Some Theoretical and Experimental Observations on Permutation Spaces and Similarity Search [*]

Giuseppe Amato, Fabrizio Falchi, Fausto Rabitti, and Lucia Vadicamo

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo",
via G. Moruzzi 1, Pisa 56124, Italy
{firstname}.{lastname}@isti.cnr.it

Abstract. Permutation based approaches represent data objects as ordered lists of predefined reference objects. Similarity queries are executed by searching for data objects whose permutation representation is similar to the query one. Various permutation-based indexes have been recently proposed. They typically allow high efficiency with acceptable effectiveness. Moreover, various parameters can be set in order to find an optimal trade-off between quality of results and costs.
In this paper we studied the permutation space without referring to any particular index structure focusing on both theoretical and experimental aspects. We used both synthetic and real-word datasets for our experiments. The results of this work are relevant in both developing and setting parameters of permutation-based similarity searching approaches.

Keywords: permutation-based indexing, similarity search, content based image retrieval

1 Introduction

Representing dataset objects as lists of preselected pivots ordered by their closeness to each object is a recent approach that have been proved to be very useful in many recent approximate similarity search techniques [3,8,14,20]. These approaches share the intuition that similarity between objects can be approximated by comparing their representation in terms of permutations. The quality of the obtained results have proved that whenever the permutations of two objects are similar then the two objects are likely to be similar also with respect to the original distance function.

In this paper, we studied the permutation space withouth relying on any specific indexing structure with the goal of making theoretical and experimental observations that can be of help in both setting parameters of existing permutation based approaches and developing new one.

[*] Preprint of https://doi.org/10.1007/978-3-319-11988-5_4

2 Related Work

Predicting the closeness between objects on the basis of ranked lists of a set of pivots was originally and independently proposed in [8] and [4]. In [8] data objects and queries are represented as appropriate permutations of a set of reference objects, called *permutants*, and their similarity is approximated by comparing their representations in term of permutations. As distance between permutations, Spearman rho, Kendall Tau and Spearman Footrule were tested. Spearman rho revealed better performance.

The MI-File approach [4,3] uses an inverted file to store relationships between permutations. Spearman Footrule Distance is used to estimate the similarity between the query and the database objects. To reduce the storage, each object is encoded using the only nearest reference points and further approximations and optimizations are adopted to improve both efficiency and effectiveness.

The Permutation Prefix Index (PP-Index), was proposed in [13,14]. PP-Index associates each indexed object with a short prefix of predefined length of the full permutation. The prefixes are indexed by a *prefix tree* kept in main memory and all the relevant information relative to the indexed objects are serialized sequentially in a *data storage* kept on disk. PP-index uses the permutations prefixes in order to quickly retrieve a candidate set of objects that are likely to be at close distance to the query. The result set is then obtained using the original distance function by a sequential scan of the candidate set.

In [20], the concept of Locality-sensitive Hashing (LSH) was extend to a general metric space by using a permutation approach. In [19], a quantized representation of the permutation lists with its related data structure was proposed and a specific data structed, namely the Metric Permutation Table, was also defined. In [22] authors presented the *neighboord approximation* (NAPP) techinique whose main idea is to represent each object by the set of its nearest pivots and approximate the similarity between objects on the basis of the number of shared pivots. Three strategies for parallelization of permutation-based indexes using inverted files were presented in [18]. Posting lists decomposition, reference points decomposition, and multiple independent inverted files were studied and compared.

In [2], various pivot selection techniques were tested on three permutation-based indexing approaches (i.e., [8,3,14]). The results revealed that each indexing approach has its own best selection strategies but also that the random selection of pivots, even if never the best, results in good performance.

In [17,1] a Surrogate Text Representation (STR) derivated from the MI-File has been proposed. The conversion of the permutations in a textual form allows using off-the-shelf text search engines for similarity search.

3 Permutation-based representation

Given a a domain $\mathcal{D}$, a *distance function* $d : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ and a fixed set of objects $P = \{p_1 \ldots p_n\} \subset \mathcal{D}$ that we call pivots, we define a permutation-based representation Π_o (briefly permutation) of an object $o \in \mathcal{D}$ as the list of pivots identifiers ordered by their closeness to o, with the pivots being a fixed set of objects.

