A Score Fusion Method Using a Mixture Copula

Takuya Komatsuda⋆[1], Atsushi Keyaki[2], and Jun Miyazaki[2]

[1] Hitachi,Ltd. takuya.komatsuda.ur@hitachi.com
[2] Department of Computer Science, School of Computing,
Tokyo Institute of Technology
keyaki@lsc.cs.titech.ac.jp,miyazaki@cs.titech.ac.jp

Abstract. In this paper, we propose a score fusion method using a mixture copula that can consider complex dependencies between multiple relevance scores in order to improve the effectiveness of information retrieval. The combination of multiple relevance scores has been shown to be effective in comparison with a single score. Widely used score fusion methods are linear combination and learning to rank. Linear combination cannot capture the non-linear dependency of multiple scores. Learning to rank yields output that makes it difficult to understand the models. These problems can be solved by using a copula, which is a statistical framework, because it can capture the non-linear dependency and also provide an interpretable reason for the model. Although some studies apply copulas to score fusion and demonstrate the effectiveness, their methods employ a unimodal copula, thus making it difficult to capture complex dependencies. Therefore, we introduce a new score fusion method that uses a mixture copula to handle the complicated dependencies of scores; then, we evaluate the accuracy of our proposed method. Experiments on *ClueWeb'09*, a large-scale document set, show that in some cases, our proposed method significantly outperforms linear combination and others existing methods that use a unimodal copula.

Keywords: copulas, information retrieval, dependencies between relevance scores

1 Introduction

Given a user query, search systems calculate the relevance scores of documents with respect to the query and return a list of documents ranked by relevance. In order to improve search accuracy, many IR models that calculate relevance scores have been proposed [3,22,27,30,32–38]. Owing to the diverse and complex nature of the information needs of users, it is difficult to determine an appropriate IR model that always yields the most accurate search results. In order to address this challenge, many studies have combined multiple relevance scores obtained from multiple IR models [11, 18, 20].

Relevance scores can be combined using various approaches, such as function-based methods [2, 9, 10, 41, 43], learning to rank [6, 7, 21, 28], and score fusion

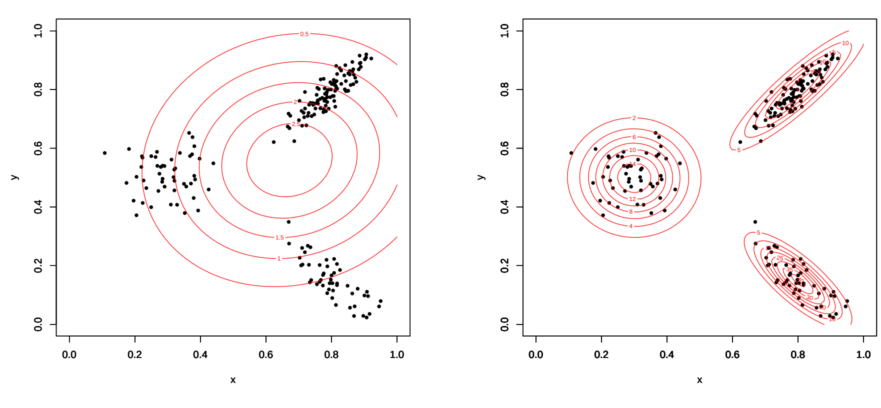
⋆ This work was conducted while the author was at Tokyo Institute of Technology.

methods using a copula [15]. Linear combination, which is one of the representative function-based methods, cannot capture the non-linear dependency between relevance scores. In addition, the output of learning to rank is complex with respect to understanding the model. These problems can be solved with a copula, which is a statistical framework used for analyzing complex multi-dimensional dependencies [15]. A copula is a model that represents the relationship between a multidimensional distribution and the marginal distributions. By applying a copula to a score fusion method, we can build a model that captures the non-linear dependency and is easy to understand intuitively.

Existing score fusion methods using a copula [15] cannot capture complex dependencies easily because these methods employ a unimodal copula that is assumed to model a unimodal distribution. For example, Figure 1 shows a distribution of two relevance scores. In the figure, each point denotes a document; the vertical axis represents the relevance scores x from an IR model X, and the horizontal axis represents the relevance scores y from a model Y. The set of documents exhibits some correlations locally around $(x, y) = (0.3, 0.5)$, $(0.8, 0.1)$, and $(0.8, 0.8)$. From Figure 1(a), the contour plot of the distribution estimated by using a unimodal copula cannot capture these correlations.

