthe expected portfolio return is given by

$$\sum_{i=1}^{N} W_i \mu_i \qquad (1)$$

and the risk is given by

$$\sum_{i=1}^{N}\sum_{j=1}^{N} W_i W_j \sigma_{ij} \qquad (2)$$

The bi-objective portfolio optimization function formulated as a single criterion optimization function in what is known as the *weighted formulation* is given by

$$\mathrm{Min}\left(\lambda \sum_{i=1}^{N}\sum_{j=1}^{N} W_i W_j \sigma_{ij} - (1-\lambda)\sum_{i=1}^{N} W_i \mu_i\right) \qquad (3)$$

where λ is the *risk aversion parameter.* When λ tends to 0, the objective function maximizes returns, in other words, shifts weights towards stocks that yield high returns and when λ tends to 1, the objective function minimizes risk, in other words, shifts weights towards stocks with minimum risk.

The basic and bounding constraints imposed on the problem model are given as:

$$\sum_{i=1}^{N} W_i = 1,\ \lambda \in [0,1] \text{ (basic constraints)} \qquad (4)$$

$$-a < W_j < b, \quad \text{(bounding constraints)} \qquad (5)$$

The marginal risk constraint is defined as

$$W_i . m_i \quad \leq \quad x\% \ of \ \sigma_P, \quad i = 1,2,...N, \qquad (6)$$

(marginal risk constraint)

where,

portfolio risk

$$\sigma_P = \sqrt{\bar{W}'.V.\bar{W}} \qquad (7)$$

V is the variance-covariance matrix,

x% is the risk tolerance limit

and $\bar{m}$, the Marginal Contribution to Risk is given by,

$$\bar{m} = (m_1, m_2, ...m_N)' = \frac{(V.\bar{w})}{\sqrt{\bar{w}'.V.\bar{w}}} \qquad (8)$$

As can be observed the problem model turns out to be non convex quadratic constrained quadratic programming problem that looks up to metaheuristic methods for its solution. To tackle the nonlinear constraint representing the marginal risk control (equations (6-8)), the original mathematical formulation was revised to include Joines and Houck's [1994] dynamic penalty functions. The revised formulation of the objective function and the marginal risk constraint is as follows:

$$\min\left((\lambda \sum_{i=1}^{N}\sum_{j=1}^{N} W_i W_j \sigma_{ij} - (1-\lambda)\sum_{i=1}^{N} W_i \mu_i) + \psi(\bar{W}, \bar{m}, t)\right) \qquad (9)$$

where $\psi(\bar{m}, \bar{m}, t)$ the constraint violation function is given as,

$$\psi(\bar{W},\bar{m},t) = (C.t)^{\alpha}\left(\sum_{k=1}^{N} G_k.\left(\phi_k(W_k,m_k)\right)^{\beta}\right) \quad (10)$$

$$\varphi_k(W_k,m_k) = W_k.m_k \quad - \quad x\% \ of \ \sigma_P$$

G_k is the Heaviside Operator such that

$$G_k = 0, \quad for \quad \varphi_k(w_k,m_k) \le 0, \quad \text{and}$$
$$= 1, \quad for \quad \varphi_k(w_k,m_k) > 0 \quad (11)$$

Here C, α, β are coefficients on which the quality of the solution depends and *t* is the generation counter.

The revised mathematical formulation for the marginal risk constrained portfolio optimization problem is given by equations (9-11) as the objective function and equations (4-5) as its constraints.

Thus, while the notorious marginal risk constraint was tackled using dynamic penalty functions, the rest of the constraints viz., basic, bounding and short selling constraints were tackled using *weight standardization functions* which on the whole, enabled the metaheuristic solution strategies to maneuver the population of chromosomes/individuals through feasible solution space rather than candidate solution space.

3 Metaheuristic methods for the solution of the portfolio optimization problem

In this section, the two metaheuristic methods of Evolution Strategy with Hall of Fame and Differential Evolution with Hall of Fame, have been detailed.

3.1 Evolution Strategy with Hall of Fame

The Evolution Strategy with Hall of Fame (ES HOF) is a population based search strategy that belongs to the genre of Evolutionary algorithms. ES HOF is an adaptation of the Evolution Strategy (ES) proposed by Pai and Michel [2009] to solve a complex constrained portfolio optimization problem.

The ES HOF like any of its evolutionary algorithm based counterparts generates a random initial population of chromosomes and employs the genetic inheritance operators of Arithmetic Variable point Cross Over and Real number Uniform Mutation [Osyszka, 2002] to generate the offspring population of chromosomes. In each generation, the best fit among the μ parents ($s \neq 0$) and λ offspring ($u \neq 0$) are selected for the next population with size N, in the ratio of s:u such that

$$s + u = N \text{ and } s,\ u \neq 0 \quad (12)$$

The fitness function is set as the objective function and thus that chromosome with the best fitness function value is the one that yields the minimum objective function value. Also, to extend elitism in time, the algorithm adopts the mechanism of Hall of Fame which accommodates the best among the best of the chromosomes that have been generated this far. During each generation, the best among the parent and offspring chromosomes in the population compete to enter the Hall of Fame. On satisfying the convergence criterion, when the evolutionary algorithm terminates, that chromosome in the Hall of Fame yields the solution to the optimization problem.

3.2 Differential Evolution (rand/1/bin) with Hall of Fame

Differential Evolution (rand/1/bin) with Hall of Fame (DE HOF) though population based like its Evolutionary algorithm based counterparts, belongs to a different genre of metaheuristics for they are quite different from them with regard to their search processes [Engelbrecht, 2007]. The Differential Evolution based metaheuristics unlike their Evolution Algorithm counterparts employ distance and direction functions modeled using vector differentials, for their search processes.