Formally, the permutation-based representation $\Pi_o = (\Pi_o(1), \Pi_o(2), ..., \Pi_o(n))$ lists the pivot identifiers in an order such that $\forall j \in \{1, 2, \ldots, n{-}1\},\ d(o, p_{\Pi_o(j)}) \leq d(o, p_{\Pi_o(j+1)})$, where $p_{\Pi_o(j)}$ indicates the pivot at position j in the permutation associated with object o.

Denoting the position of a pivot p_i, in the permutation of an object $o \in \mathcal{D}$, as $\Pi_o^{-1}(i)$ so that $\Pi_o(\Pi_o^{-1}(i)) = i$, we obtain an equivalent representation Π_o^{-1}:

$$\Pi_o^{-1} = (\Pi_o^{-1}(1), \Pi_o^{-1}(2), ..., \Pi_o^{-1}(n))$$

This representation is very useful for essentially two reasons: first, $\Pi_o^{-1} \in \mathbb{R}^n$ allowing representing permutation in the Cartesian coordinate system; second, the Euclidean distance between two objects x, y represented as Π_x^{-1} and Π_y^{-1} is equivalent to the Spearman rho distance between Π_x and Π_y (see Section 3.1).

3.1 Comparing permutations

The idea of approximating the distance $d(x, y)$ between any two objects $x, y \in \mathcal{D}$ by comparing their permutation-based representation $\Pi_x, \Pi y$ was originally proposed in [8]. As distance between permutations, Spearman rho, Kendall Tau and Spearman Footrule were tested. Spearman rho revealed better performance. Given two permutations Π_x and Π_y, Spearman rho is defined as:

$$S_\rho(\Pi_x, \Pi_y) = \sqrt{\sum_{1 \leq i \leq n} (\Pi_x^{-1}(i) - \Pi_y^{-1}(i))^2}$$

Following the intuition that the most relevant information of the permutation Π_o is in the very first, i.e. nearest, pivots, Spearman rho distance with location parameter $S_{\rho,l}$ defined in [15], intended for the comparison of top-l lists, has been also proposed.

$S_{\rho,l}$ differs from S_ρ for the use of an inverted truncated permutation $\tilde{\Pi}_o^{-1}$ that assumes that pivots further than $p_{\Pi_o(l)}$ from o being at position $l + 1$. Formally, $\tilde{\Pi}_o^{-1}(i) = \Pi_o^{-1}(i)$ if $\Pi_o^{-1}(i) \leq l$ and $\tilde{\Pi}_o^{-1}(i) = l + 1$ otherwise.

It is worth to note that only the first l elements of the permutation Π_o are needed, in order to compare any two objects with the $S_{\rho,l}$.

4 Theoretical observations

As mentioned in Section 3, the permutation-space representation Π_o^{-1} belongs to $\mathbb{R}^n$. Moreover, the Spearman rho distance between two permutations Π_x and Π_y results in a Euclidean distance between Π_x^{-1} and Π_y^{-1}. In the following we consider the Π_o^{-1} representation in a Cartesian coordinate system.

If we consider the case $n = 3$, the set of all possible permutation-based representation (i.e., the set of all permutations on 3 elements) is formed by

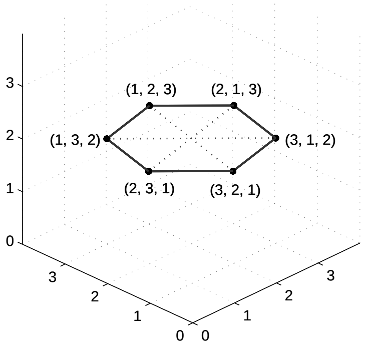


Fig. 1. The six points in $\mathbb{R}^3$ obtained by permuting the coordinate of the vector $(1, 2, 3)$

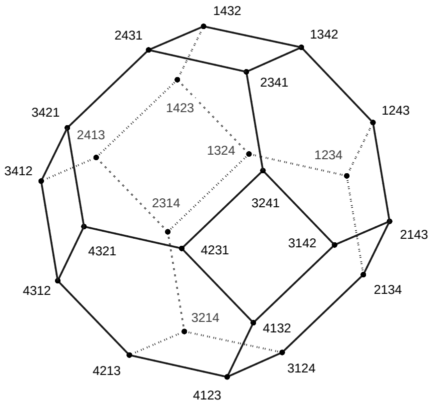


Fig. 2. Permutahedron with $4! = 24$ vertices

$\{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$. It is easy to see that all this points lie on the plane $x + y + z = 6$ and represent the vertices of a regular hexagon as depicted in Figure 1.