In this paper, we propose a score fusion method using a mixture copula that consists of multiple unimodal copulas. Mixture copulas can capture complex dependencies; therefore, they can estimate multimodal distribution accurately. Figure 1(b) shows that the distribution estimated by a mixture copula can capture complex dependencies between multiple relevance scores. Further, we evaluate the effectiveness of our proposed method by demonstrating that the consideration of complex dependencies can improve search accuracy.

Section 2 provides a basic introduction to copulas, and Section 3 reviews related work. Section 4 describes a score fusion method using a mixture copula, and Section 5 presents the evaluation method and results for the proposed method. In Section 6, the conclusion is stated, and plans for future work are described.



(a) Distribution using a Unimodal Copula **(b)** Distribution using a Mixture Copula

Fig. 1: Examples of Complex Distribution

2 Copulas

Before applying copulas to an IR model, we provide a basic introduction of copulas that have been used in other research fields such as finance. For more detail, refer to the book by Nelsen [25].

2.1 Definitions and Properties

Copulas are models that describe the relationship between a multivariate distribution and the marginal distributions. Let X be a k-dimensional random vector $X = (x_1, x_2, ..., x_k)$. Further, let a function $F_k(x)$ be a marginal cumulative distribution function for an element x_k of the random vector X, where $F_k(x) = P[X_k \geq x]$. Then, we can map X to a k-dimensional unit cube $[0, 1]^k$ as $U = (u_1, u_2, ..., u_k) = (F_1(x_1), F_2(x_2), ..., F_k(x_k))$. A k-dimensional copula C is described as a joint cumulative distribution function of the normalized random vector U. Most importantly, it has been proved that there exists a copula C that satisfies $F(x_1, x_2, ..., x_k) = C(F_1(x_1), F_2(x_2), ..., F_k(x_k))$ in any k-dimensional joint cumulative distribution function $F(x_1, x_2, ..., x_k)$ [25]. This general fact indicates the high applicability of copulas. In addition, copulas facilitate our analysis of the structure of joint distribution because we separately estimate each marginal distribution $F_k(.)$ and the dependency structure between the marginal distributions.

2.2 Copulas and Dependency of Relevance Scores

Let us introduce the constraint for copulas, assuming that the dependency between relevance scores is for extreme conditions such as independent, completely positive correlation, and completely negative correlation.

When the dependency between relevance scores is independent, the copulas are described as independent copulas C_{indep}.

$$C_{indep}(U) = \exp(-\sum_{i=1}^{k} -\log u_i)$$

Thus, independent copulas are equivalent to the product of all elements of U. It must be noted that while independence is frequently assumed in IR theory, it is a naive assumption.

When the dependency between relevance scores is a completely positive correlation, the copulas can be represented by the formula below.

$$C_{coMono}(U) = min\{u_1, u_2, ..., u_k\}$$

When the dependency between relevance scores is a completely negative correlation, the copulas can be represented by the formula below.

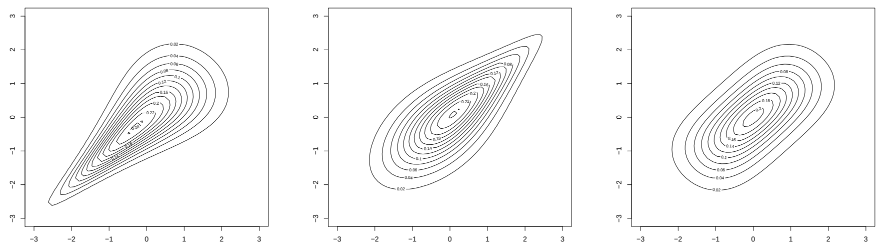
$$C_{counterMono}(U) = max\{\sum_{i=1}^{k} u_i + 1 - k, 0\}$$

Copulas have parameters that provide an interpretable reason for the dependency structure of marginal distributions. Therefore, we can understand joint distribution clearly. The parameters can be estimated by using maximum likelihood estimation or the Monte Carlo method [8].

2.3 Typical Families of Copulas

Families of copulas are of various types, such as elliptical copulas, Archimedean copulas and empirical copulas.