DE (*rand/1/bin*) is a category of Differential Evolution [Storn and Price, 1997] where *rand* indicates the random selection of target vectors, *1* the number of difference vectors used and *bin* the binary cross over method used.

The DE HOF algorithm begins with an initial population of individuals. Two control parameters viz., Scaling factor $\beta \in (0,\infty)$ which controls the amplification of the differential variations and p_r the probability of recombination which has a direct influence on controlling the diversity of the population, are initialized.

The mutation operator produces a trial vector $\bar{u}_i(t)$ for each individual of the current population by mutating a target vector $x_a(t)$ with a weighted differential as given below:

$$\bar{u}_i(t) = x_{i1}(t) + \beta(x_{i2}(t) - x_{i3}(t)) \qquad (13)$$

where

$x_{i1}(t),\ x_{i2}(t)$ and $x_{i3}(t)\ i \neq i1, i2 \neq i3$
are randomly chosen individuals.

The trial vector is then used by the crossover operator to produce offspring. In this work we employ a binary cross over operator given as below:

$$x_{ij}'(t) = \begin{cases} u_{ij}(t), & j \in \tau \\ x_{ij}(t), & otherwise \end{cases} \qquad (14)$$

Here $x_{ij}(t)$ and $u_{ij}(t)$ are the j^{th} elements of the vectors and which are the parent vector $\bar{x}_i(t)$ and $\bar{u}_i(t)$ trial vector respectively, τ is the set of element indices that undergoes perturbation and $x_{ij}'(t)$ is the j^{th} element of the offspring $\bar{x}_i'(t)$.

A *deterministic selection* where the better of the two, for each pair of parent and offspring individuals chosen from the parent and offspring population, is selected for stepping into the next generation, is adopted for creating the new generation of population. The best individual amongst those in the new population competes to enter the Hall of Fame which accommodates the best individual found thus far. The new population is now treated as the current parent population and the generations proceed until the convergence criterion is met with.

When the algorithm terminates, that individual in the Hall of Fame yields the optimal solution to the optimization problem.

4 Constraint Handling

The marginal risk constrained portfolio optimization problem includes the basic, bounding, short selling and marginal risk constraints.

The marginal risk constraint which is a non-linear constraint is tackled using Joines and Houcke's [1994] penalty function strategy. To tackle the rest of the constraints a *weight standardization procedure* had to be evolved.

Metaheuristic methods adopt weight standardization where the portfolio weights adjust among themselves to ensure that the entire population of parents or offspring satisfy their appropriate constraints, as a consequence of which the metaheuristic strategy is propelled to traverse through feasible solution space rather than candidate solution space.

Handling constraints using weight standardization approaches have already been reported in the literature. Thus, while Chang *et al.*, [2000] devised an elegant method of weight adjustment to ensure that the population set satisfied bounding constraints, Pai and Michel [2009] devised refined weight adjustment algorithms for a class of portfolio optimization problems that involved class constraints, constraints related to popular investment strategies such as 130-30 investment strategy [Pai and Michel, 2012], Risk Budgeting [Pai and Michel, 2011] and Equity Market Neutral Portfolios [Pai and Michel, 2012].

In this work, the weight standardization algorithms work to enable the population of chromosomes/individuals satisfy their basic, bounding and short selling constraints. Considering the long-short nature of the portfolio set, the standardization involved adjustment of weights in the real space. Both the metaheuristic methods viz., ES HOF and DE HOF resorted to the same weight standardization algorithms to propel them traverse through feasible solution space.

Algorithm PORTFOLIO_WEIGHT_STDZN() is an adaptation of the generic weight standardization algorithm discussed by Pai and Michel [2012] to tackle the basic and bounding constraints of a long-short portfolio which however was employed to solve a different portfolio optimization problem, viz., optimization of 130-30 long-short portfolios. We briefly review the two major functions of the algorithm that serves to handle the constraints.

The PORTFOLIO_WEIGHT_STDZN() algorithm first checks for the feasibility of the bounds. If $\sum_i a > 1$ or $\sum_i b < 1$ where *(a,b)* are the bounds interval, then the solution is deemed infeasible. Thereafter, the following two major functions viz., *LOW_BOUNDS_STDZN* and *UP_BOUNDS_STDZN* work to enable each chromosome/individual satisfy their respective lower and upper bounds subject to the basic constraint that the sum of the weights must equal 1.

(i) *LOW_BOUNDS_STDZN():* Here weights are adjusted in such a way that $w_i \geq a$ and $\sum_i w_i = 1$ for each chromosome/individual of the population.

(ii) *UP_BOUNDS_STDZN():* Here weights are adjusted in such a way that $w_i \leq b$ and $\sum_i w_i = 1$ The bounds (a,b) are set by the investor.

***LOW_BOUNDS_STDZN* function:** For a given set of weights $\bar{W} = (w_1, w_1, w_1, \ldots w_N)$ in a chromosome or individual, let Q denote those weights that satisfied their lower bounds and R those weights w_t which fell short of their lower bounds and hence are upgraded to their lower bounds (i.e.) $w_t = a$.