Consider now the $n = 4$ case: the vectors of all possible Π_o^{-1} lie in a three-dimensional subspace of $\mathbb{R}^4$ and are the vertices of a truncated octahedron (see Figure 2).

In general, the $n!$ points x obtained by permuting the coordinates of the vector $(1, 2, \ldots, n)$, form the vertices of a $(n-1)$-dimensional polytope embedded in a n-dimensional space, referred to as *permutahedron* (also spelled *permutohedron*) [23,16]. In fact, given that both the sum of vector values x_i (i.e., $\Pi_o^{-1}(i)$) and their squared values are fixed, all the vertices lie on both the hyperplane

$$x_1 + x_2 + \cdots + x_n = \frac{n(n+1)}{2}$$

the $n-$sphere

$$x_1^2 + x_2^2 + \cdots + x_n^2 = \frac{n(n+1)(2n+1)}{6}.$$

That is they lie on the intersection between an hyperplane and a sphere both in $\mathbb{R}^n$, i.e., on a $n-1$ sphere residing in n-dimensional space.

The permutahedron is a very interesting convex polytope. It is centrally symmetric and its vertices can be identified with the permutation of n objects in such a way that two vertices are connected by an edge if and only if the corresponding permutations differ by an adjacent transposition. It is rather easy to see that the squared Euclidean distance between any two vertices is an even integer, moreover, for $n > 4$, the squared distances constitute every even integer up through the maximum possible value, that is $\frac{1}{3}(n^3 - n)$ [21,23].

As observed in [21], standing on any vertex of a permutahedron and looking around at neighbouring vertices, the view of the surrounding space is the same:

there would be $n-1$ adjacent vertices evenly distributed around the observation vertex, which Euclidean distance is $\sqrt{2}$. Furthermore, the number of vertices and their relative positions within a generic ϵ-ball neighbourhood is independent of the observation vertex.

The *permutahedron* precisely illustrate how the permutation-based representation are positioned in the space were the Euclidean distance is equivalent to the Spearman rho. It is worth to mention that the Spearman Footrule, sometimes used in permutation based-indexing, results in a L1 (also Manattan) distance in the same space. However, it does not help very much in understanding the distance distribution.

In order to understand the Spearman rho distance distribution it is useful to use its not-squared root variant (S_ρ^2) because of its interesting distribution properties. In [11] it was shown that S_ρ^2 distance has:

- mean: $\frac{1}{6}(n^3 - n)$
- variance: $\frac{1}{36}n^2(n-1)(n+1)^2$
- maximum value: $\frac{1}{3}(n^3 - n)$

Unfortunately, S_ρ^2 is not a metric. However, due to the monotony of the square root function, there are not changes in the order of the results of a k-NN search with respect to the ones that can be obtained with S_ρ. Moreover, normalized by its means and variance, S_ρ^2 has a limiting normal distribution [12]. Chávez's intrinsic dimensionality [10] of the permutation space with squared Spearman rho distance is $\frac{1}{2}(n-1)$.

5 Performance evaluation of the permutation space

For our experiments we did not use any specific index approach. In fact, we performed sequential scan of permutation-based representation archives in order to retrieve most similar objects with respect to the query by using the Spearman rho distance function.

5.1 Datasets and Groundtruth:

Random float vectors As synthetic dataset we considered random generated vectors of floats of various dimensionalities d between 2 and 10. For each dimension we randomly generated float between 0 and 1. As distance measure for comparing any two vectors we used the Euclidean distance.

CoPhIR As real-word dataset we used CoPhIR dataset [7], which is the largest multimedia metadata collection available for research purposes. It consists of 106 millions images crawled from Flickr. We run experiments by using as distance function d a linear combination of the five distance functions for the five MPEG-7 descriptors that have been extracted from each image. We adopted the weights proposed in [5]. As the ground truth, we have randomly selected 100 objects from the dataset as test queries and we have sequentially scanned the CoPhIR to compute the exact results. The queries were removed from the dataset itself.

5.2 Pivots selection

For the CoPhIR dataset we randomly selected 10,000 pivots from the whole 106M objects collection. We then created subsets of this first selection. In the following we report experiments obtained on a subset of the entire CoPhIR collection. Thus it happens that some pivots are also in the dataset while some are not.

Pivots for the random float vectors were randomly generated without selecting between the objects in the dataset.

Variuos pivots selection strategies have been proposed for permutation-based indexing [2]. Experimental results have shown that while each specific index strategies have its own best selection approach, the random selection is always a good choice.