- **Elliptical Copulas**
  An elliptical copula is a copula derived from standard distribution, such as Gaussian distribution and t distribution. Equation (1) shows the formula for a Gaussian copula.
  $$C_{Gaussian}(U) = \Phi_{\Sigma}(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_k)) \tag{1}$$
  where Φ_{Σ} denotes a cumulative distribution function of standard normal distribution, and Φ^{-1} denotes its inverse function. A Gaussian copula requires a parameter $\Sigma \in R^{k\times k}$, which shows the observed covariance matrix.
- **Archimedean Copulas**
  Let ϕ be a continuous, strictly decreasing function from $\mathbf{I}$ to $[0, \infty]$ such that $\phi(1) = 0$. Then,
  $$C_{\phi}(U) = \phi^{-1}(\phi(u_1) + \phi(u_2) + ... + \phi(u_k)), U \in (0, 1]^k$$
  This formula represents a k-dimensional Archimedean copula, where ϕ is a generator of C_{ϕ}. For $\phi(t) = \frac{t^{-\theta}-1}{\theta}, (-logt)^{\theta}, -log\frac{e^{\theta t}-1}{e^{\theta}-1}$, the copulas are called Clayton copulas, Gumbel copulas and Frank copulas, respectively. Further, Clayton, Gumbel and Frank copulas are defined by Equations (2), (3) and (4), respectively, and their corresponding contour plots are shown in Figures 2(a), 2(b) and 2(c), respectively.
  $$C_{Clayton}(U) = (1 + \theta(\sum_{i=1}^{k} \frac{1}{\theta}(u_i^{-\theta} - 1)))^{\frac{-1}{\theta}} \tag{2}$$
  $$C_{Gumbel}(U) = \exp(-(\sum_{i=1}^{k}(-log(u_i))^{\theta})^{\frac{1}{\theta}}) \tag{3}$$
  $$C_{Frank}(U) = \frac{1}{\theta}log(1 + \frac{\prod_{i=1}^{k}(\exp(-\theta\ u_i) - 1)}{\exp((-\theta) - 1)^{k-1}}) \tag{4}$$
  As seen in Figure 2, different copulas have different features. For example, in Figure 2(a), we assume that for a Clayton copula, the dependency of the lower region is strong whereas the dependency of the upper region is independent. The use of a Clayton copula is effective if the dependency of relevance scores is strong in cases where relevance scores is low.
- **Empirical Copulas**
  An empirical copula refers to a copula that is derived from an empirical joint distribution whose marginal distributions are estimated by empirical distribution. Thus, an empirical copula is a nonparametric joint distribution that is based on observations, without assuming any specific distribution. A k-dimensional empirical copula $\hat{C}(U)$ is described as Equation (5)
  $$\hat{C}(U) = \frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{k}\mathbf{1}\{t_i^n \leq u_i\} \tag{5}$$



(a) Clayton Copulas **(b)** Gumbel Copulas **(c)** Frank Copulas

Fig. 2: Contour Plots of 2-dimension Joint Distribution using different Copulas

where N denotes the number of observations to estimate the empirical copula, and t_i^n represents a score of the i-axis of the n^{th} observation. The probability of a k-dimensional joint cumulative distribution derived from an empirical copula is calculated by dividing the number of training data, such that $(t_1^n \leq u_1, t_2^n \leq u_2, ..., t_k^n \leq u_k)$ by the number of all training data N.

We need to select an appropriate model from the various types of copulas. A model of copulas can be selected based on certain criteria such as tail dependence coefficient and rank correlation coefficient [17]. The tail dependence coefficient is an indicator of the dependence structure at the end points of the probability, i.e., for probability around 0 or 1. If we use the tail dependence coefficient for the selection of a model, it implies that we focus on the dependency between high relevance (or low relevance). The rank correlation coefficient is an indicator of the dependence structure in the entire distribution. If we select a model based on the rank correlation coefficient, it implies that we focus on the average dependency in the overall distribution.

2.4 Mixture Copula

A mixture copula is a copula that is composed of several copulas. By using a mixture copula, we can build a multimodal joint distribution that enables us to capture a complex dependency.

A mixture copula is described as the weighted sum of k copulas as shown in Equation (6).

$$C_{mix}(U) = \sum_{i=1}^{k} p_i C_i(U) \tag{6}$$

In order to construct a mixture copula, each copula C_i and its weight p_i should be estimated. These parameters can be estimated by using approaches based on clustering [12, 40]. Thus, we can construct a mixture copula by performing the following steps: **(1)** the training data to estimate the mixture copula is split into k clusters, **(2)** the data in each cluster is fit to a unimodal copula. The number of clusters k is determined in advance. One of the methods to decide the value of k is the usage of an information criterion such as AIC (Akaike Information Criterion) [5].