Let $T = \sum_{s \in Q} |w_s|$ be the absolute sum of weights in Q and $F = 1 - \sum_{i=1}^{N} a$ be the free proportion of weights F, available after according the N weights their minimum due subject to the basic constraint $\sum_{i=1}^{K} w_i = 1$. The free proportion F is now redistributed to the weights in Q proportional to their existing weights as $w_s = a + |w_s| \cdot \frac{F}{T}, s \in Q$ if $T \neq 0$, or equally distributed as $w_s = a + \frac{F}{|Q|}, s \in Q$ if $T = 0$.

The adjustment ensures that each of the weights satisfy their lower bounds subject to the basic constraint that the sum of the weights equals 1. ***UP_BOUNDS_STDZN* function:** The*UP_BOUNDS_STDZN*ensures that the weights satisfy their upper bounds subject to the basic constraint that the sum of the weights equals 1.

Let Q denote the index set of those weights which satisfy their upper bounds and R denote the index set of those weights $w_t$which exceeded their upper bounds and therefore have been levelled off such that w_t =*b*.

Compute $T = \sum_{s \in Q} |w_s|$, the absolute sum of weights in Q, and $F = 1 - \sum_{s \in Q} a - \sum_{t \in Q} b$, the free proportion of weights F, available after according those weights belonging to R their maximum due (*b*) and those weights in Q their minimum due (*a*), subject to the basic constraint that the sum of weights equals 1.

Distribute the free proportion F to the weights in Q, proportional to their existing weights as $w_s = a + |w_s| \cdot \frac{F}{T}, s \in Q, if T \neq 0$, or assign it completely to an arbitrarily chosen weight $w_p, p \in Q$ such that $w_p = F$, otherwise.

If the adjustment still leaves weights exceeding their upper bounds, repeat the process by levelling of such weights to their maximum upper bound, migrating these weights to the set R and redistributing the new free proportion of weights amongst the current set of weights in Q, until there are no more such weights left in Q.

At the end of the adjustments, the weight set w_i i=1,2, $\ldots N$ satisfy both their upper and lower bounds subject to the basic constraint that the sum of weights equals 1.

During each generation, each of the parent and offspring chromosome population standardize their weights to satisfy their respective constraints thereby facilitating a metaheuristic search through feasible solution space.

5 Metaheuristic optimization of the marginal risk constrained portfolios

In this section we detail the optimization of the marginal risk constrained portfolios by the two metaheuristic methods viz., ES HOF and DE HOF.

The common inputs to the problem solution irrespective of the metaheuristic strategy, are the mean returns μ_i the variance-covariance matrix V for the N assets in the portfolio, the risk budget x%, the risk aversion parameter value λ and the population size M. The coefficients C, α, β required for the penalty function strategy are also initialized.

The convergence criterion fixed for both the metaheuristic strategies was the number of generations.

5.1 ES HOF based portfolio optimization

The ES HOF begins by generating a random initial population of chromosomes each of which represents a set of portfolio weights. The population is standardized to satisfy their basic, bounding and short selling constraints by invoking the function PORTFOLIO_WEIGHT_STDZN().

Set fit_{HOF} the fitness value of the chromosome occupying the Hall of Fame to ∞(a large number).

Compute $\psi(\bar{W}, \bar{m}, t)$ the constraint violation function using equations (10-11) and the fitness function values for each chromosome in the population using equation (9). Record the penalty function values for the chromosome.

Generate offspring population by undertakingthe Arithmetic Variable point Cross overoperation with a cross over rate ρ and Real Number Uniform Mutation for a specified mutation rate τ. Standardize the offspring population by invoking PORTFOLIO_WEIGHT_STDZN(). Compute the penalty function values and the fitness function values of the offspring population.

The best of the parent and offspring population in the ratio s:u, are selected as the members for the new generation.

ES HOF now triggers its generation cycle by reproducing offspring chromosomes, standardizing them, enabling the best of the offspring to compete with the chromosome in the Hall of Fame and exporting the best of the parent and offspring population to the subsequent generation, until the convergence criteria is met with.

When the strategy terminates, that chromosome in the Hall of Fame yields the optimal solution. The portfolio weights obtain the optimal risk return couple for the specific risk aversion parameter.

Algorithm
Marginal_Risk_PortfolioOptmzn_**ESHOF()**details the working of the ES HOF in obtaining the solution to the marginal risk constrained portfolio optimization problem.

5.2 DE HOF based portfolio optimization

The DE HOF begins its operation by initializing the control parameters β (scaling factor) and p_r(probability of recombination). An initial random population of individuals are generated. The parent population is standardized to satisfy their basic, bounding and short selling constraints by invoking PORTFOLIO_WEIGHT_STDZN().fit_{HOF} the fitness value of the individual occupying the Hall of Fame is initialized to ∞ (a large number).

DE HOF invokes the mutation operator to generate the trial vectors and the cross over operator to generate the offspring population. The offspring population is standardized using PORTFOLIO_WEIGHT_STDZN()and the penalty function values and the fitness function values are recorded.

The deterministic selection operator is invoked to select the best of the parent and offspring population to the next generation. The best among the choices made competes to enter the Hall of Fame.

DE HOF now triggers its generation cycle by generating new offspring individuals from the new generation selected, standardizing them, selecting the best among the parent and the offspring to the next generation while allowing the best among the choices made to compete with the Hall of Fame until the convergence criterion is met with.

Algorithm
Marginal_Risk_PortfolioOptmzn_**DEHOF()** details the working of the DE HOF in obtaining the solution to the marginal risk constrained portfolio optimization problem.

6 Experimental studies

This section details the various experiments that were undertaken to analyze the results and performance of the two metaheuristic methods. The studies were undertaken on the Bombay Stock Exchange BSE 200 data set (March 1999-March 2009). The risk budget was fixed at 12.5% of the total portfolio risk. The portfolio considered was a large portfolio with 30 assets.