5.3 Parameters

In this section we summarize the parameters that have to be set for each specific experiment.

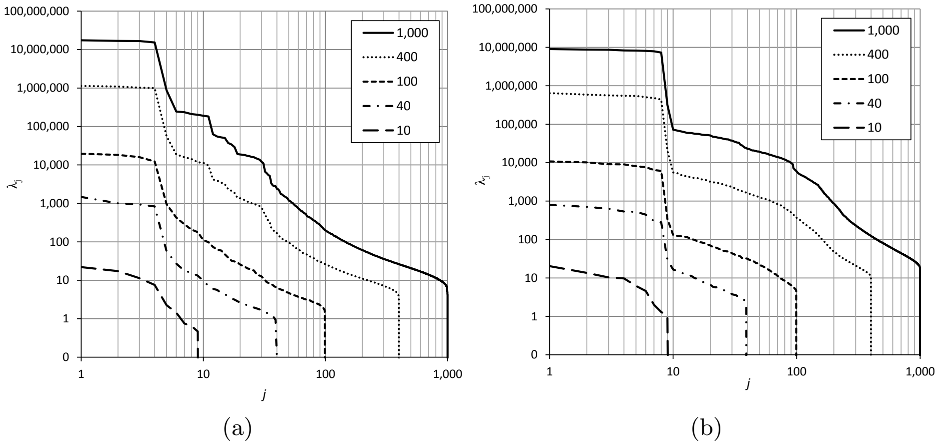
***d* - float vectors dimensionality** This parameter is only necessary to indicate which random float vector dataset was used for the specific experiment. Experiments are reported for $d = 2, 4, 6, 8$.

***m* - dataset size** For both the synthetic and the CoPhIR dataset we recursively selected a subset of the collection. We performed experiments up to 1M and 10M objects for the random float vectors and CoPhIR datasets respectively.

***n* - number of pivots** The max number of pivots we used was 10,000. The smallest set of pivots have been obtained recursively selecting a subset of the larger collection.

***l* - permutation length** Various values of l for the Spearman rho with location parameter (see Section 3.1) where tested. Please note that $l = n$ results in the standard Spearman rho distance.

***a* - amplification factor** When a k-NN search is performed, a candidate set of results of size $k' = a * k$ is retrieved considering the similarity of the permutations based on S_ρ. This set is then reordered considering the original distance $d : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$.



(a) (b)

Fig. 3. Variances (eigenvalues) $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, for various number of pivots n, corresponding to each principal component of the permutation obtained from the random float vectors of dimensionality 4 (a), 8 (b)

5.4 Evaluation Measure

Permutation-based indexing approaches, typically re-rank a set of approximate results using the original distance. In this work we did the same. Thus, if the k-NN results list $\tilde{\mathcal{R}}_k$ returned by a search technique has an intersection with the ground truth $\mathcal{R}_k$, the objects in the intersection are ranked consistently in both lists. The most appropriate measure to use is then the *recall*: $|\tilde{\mathcal{R}}_k \cap \mathcal{R}_k|/k$. In the experiments we fixed the number of results k to 10.

5.5 Principal Component Analysis

While PCA can not be performed on a generic domain $\mathcal{D}$ that can have a non metric distance and/or being a non vector space, once the permutation-based representation has been obtained it is always possible to run PCA on the $\mathit{\Pi}_o^{-1}$. We did this for both the random float vectors and CoPhIR dataset.

In Figure 3, we show the eigenvalues of each principal component of the permutations obtained for various number of pivots n. The dimensionality of the float vectors was 4 for (a) and 8 for (b). Please note, that both axes have logarithmic scale. With 1,000 pivots it is clear in both cases what the original dimensionality of the vector space was. In fact, there is a large drop in the eigenvalues passing from the 4th and 5th eigenvectors in (a), and from 8th and 9th in (b). The results also show that with more pivots we obtain a permutation-based representation that better fix the original data complexity.

We did the same for the CoPhIR dataset reporting the results in Figure 4. It is interesting to see that, in the logarithmic scale, the eigenvalues linearly decrease. However, CoPhIR did not reveal any specific dimensionality.