3 Related Work

3.1 IR models

Many IR models that calculate the relevance scores of a document have been proposed [3,22,36–38]. BM25, one of the classical probability models [30,32], has demonstrated high effectiveness [33]. Ponte and Croft proposed a probabilistic language model [27] which is developed in a mathematical framework, while vector space models [34,35] and classical probability models have been proposed as heuristic approaches.

3.2 Fusion of IR Models

Although many IR models exist, it is difficult to determine an appropriate model because the information needs of a user are diverse and complex. In order to address this challenge, various studies have focused on the fusion of multiple relevance scores calculated by several IR models. For example, some meta search engines have attempted to improve the accuracy by combining the results from multiple engines. These studies are known as score fusion of relevance scores [1, 11, 18, 20, 24].

The retrieval of structured documents such as XML posed a challenge in combining the structure information with the relevance scores of a document with respect to a query. Robertson et al. explain the difficulty in combination of document's structure information [31]. In the case of information retrieval for children, it is important to consider the credibility and readability of the document, as well as the relevance scores for a query [13].

3.3 Fusion of Relevance Scores

Advances of Score Fusion Although score fusion is often achieved by obtaining the sums or products of results from individual systems [18], probabilistic approaches also exist [11, 20]. Aslam and Montague proposed a probabilistic model based on ranking [1]. They improve their model by incorporating a majority method [24]. Further, they attempted to make the model robust against outliers by normalizing the scores [23].

Linear Combination Vogt et al. introduced linear combination in information retrieval [41]. The suitability of linear combination has been demonstrated [2, 9, 10, 43]. Gerani et al. applied nonlinear transformation to relevance scores before applying linear combination [19]. Gerani et al. showed that their method outperformed standard linear combination. This result demonstrates the need for a model that can capture complex and nonlinear dependency.

Learning to rank Learning to rank is a ranking model that uses machine learning and enables easy unification to obtain one score from a large number of document features [6, 7, 21, 28]. This approach extracts the features of relevant documents from a set of documents that are labeled as either relevant or irrelevant. The disadvantage of this approach is the difficulty in understanding the resulting model.

Copulas In general, copulas are widely used in quantitative finance and in portfolio management [4, 17]. Some recent studies have applied copulas in other research fields [26,29,39]. Vrac et al. applied a mixture copula to a global climate dataset and showed that a mixture copula can group the climate of the world correctly in terms of meteorology [42].

Eickhoff et al. applied copulas to score fusion and their proposed method outperformed the baselines as a result of combining two relevant features in some cases [15]. They verified the effectiveness of the approach when the number of relevant features was increased from 2 to 136. The result showed that their proposed method is more effective than linear combination as the number of relevant criteria increases [14]. In addition, they applied copulas to language models in which independence is frequently assumed. Their proposed method showed that it has competitive performance when compared with naive language models and some learning to rank methods [16].

Although the approaches proposed by Eickhoff et al. have demonstrated the effectiveness of copulas, their methods cannot estimate a joint distribution correctly when the distribution is complex, as shown in Figure 1(a), because the complexity of the dependency makes it difficult to estimate the joint distribution precisely. One of the solutions to the problem is the use of multiple copulas to estimate a multimodal distribution.

4 Proposed Approach

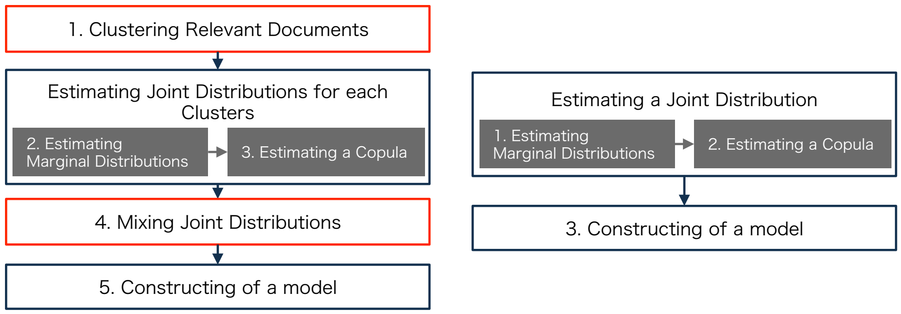
We propose a score fusion method that uses a mixture copula. We use mixture copulas to precisely estimate a joint distribution of relevant documents which has some strong correlations locally as shown in Figure 1(b). In Figure 1(b), for example, the distribution is accurately captured by three copulas, although its right two areas have strong correlations locally.