6.1 Experiment 1: *Testing individual performance consistency of ES HOF and DE HOF over different runs for the marginal risk constrained portfolio optimization problem*

In this experiment the performance consistency of the two metaheuristic methods were tested over various runs for a specific portfolio set. Considering the fact that metaheuristic search strategies evolve out of a randomly generated population of chromosomes/individuals, it has turned out to be essential to study the consistency of behaviour of the strategies and hence the need for this study. Tables 1 and 2 show the control parameters set by ES HOF and DE HOF respectively, during the runs.

Both the strategies were implemented over a specific portfolio set of 30 assets and the results observed for various runs for 51 different values of the risk aversion parameter $\lambda \in [0,1]$. For each value of λ, the optimal chromosome/individual available in the Hall of Fame was extracted and the corresponding annualized risk (%) and expected annual portfolio return (%) values were computed. Figure 1 and Figure 2 show the plots of the risk-return couples graphed for various values of λ , for a specific portfolio of assets, by ES HOF and DE HOF respectively. A visual inspection does reveal the proximity

```
Algorithm Marginal_Risk_PortfolioOptmzn_ESHOF()
Obtain mean returns μ_i and the variance-covariance matrix V for the N
assets in the portfolio;
Initialize coefficients  C,α, β;
Set risk budget x%, risk aversion parameter value λ, population size M;
Set Hall of Fame to null and initialize fit_HOF the fitness value of the
chromosome occupying the Hall of Fame to ∞ ( a large number)
i = 0;         /* Set generation index  i to  0 */

Randomly generate initial (parent)  population of size M;

Standardize the parent population by invoking PORTFOLIO_WEIGHT_STDZN();
            /* P_i = {x_1, x_2, ....x_M} represents a set of feasible solutions to the
          marginal risk constrained portfolio optimization problem*/

Compute the penalty functions and fitness function values of the
standardized parent population P_i;
Perform Arithmetic Variable point Cross over with a cross over rate ρ
and Real Number Uniform Mutation for a  specified mutation rate τ, to
generate the first offspring population;
Standardize the offspring population by invoking
PORTFOLIO_WEIGHT_STDZN();
               /* Let O_i represent the feasible offspring solution set*/

Compute the penalty functions and fitness values of the offspring
population  O_i ;/* the parent and the first generation offspring are
          ready for reproduction*/
repeat
          i = i+1;
          Reproduction: Select the high fit parent chromosomes and
          offspring chromosomes from population P_(i-1) and O_(i-1) respectively
          in the ratio of s:u (s<u)to form the mating pool;
          Call the current pool of chromosomes P_i ;

          Perform Arithmetic Variable point Cross over with a cross over
          rate ρ and Real Number Uniform Mutation for a    specified
          mutation rate τ, to generate the offspring population;
          Standardize the offspring population by invoking
          PORTFOLIO_WEIGHT_STDZN();
          Compute the penalty functions and fitness values of the
          feasibleoffspring population O_i and let fit_offspring be the
          fitness value of the best fit chromosome O_t among the population
          of offspring O_i ;
          if((fit_offspring<fit_HOF)and the penalty functions of O_t is zero)
          then
          induct O_t into the Hall of Fame and set fit_HOF = fit_offspring;
until convergence criterion is met with;
Obtain optimal solution from the chromosome in the Hall of Fame and
fit_HOF ;
endMarginal_Risk_PortfolioOptmzn_ESHOF()
```

Algorithm Marginal_Risk_PortfolioOptmzn_**DEHOF()**

Obtain mean returns μ_i and the variance-covariance matrix V for the N assets in the portfolio;

Initialize coefficients C,α, β;

Set risk budget x%, risk aversion parameter value λ, population size M;

Set Hall of Fame to null and initialize fit_{HOF} the fitness value of the chromosome occupying the Hall of Fame to ∞ (a large number)

t = 0;/ Set generation count t to 0 */*

Initialize the control parameters : β (scaling factor) and p_r(probability of recombination);

Generate an initial population of individuals $P^{(t)}$representative of the portfolio weights;

Standardize the parent population $P^{(t)}$ by invoking PORTFOLIO_WEIGHT_STDZN();

Compute the penalty functions and fitness function values $F^{(t)}$ of the standardized parent population $P^{(t)}$;

while termination condition is not satisfied **do**

  Compute the trial vector $U^{(t)}$ by applying the mutation operator;

  Create the offspring population $O^{(t)}$ by applying the Cross over operator;

  Compute penalty function values and fitness values $F'^{(t)}$of the offspring population $O^{(t)}$;

  Apply selection operator on $F^{(t)}$and $F'^{(t)}$ to determine which individuals will move to the next generation $P^{(t+1)}$;
  t = t+1; */* increment generation count */*

  Retain penalty function and fitness function values $F^{(t)}$of the new population $P^{(t)}$;

  Choose the best fit individual $I^{(best)}$ from $P^{(t)}$ and let $I^{(fit)}$ be its fitness value;

  if ($I^{(fit)}<fit_{HOF}$)**and** *(the penalty function values of* $I^{(best)}$are zero)

  then

  induct $I^{(best)}$into the Hall of Fame and set$fit_{HOF} = I^{(fit)}$;
  endif

endwhile

  Return the individual in the Hall of Fame and fit_{HOF} as the optimal solution;

endMarginal_Risk_PortfolioOptmzn_**DEHOF()**

of the results obtained during various runs.