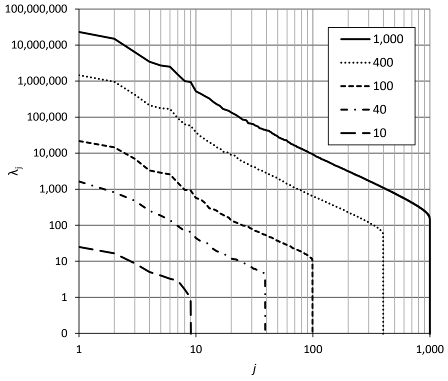


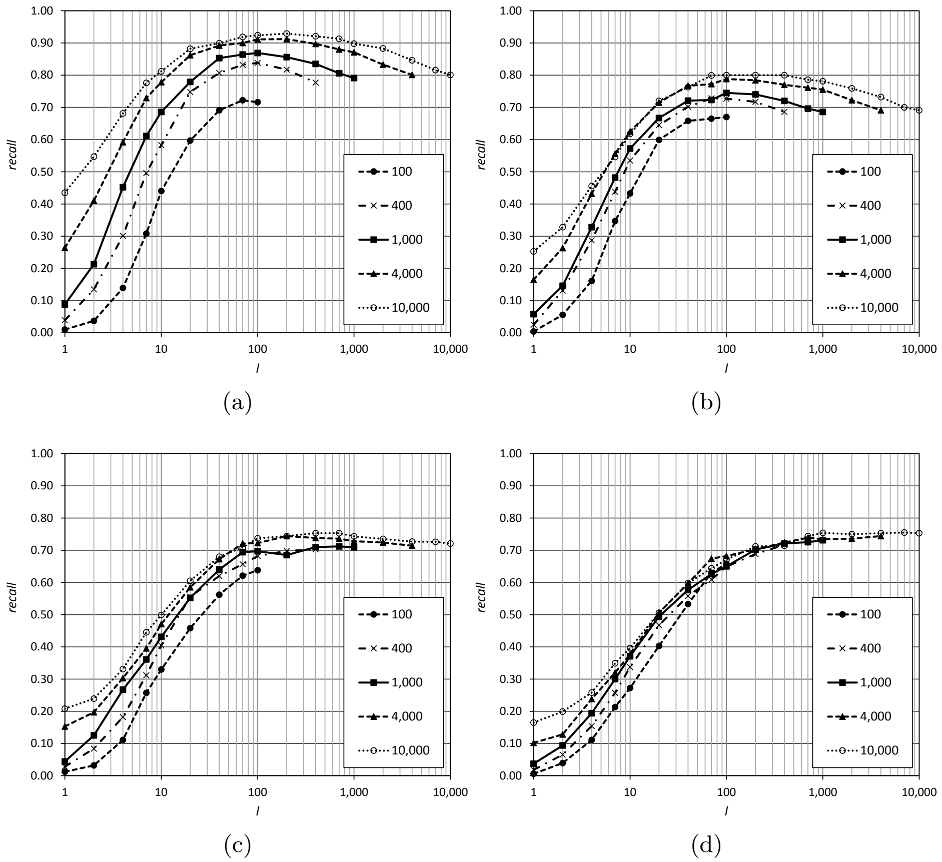
Fig. 4. Variances (eigenvalues) $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, for various number of pivots n, corresponding to each principal component of the data obtained from the mapping in the permutation space of the CoPhIR 30,000 objects

In [6], it was shown that the combined distance function that we are also using in our experiments, results on the CoPhIR dataset in a near normal distribution with an intrinsic dimensionality, measured following the approach presented in [9], of about 13. Unfortunately, the same information can't be induced from Figure 4. Some non-linearity can be seen around 6 and 9, but performing PCA on the CoPhIR doesn't allow to understand the intrinsic dimensionality of the dataset as well as it allowed to understand the real dimensionality of the random generated float vectors.

5.6 Recall

In this section we relate the various parameters presented in 5.3 to the *recall* obtained on k-NN searching for $k = 10$. As mentioned before, results were obtained sequentially scanning archives of permutations by using the Spearman rho with and without location parameter l. Please note that $l = n$, i.e. for location parameter l equal to the number of pivots, the Spearman rho with an without location parameter are equivalent.