In our method, first, a mixture copula is estimated; then, models of score fusion are constructed by using the mixture copula. As we mentioned in Section 2.4, a mixture copula is composed of multiple unimodal copulas. During the process, clustering is useful in estimating a mixture copula [12, 40]. We use a clustering approach in our method.

The process for constructing the proposed model is described below;

1. Apply a clustering algorithm to relevant documents.
2. In each cluster, estimate a joint distribution using a unimodal copula.
3. Combine the unimodal copulas estimated in the previous step and construct a mixture copula.
4. Create a score fusion method by using the estimated mixture distribution.

Figure 3(a) shows the process for constructing a model by using our method, and Figure 3(b) shows the process that uses the method of Eickhoff et al [15]. Our proposed method begins with the clustering of relevant documents and estimates joint distribution of relevant documents in individual clusters using unimodal copulas, whereas Eickhoff et al. estimate a joint distribution of relevant documents using one unimodal copula.



(a) Our Method **(b)** Method of Eickhoff et al.

Fig. 3: Processes for Constructing a Model using Copulas

4.1 Clustering Relevant Documents

A group-average agglomerative clustering method is used as a clustering algorithm. The group-average agglomerative clustering method merges two clusters whose distance is the closest until the number of clusters reaches k determined in advance. The distance between two clusters is defined as the average of distances between all pairs of documents except for pairs from the same cluster. Although we employ a group-average agglomerative clustering method, any clustering algorithms can be used in our method.

4.2 Estimating Joint Distribution in Individual Clusters

In individual clusters, joint distributions are obtained by performing the following two steps: **(1)** estimating the marginal distributions, and **(2)** estimating the dependency structure between the marginal distributions by using a copula.

We infer a marginal distribution by using a marginal distribution function. Equation (7) shows a Gaussian cumulative distribution function as an example of a marginal distribution function:

$$\hat{F}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp(-\frac{(t-m)^2}{2\sigma^2})\, dt \tag{7}$$

where m denotes the mean of the distribution and σ represents the variance of the distribution. A cumulative distribution function is used because copulas require a random vector whose component is a cumulative score.

Next, we estimate the dependency between the marginal distributions by using copulas. The best copula is determined by comparing the performance of each copula. As mentioned in Section 2, although we must select an appropriate copula from various copulas, this challenge will be addressed in the future. In order to estimate the parameters of a copula, we use maximum likelihood estimation.

4.3 Constructing a Mixture Copula

Individual unimodal copulas are mixed by applying Equation (6). The weight p_i of a mixture copula is considered to be the probability that indicates how much each document contributes to the i^{th} cluster. Therefore, we set weight p_i as the ratio between the number of relevant documents assigned to the i^{th} cluster and the total number of relevant documents.

4.4 A Model for Score Fusion

A method to apply copulas to a score fusion method in an effective manner is unknown. Thus, we propose two models and evaluate their effectiveness.

The first model is a cumulative joint distribution function estimated by a mixture copula, which is in Equation (6). In comparison with a unimodal copula, a mixture copula can be precisely fitted to data when the dependency structure is complex. We evaluate its effect on accuracy improvement when estimating joint distribution more precisely.

The second model is the product of the likelihood of a normalized random vector U and a cumulative score derived from a mixture copula $C_{mix}(U)$, as shown in Equation (8).

$$C_{mix-prod} = C_{mix}(U)\prod_{i=1}^{k} u_i \tag{8}$$

The likelihood of U is calculated based on an assumption that individual components of U occur independently. This assumption is very naive. In order to consider the dependency of the components, we multiply the likelihood by the mixture copula $C_{mix}(U)$ for which the correlation among individual components is considered. Eickhoff et al. multiplied the likelihood by a single copula; however, we use a mixture copula instead of a single copula.

5 Evaluation

We evaluate the effectiveness of our models proposed in Section 4.4 when combining the relevance scores of two IR models.