Table 1. Control parameters of the ES HOF strategy for marginal risk constrained portfolio optimization problem

Population Size (P)	300
Chromosome length	30
Number of generations	8000
Risk Aversion parameter(λ) values	51 points between [0,1]
Cross over rate	0.61
Mutation rate	0.01
Parent child ratio for reproduction (s:u)	1:2

Table 2. Control parameters of the DE HOF strategy for marginal risk constrained portfolio optimization problems

Population Size (P)	300
Chromosome length	30
Number of generations	8000
Risk Aversion parameter(λ) values	51 points between [0,1]
Scaling factor (β)	0.5
Probability of recombination (p_r)	0.87

6.2 Experiment 2: *Performance comparison of the optimal results yielded by ES HOF and DE HOF*

In this experiment the optimal risk return couples yielded by ES HOF and DE HOF methods were compared. The two meta-heuristic methods were executed for different runs over different portfolios with different choices of assets made by the investor.

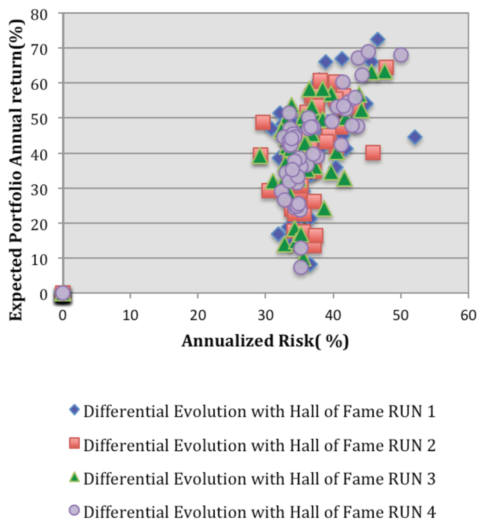

Figure 1. Plot of optimal risk-return couples for various runs using DE-HOF for a specific marginal risk constrained portfolio (N=30) of BSE200 data set (March 1999- March 2009), for various values of risk aversion parameter $\lambda \in [0,1]$

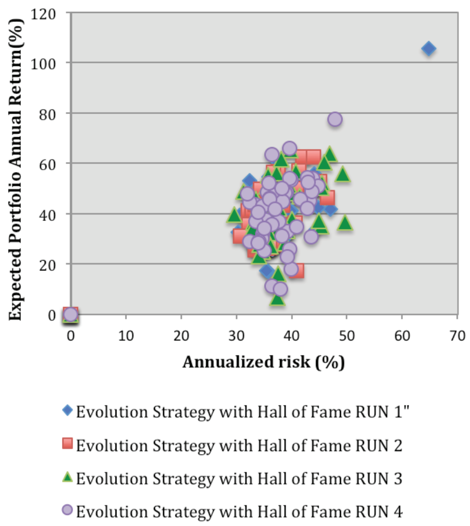

Figure 2. Plot of optimal risk-return couples for various runs using ES-HOF for a specific marginal risk constrained portfolio (N=30) of BSE200 data set (March 1999- March 2009), for various values of risk aversion parameter $\lambda \in [0,1]$

Pai and Michel [2007] devised an efficient method where k-means clustering was adopted to tackle the notorious cardinality constraint that was turning the cardinality constrained portfolio optimization problem into a mixed integer problem rendering it difficult for direct solving using traditional methods. After several experimental studies, the characteristics of mean and covariance of daily returns of the assets were zeroed down upon to cluster the original universe of N assets into K groups, for an investor's choice of K. The set of assets chosen from each of the K clusters, termed *investable universe* not only served to eliminate the cardinality constraint and simplify the mathematical model but also served to yield *reliable* portfolio sets, which was extensively demonstrated in their experimental studies.

In this experiment the universe of assets was *k*-means clustered into 30 groups and various investable universes were selected, choosing one asset from each of the clusters. The two metaheuristic methods were run over the investable universes for various runs.

Figure 3 and Figure 4 show the results obtained by the two methods for a run over two investable universes of 30 assets. A visual inspection indeed shows the proximity of the solutions obtained by the two methods considering the fact that both the methods hail from different genres of metaheuristics.

6.3 Experiment 3: *Convergence behavior of ES HOF and DE HOF strategies for the marginal risk constrained portfolio optimization problem*

The convergence behavior of a metaheuristic algorithm conventionally refers to the convergence of the objective function that is being minimized. Figure 5 and Figure 6 show a plot of the fitness function values (minimal objective function value attained by a population) obtained during each generation, for a specific run of the ES HOF and DE HOF strategies respectively on a specific marginal risk constrained portfolio for risk aversion parameter λ=0.001. A visual inspection of the plots show the convergence of the fitness function values.

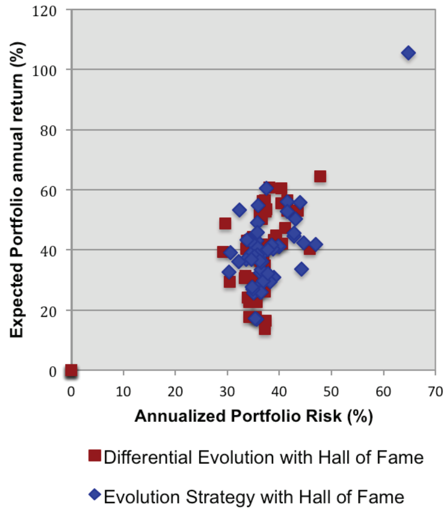

Figure 3. Plot of optimal risk-return couples for a specific run of ES HOF and DE-HOF for a specific marginal risk constrained investable universe of 30 assets (Investable Universe 1) of BSE200 data set (March 1999- March 2009), for various values of risk aversion parameter $\lambda \in [0,1]$