In Figure 5, we report the *recall* obtained on the random float vector datasets of 2 (a), 4 (b), 6 (c), 8 (d) dimensionalities, varying the location parameter l and for various number n of pivots. In these experiments we fixed the amplification factor $a = 1$. The most interesting result is that for small dimensionalities (2 and 4) there is a maximum *recall* that can be obtained varying l. In other words, $l = n$ it is not always the best solution, but there is an optimal l that appears not to vary for $n > l$. It also interesting to see that this optimal l varies significantly with the dimensionality of the original vector space. For 8 dimension vectors we are not even able to see this effect in the results. Probably, in this case the optimal l is well above 10,000 which is the max number of pivots we



(a) (b)

(c) (d)

Fig. 5. *Recall* varying *l* for various number of pivots obtained on 100,000 random float vectors of dimensionality 2 (a), 4 (b), 6 (c), 8 (d)

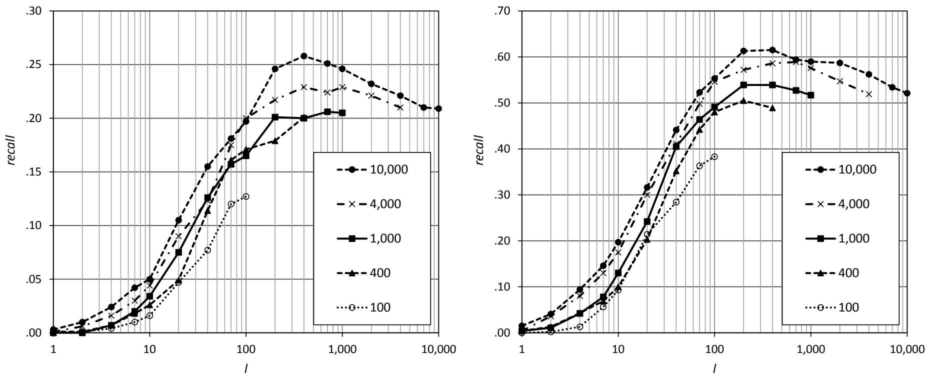
tried. Another important observation is that the differences between the recall obtained by the various set of pivots tend to be smaller for higher d.

The same type of experiments were conducted on the CoPhIR dataset. In Figure 6, we report the recall obtained for $a = 1$ (a) and $a = 10$ (b). As for the random float vectors, it appears to be an optimal l that does not vary significantly with n. While the amplification factor does significantly impact the overall *recall*, the optimal l still remain almost the same. These results are consistent with the ones obtained on the random float vectors for dimensionality of about 4. In terms of indexability with respect to the permutation-based approach, CoPhIR appears to be as complex as random generated vectors of dimensionality between 4 and 6. In fact, we shown in Figure 5 that for random float vectors of 8 dimensions, the optimal l equals the number of pivots.

In Figure 7, the *recall* obtained varying the number of pivots for various size of CoPhIR subsets is reported. In this case we use $a = 1$. In Figure 7 (a), we show the results obtained for $l = n$ (i.e., the standard Spearman rho). In Figure 7 (b), we report the recall obtained for the optimal l which depends on both n and dataset size. Comparing these two figures it is evident that higher *recall* can be obtained increasing the number of pivots only if the optimal location parameter l is used. However, in our experiments, we had very near optimal results by using $l = 200$ (as can be seen for 10M objects in Figure 6). The intuition is that after a certain number of pivots, information regarding distant pivots is not only useless but distracting. Pleas note that the experiments performed on the random vectors indicate that the distant pivots are useful when the dimensionality of the dataset is above 8 (up to 10,000 pivots). Thus, while the observations made on the CoPhIR datasets are useful for understanding its characteristics and the fact that it exists an optimal l for a specific dataset, $l = 200$ is a near optimal solution only for the CoPhIR dataset and it probably reflects its complexity which appears to be lower than the intrinsic dimensionality evaluated in [6] following the [9] approach.

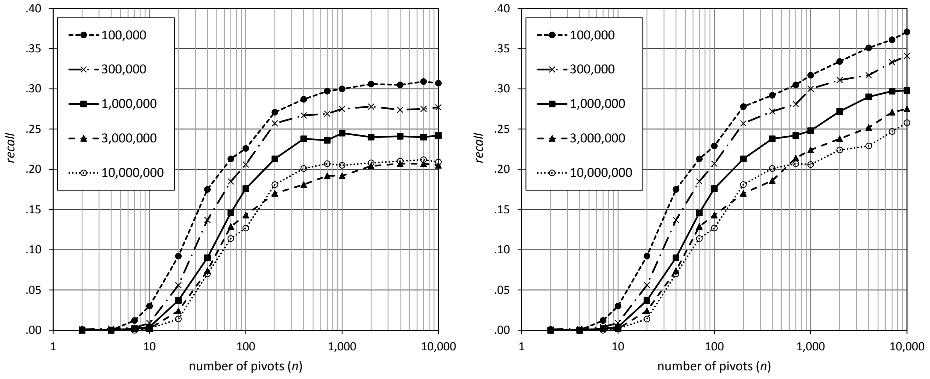
In Figure 8, we show the *recall* obtained varying the size of the CoPhIR subset for various number of pivots, optimal l (different for each combination of number of pivots n and dataset size) and $a = 10$. This graph is useful for understanding the loss in *recall* when the dataset increase. The results show that there is almost a linear dependency between the number of pivots needed to achieve a given quality of results and the dataset size.

In Figure 9, we fixed both the number of pivots ($10,000$) and the dataset size (10M) reporting the *recall* varying a for various l. As obvious, the larger the amplifier factor a the better the quality of the results. Please note that l and a are the most relevant parameters in trading efficiency versus effectiveness in permutation based indexes. In fact, the shorter the permutation Π_o, the fewer the information to be stored for each object. Moreover, the less the amplification factor a, the smaller the number of objects to be retrieved for each search.



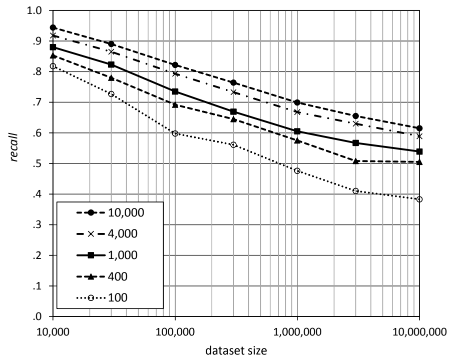
(a) (b)

Fig. 6. *Recall* varying location parameter l for various number of pivots and $a = 1$ (a) and 10 (b) on CoPhIR 10M objects



(a) (b)

Fig. 7. *Recall* varying the number of pivots for various dataset sizes (all subsets of CoPhIR) obtained without location paramter l (a) and with the optimum l (b).



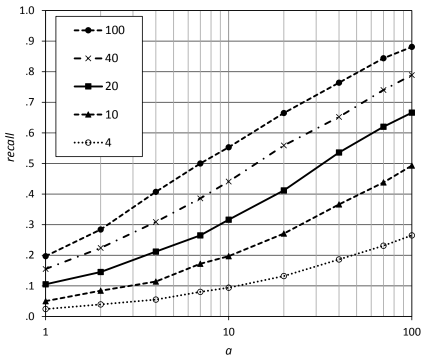


Fig. 8. *Recall* varying dataset size (all subsets of CoPhIR) for various l and $a = 10$.

Fig. 9. *Recall* varying a for various l, number of pivots $n = 10,000$ and size of the CoPhIR subsets 10M.

6 Conclusion

In this work we studied the permutation space focusing on both theoretical and experimental aspects not relying on any specific index structure. We used both synthetic and the CoPhIR dataset for the experiments varying various parameters that are typically used for trading-off between efficiency and effectiveness.

We first made some observations on the permutation space generating random permutations in order to understand its specific characteristic. We showed that the points are vertices of a permuthaedron, that using a squared Spearman rho results in Gaussian distance distribution.

The experiments conducted using random float vectors of various dimensionality shown that the complexity of the dataset affects the optimal value of l in terms of *recall* and that the dimensionality of the original vector space can be argued by performing PCA on the permutation space.

Also in the case of the CoPhIR dataset we found that it exists an optimal l for each specific number of pivots. Moreover, this optimal l is very stable and typically around 200. Thus, we believe that the optimal length of the permutations is in relation with the intrinsic complexity of the dataset even if this complexity can not be clearly seen performing PCA on the permutation space.

The experiments also revealed a linear dependency between the number of pivots and dataset size. Other results were shown considering l and amplifier factor a combination considering that they are the most useful parameters in trading-off efficiency and effectiveness in permutation indexes.