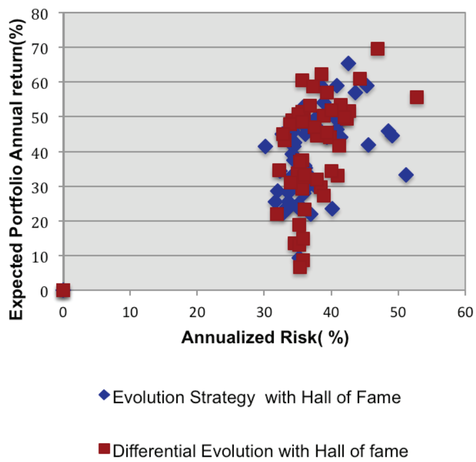

Figure 4. Plot of optimal risk-return couples for a specific run of ES HOF and DE-HOF for a specific marginal risk constrained investable universe of 30 assets (Investable Universe 2) of BSE200 data set (March 1999- March 2009), for various values of risk aversion parameter $\lambda \in [0,1]$

Alternately, Vitaliy Feoktistov[2006] introduced a new performance measure termed P-measure that served to expand the convergence measure of the objective function. P-measure represents the dynamics of grouping the individuals/ chromosomes of the population around the optimum. In other words, the more densely individuals are populated in a generation, the better convergence is intended. P-measure thus measures the radius of the population as follows:

The Euclidean distance measure between the centre of the population and the individual farthest from it is determined. The centre of the population (*barycentre)* is calculated as the average vector of all individuals in the population whose size is N,

$$C_p = \sum_{i=1}^{N} \frac{d_i}{N} \tag{15}$$

where d_i are the individuals that comprise the population. The P-measure is defined as

$$P_m = \max \left\| d_i - C_p \right\|_{Euclidean} \quad i = 1,2,\ldots N \tag{16}$$

Figure 7 and Figure 8 show the P-measures computed for the DE-HOF and ES-HOF algorithms for the same marginal risk constrained portfolio, for a risk aversion parameter value λ=0.001 for the same run graphed in Figure 4.

6.4 Experiment 4: *Comparison of the technical efficiencies of the optimal portfolios yielded by DE HOF and ES HOF*

A visual inspection of the plot of the risk return couples of the optimal portfolios, obtained by DE HOF and ES HOF and the convergence behavior of the two metaheuristic strategies, more or less displayed similar behaviour, despite the fact that the two metaheuristic strategies belonged to two different genres of metaheuristics.

Hence in this experiment, the technical efficiencies of the optimal portfolios obtained by DE HOF and ES HOF were studied using Data Envelopment Analysis (DEA).

DEA (Charnes et al., 1978) is a non-parametric, deterministic methodology that serves to measure the relative efficiencies of comparable functional units called has emerged as a powerful and analytical tool in measuring the relative efficiency of similar functional units. DEA is a non-parametric, deterministic methodology for determining the relatively efficient production frontier, by assessing the relative efficiencies and performances of a collection of comparable entities called Decision Making Units (DMUs) which transform inputs to outputs. DEA analyzes each DMU individually, yielding its efficiency score relative to the other DMUs in the entire set by making use of linear programming to determine the efficiency scores of each DMU relative to the others and hence is computationally intensive.

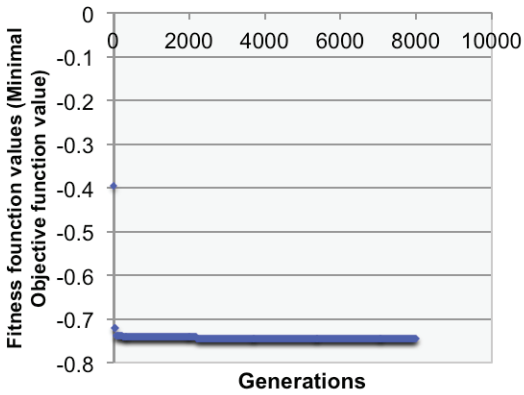

Figure 5. Convergence of the objective function:Trace of fitness function values (Minimal objective function values) obtained during the generations for a specific run of DE HOF for a specific marginal risk constrained portfolio, for $\lambda = 0.001$

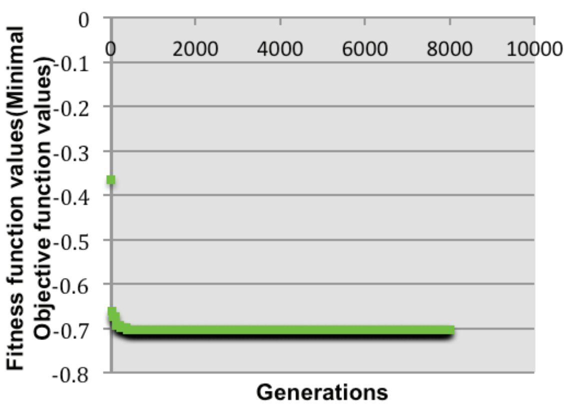

Figure 6. Convergence of the objective function: Trace of fitness function values (Minimal Objective function values) obtained during the generations for a specific run of ES HOF for a specific marginal risk constrained portfolio, for $\lambda = 0.001$

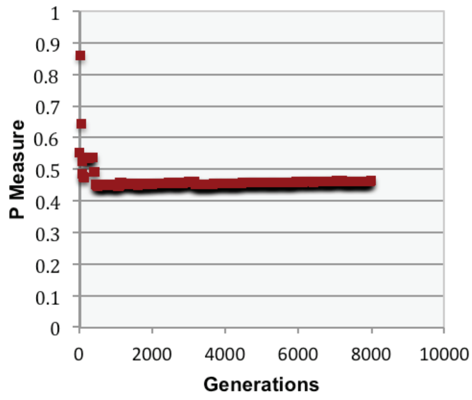

Figure 7. Population convergence (P-Measure): Trace of P-Measures obtained during the generations for various runs of DE HOF for a specific risk budgeted global asset allocation portfolio, for $\lambda = 0.001$

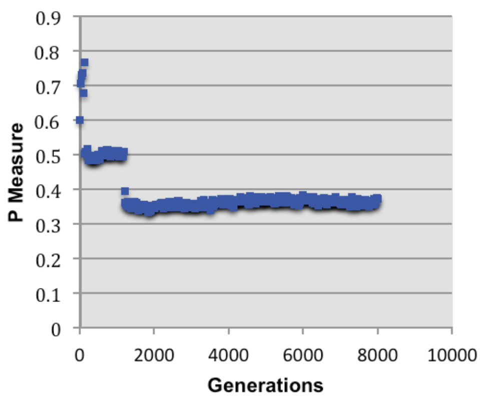

Figure 8. Population convergence (P-Measure): Trace of P-Measures obtained during the generations for various runs of ES HOF for a specific risk budgeted global asset allocation portfolio, for $\lambda = 0.001$

Though DEA encompasses various approaches, this work employs the BCC model (Banker, Charnes and Cooper, 1984) built on variable returns to scale. The linear programming problem representing the same, is given below:

$$
\begin{aligned}
&\min(\theta) \\
&\text{subject to} \\
&-y_i + Y.\lambda \geq 0 \\
&\theta.x_i - X\lambda \geq 0 \\
&\sum \lambda = 1 \\
&\lambda \geq 0, \quad \theta : \mathit{free}
\end{aligned}
\tag{17}
$$

where θ is the efficiency score, λ are the dual variables, (x_i, y_i) is the amount of input utilized and output produced respectively by the DMU i whose efficiency is to be computed and $(X,Y) = ((x_1,y_1),(x_2,y_2),\ldots(x_n,y_n))$ are the amounts of input utilized and output produced, by each of the n DMUs in the system.

As can be inferred, the linear programming system will have to be solved n times to determine the efficiency score of each of the nDMUs. In general, an efficiency score of 1 implies the DMU is considered efficient.

For the marginal risk constrained portfolio optimization problem, to measure the efficiency of portfolios using a DEA, each risk return couple of the optimal portfolios obtained for various risk aversion parameter $\lambda \in [0,1]$, by the two metaheuristic strategies, were treated as a DMU with risk as the input utilized and return as the output produced by it. The DMUs of all the competing risk return couples are put together and the efficiency score of each of the DMUs relative to all others in the set are obtained by solving equation (17).

The objective behind employing DEA for measuring the technical efficiency of portfolios was to investigate the quality of the portfolios yielded by DE HOF and ES HOF. Thus, for a specific portfolio set, all risk return couples obtained for the optimal portfolios during various runs were pooled together as the set of DMUs. The efficiency scores of each of the DMUs in each of the sets were obtained by solving the linear programming system of equation (14).

Table 3 illustrates a summary of the statistical measures of dispersion for the efficiency scores of the optimal portfolios for a specific portfolio of the BSE200 data set obtained by DE HOF and ES HOF.

Using the observations listed in Table 3, thestatistical hypotheses that there is no significant difference in the average efficiency scores of the two competing methods viz., DE HOF and ES HOF was accepted at 1% level of significance using a test statistic z for large samples.

Table 3. Summary of the DEA analysis of marginal risk constrained optimal portfolios of the BSE200 data set, obtained by DE HOF and ES HOF

	DE HOF	ES HOF
Number of DMUs (n) considered by the DEA	204	204
Average efficiency score (μ)	0.6411	0.6224
Standard deviation of the efficiency scores (σ)	0.1838	0.1648
Coefficient of Variation	3.4873	3.7757

A statistical testing of hypotheses using paired t-tests was applied for the efficiency scores obtained by the DEA, for the DE HOF and ES HOF based optimal portfolios. The objective was to find out if there was any significant difference in the means of the efficiency scores with regard to the two metaheuristic based methods, and if so to find out which of the two methods was efficient. The null hypotheses and alternate hypotheses set to a confidence level of 95% were:

H0: There is no significant difference in the means of the efficiency scores yielded by DE HOF and ES HOF

H1: There is a significant difference in the means of the efficiency scores yielded by DE HOF and ES HOF.

The test revealed that there was no significant difference in the means of the efficiency scores yielded by DE HOF and ES HOF concluding that both the metaheuristic methods despite belonging to different genres of metaheuristics had only equalled themselves in their efficiency.

7 Conclusions

In this work, a metaheuristic optimization of marginal risk constrained portfolios was undertaken. Two metaheuristic strategies, viz., DE HOF and ES HOF belonging to two different genres of evolution computation and augmented with weight standardization procedures to enable efficient search through feasible solution space and ensure faster convergence, were applied to solve the problem. The experimental studies were undertaken on portfolio sets selected from the Bombay Stock Exchange data set (BSE 200 data set March 1999-March 2009).