References

1. Amato, G., Bolettieri, P., Falchi, F., Gennaro, C., Rabitti, F.: Combining local and global visual feature similarity using a text search engine. In: Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on. pp. 49 –54. IEEE Computer Society (2011)
2. Amato, G., Esuli, A., Falchi, F.: Pivot selection strategies for permutation-based similarity search. In: Brisaboa, N., Pedreira, O., Zezula, P. (eds.) Similarity Search and Applications, Lecture Notes in Computer Science, vol. 8199, pp. 91–102. Springer Berlin Heidelberg (2013)
3. Amato, G., Gennaro, C., Savino, P.: Mi-file: using inverted files for scalable approximate similarity search. Multimedia Tools and Applications pp. 1–30 (2012)
4. Amato, G., Savino, P.: Approximate similarity search in metric spaces using inverted files. In: Proceedings of the 3rd international conference on Scalable Information Systems. pp. 28:1–28:10. InfoScale '08, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2008)
5. Batko, M., Falchi, F., Lucchese, C., Novak, D., Perego, R., Rabitti, F., Sedmidubsky, J., Zezula, P.: Building a web-scale image similarity search system. Multimedia Tools and Applications 47(3), 599–629 (2010)
6. Batko, M., Kohoutková, P., Novak, D.: CoPhIR image collection under the microscope. In: Skopal, T., Zezula, P. (eds.) Similarity Search and Applications, 2009. SISAP '09. Second International Workshop on. pp. 47–54. IEEE Computer Society (2009)

7. Bolettieri, P., Esuli, A., Falchi, F., Lucchese, C., Perego, R., Piccioli, T., Rabitti, F.: CoPhIR: a test collection for content-based image retrieval. CoRR abs/0905.4627 (2009)
8. Chávez, E., Figueroa, K., Navarro, G.: Effective proximity retrieval by ordering permutations. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30(9), 1647–1658 (2008)
9. Chávez, E., Navarro, G.: Measuring the dimensionality of general metric spaces. Department of Computer Science, University of Chile, Tech. Rep. TR/DCC-00-1 (2000)
10. Chávez, E., Navarro, G., Baeza-Yates, R., Marroquín, J.L.: Searching in metric spaces. ACM Computing Surveys 33(3), 273–321 (2001)
11. Diaconis, P.: Group representations in probability and statistics, Lecture Notes-Monograph Series, vol. 11. Institute of Mathematical Statistics (1988)
12. Diaconis, P., Graham, R.L.: Spearman's footrule as a measure of disarray. Journal of the Royal Statistical Society. Series B (Methodological) 39(2), 262–268 (1977)
13. Esuli, A.: MiPai: Using the PP-index to build an efficient and scalable similarity search system. In: Skopal, T., Zezula, P. (eds.) Similarity Search and Applications, 2009. SISAP '09. Second International Workshop on. pp. 146–148. IEEE Computer Society (2009)
14. Esuli, A.: Use of permutation prefixes for efficient and scalable approximate similarity search. Information Processing & Management 48(5), 889–902 (2012)
15. Fagin, R., Kumar, R., Sivakumar, D.: Comparing top k lists. In: Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 28–36. SODA '03, Society for Industrial and Applied Mathematics (2003)
16. Gaiha, P., Gupta, S.K.: Adjacent vertices on a permutohedron. SIAM Journal on Applied Mathematics 32(2), 323–327 (1977)
17. Gennaro, C., Amato, G., Bolettieri, P., Savino, P.: An approach to content-based image retrieval based on the lucene search engine library. In: Lalmas, M., Jose, J., Rauber, A., Sebastiani, F., Frommholz, I. (eds.) Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science, vol. 6273, pp. 55–66. Springer Berlin Heidelberg (2010)
18. Mohamed, H., Marchand-Maillet, S.: Parallel approaches to permutation-based indexing using inverted files. In: Navarro, G., Pestov, V. (eds.) Similarity Search and Applications, Lecture Notes in Computer Science, vol. 7404, pp. 148–161. Springer Berlin Heidelberg (2012)
19. Mohamed, H., Marchand-Maillet, S.: Quantized ranking for permutation-based indexing. In: Brisaboa, N., Pedreira, O., Zezula, P. (eds.) Similarity Search and Applications, Lecture Notes in Computer Science, vol. 8199, pp. 103–114. Springer Berlin Heidelberg (2013)
20. Novak, D., Kyselak, M., Zezula, P.: On locality-sensitive indexing in generic metric spaces. In: Proceedings of the Third International Conference on Similarity Search and Applications. pp. 59–66. SISAP '10, ACM (2010)
21. Santmyer, J.: For all possible distances look to the permutohedron. Mathematics Magazine 80(2), 120–125 (2007)
22. Tellez, E.S., Chavez, E., Navarro, G.: Succinct nearest neighbor search. Information Systems 38(7), 1019 – 1030 (2013)
23. Ziegler, G.M.: Lectures on Polytopes. Graduate Texts in Mathematics, Springer New York (1995)