The conclusions are:

i Marginal risk constrained portfolio optimization problem together with basic, bounding and long-short portfolio constraints turns itself into a non-convex quadratic constrained quadratic programming problem, difficult for direct solving using analytical methods. The metaheuristic methods of DE HOF and ES HOF both belonging to different genres of evolution computation have served to find solutions to the complex problem within reasonable time and compuational effort.

ii Both the metaheuristic methods augmented with weight standardization procedures served to navigate themselves only through feasible solution space rather than candidate solution space thereby ensuring faster convergence.

iii Both the metaheuristic methods reported consistency of performance during their individual runs.

iv For a specific portfolio set, the risk return couples of the optimal portfolios yielded by the DE HOF and ES HOF strategies for a range of risk aversion parameter values were in close proximity to one another.

v A DEA to measure the technical efficiencies of the portfolios yielded by the two metaheuristic methods,resulted in both the metaheuristic strategies reporting matching performance.

vi A statistical testing of hypothesis of the efficiency scores obtained by the two methods only revealed that there was no significant difference in their mean efficiency scores for a confidence level of 95%.

Future work would revolve round solving the marginal risk constrained portfolio optimization problem as a constrained multi-objective optimization problem.

8 Acknowlegements

This research was carried out under the grant provided by University Grants Commission, New Delhi, INDIA, Major Research Project 2010, F.No. 39-125/2010(SR) dated 28-12-2010.

9 References

1. Andries Engelbrecht, *Computational Intelligence*, John Wiley, 2007.

2. Andrzej Osyczka, *Evolutionary algorithms for single and multicriteria design optimization*, Physica –Verlag, 2002.

3. Banker R D, A Charnes and W W Cooper, Some models for estimating technological and scale inefficiencies in Data Envelopment Analysis, Management Science, vol. 30, no.9, 1078-1092, 1984.

4. T J Chang, N Meade, J B Beasley and Y M Sharaiha, Heuristics for cardinality constrained portfolio optimization, *Computers and Operations Research*, 27: 1271-1302, 2000.

5. Charnes A, W W Cooper and E Rhodes, Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2, 429-444, 1978.

6. E J Elton, M J Gruber, S J Brown and W N Goetzmann, *Modern Portfolio Theory and Investment analysis,* Wiley & Sons, In., Hoboken, NJ, 6^{th} ed., 2003

7. Fernandez Alberto and Sergio Gomez, Portfolio selection using neural networks, *Computers and Operations Research*, 34, 1177-1191, 2007.

8. Gilli Manfred and Enrico Schumann, *Heuristic optimization in financial modelling*, SSRN: http://ssrn.com/abstract=1277114, September 29, 2008.

9. R C Grinold, and R N Kahn, *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*, McGraw-Hill, New York, 1999.

10. J A Joines and C R Houck, On the use of non-stationary penalty functions to solve nonlinear constrained optimization problems with GAs, *Proceedings of the First IEEE Conference on Evolutionary Computation*, pp. 579-584, 1994.

11. G Kendall and Y Su, A particle swarm optimization approach in the construction of optimal risky portfolios, in *Proc. Of the 23^{rd} International Multiconference on Artificial Intelligence and Applications (IASTED, 2005*) pp. 140-145, 2005.

12. Maringer Dietmar, *Portfolio management with heuristic optimization*, Springer, 2005.

13. Markowitz H M, Portfolio selection, *The Journal of Finance*, 7(1), 77-91, 1952.

14. Shusang Zhu, Duan Li and Xiaoling Sun, Portfolio selection with marginal risk control, *Journal of Computational Finance*, 14(1), 3-28, 2012.

15. Storn, R. and Price, K., Differential Evolution - a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Journal of Global Optimization*, Kluwer Academic Publishers, Vol. 11, pp. 341 – 359, 1997.

16. Streichert Felix, Holger Ulmer and Andreas Zell, Evolutionary algorithms and the cardinality constrained portfolio optimization problem, *Intl. Conf. on Operations Research*, pp. 253-260, Springer, 2003.

17. Thomaidis Nikos, Timotheos Angelidis, Vassilios Vassiliadis, Georgios Dounias. Active portfolio management with cardinality constraints: An application of particle swarm optimization, *New Computational Methods for Financial Engineering* (Spl. Issue), *Journal of New Mathematics and Natural Computation*, November 09, 5(3), pp. 535-555, DOI No. 10.1142/S1793005709001519, 2009.

18. Pai Vijayalakshmi G A , Thierry Michel. Evolutionary optimization of constrained k-means clustered assets for diversification in small portfolios, *IEEE Transactions on Evolutionary Computation*, 13(3), pp. 1030-1053, 2009.

19. Vijayalakshmi Pai G A and Thierry Michel, Integrated metaheuristic optimization of 130-30 investment strategy based long-short portfolios, *Intelligent Systems in Accounting, Finance and Management*, 19, 43-74, Blackwell-Wiley, 2012.

20. Vijayalakshmi Pai G A and Thierry Michel, Differential Evolution based Optimization of Evolutionary Risk Budgeted Equity Market Neutral Portfolios, Proc. IEEE World Congress on Computational Intelligence (IEEE WCCI 2012), *2012 IEEE Congress on Evolutionary Computation*, pp. 1888-1895, Brisbane, Australia, June 2012.

21. Vijayalakshmi Pai G A and Thierry Michel, Evolutionary optimization of Risk Budgeted Long-Short Portfolios, Proc. IEEE Symposium Series in Computational Intelligence (IEEE SSCI 2011): *2011 IEEE symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr 2011),* pp. 59-66, Paris, France, April 2011.

22. Vitaliy Feoktistov, *Differential Evolution : In search of solutions*, Springer, 2006.