5.1 Setup for Evaluation

- **Dataset**
  The dataset that we use is the Category B of *ClueWeb'09*[3], excluding Wikipedia documents. *ClueWeb'09* is a dataset used for the Web Track in TREC 2009-2012. Category B, a subset of *ClueWeb'09*, contains approximately 44 million English documents.
- **Queries**
  We used 45 out of 50 queries for ad-hoc tasks of the Web Track in TREC2011. We omitted five queries: four queries that do not have relevant documents in the dataset and one query that includes a numeric term.

[3] http://lemurproject.org/clueweb09/

– **Measures for Evaluation**
The evaluation measures that we use are: Precision (P@k) and normalized Discounted Cumulative Gain (nDCG@k) in the top-k documents; Interpolated Precision (IP@i), where i is recall level; and Mean Average Interpolated Precision (MAIP). Further, we set k = 5, 10, 15, 20 and i = 0.0, 0.1,..., 0.5. P@k is the ratio between the number of relevant documents in the top-k documents and the total number of top-k documents; it is defined as:

$$\text{P@}k = \frac{|A \cap B|}{|A|} \tag{9}$$

where A is a set of top-k documents, and B is a set of relevant documents. Equation (10) shows nDCG@k where iDCG@k is the maximum score of DCG@k, defined as Equation (11). In Equation (11), rel_i denotes a binary variable for the i^{th} document, such that when the i^{th} document is relevant, rel_i equal to 1; otherwise, rel_i is equal to 0. nDCG increases as relevant documents are ranked higher.

$$\text{nDCG@}k = \frac{\text{DCG@}k}{\text{iDCG@}k} \tag{10}$$

$$\text{DCG@}k = \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2(i+1)} \tag{11}$$

IP@k is defined as shown in Equation (12) where R@r is the value of Recall in the top r documents.

$$\text{IP@}i = \max_{r}\{P@r | R@r \geq i\} \tag{12}$$

MAIP is the average of 11 points of Interpolated Precision, as defined in Equation (13).

$$\text{MAIP} = \frac{\sum_{i \in \{0, 0.1, \ldots, 1\}} \text{IP@}i}{11} \tag{13}$$

– **Cross Validation**
Some baselines and our proposed models have parameters that are estimated with training data. We trained the models with a part of the dataset and tested them with the other parts of the dataset. In our experiment, we divided the dataset into 5 parts of training data, then used each parts of training data for a test set. The accuracy of the models is calculated by the average of 5 test results.

– **Target Models for Combination**
We combine two IR models: BM25 [33] and a query likelihood model [27]. Dirichlet smoothing is applied to the query likelihood model. BM25 parameters b, k, and smoothing parameter μ are set to 0.75, 1.2 and 110, respectively. During evaluation, we change the variations of marginal distributions, copulas, and the number of clusters for a mixture copula for as shown in Table 1. In our experiments, a cluster containing only one document is omitted as an outlier. The selection of an appropriate model is a task that will be considered in future work.

Table 1: Model Parameters for Evaluation

	Values Used in the Experiment
Marginal distributions	Gaussian and Empirical Distribution
Copulas	Clayton, Gumbel, Frank, Gaussian and Empirical Copula
The Number of Clusters	2-10

5.2 Baselines

We prepare five baselines for comparison with the performance of the two models shown in Equations (6) and (8). The component x_i of a random vector X denotes a normalized relevance score, and the component u_i of a random vector U denotes a score to which a cumulative distribution function $F_i(.)$ maps the x_i.

Linear combination:

$$LIN(X) = \sum_{i=1}^{k} \lambda_i x_i \tag{14}$$

Harmonic mean:

$$HM(X) = \frac{k \cdot \prod_{j=1}^{k} x_j}{\sum_{i=1}^{k} \frac{\prod_{j=1}^{k} x_j}{x_i}} \tag{15}$$

If random vector X contains at least one component whose score is low, the final score obtained by combination using harmonic mean tends to be low. For example, $HM((0.5, 0.5)) = 0.5$, whereas $HM((0.9, 0.1)) = 0.18$.

An independent copula is given by:

$$C_{indep}(U) = \prod_{i=1}^{k} u_i \tag{16}$$

An independent copula denotes the product of U indicating a cumulative score of joint distribution based on the assumption that individual components U occur independently.

A joint distribution using a single copula is given by:

$$C_{mono}(U) = C(U) \tag{17}$$

The product of a likelihood of U and a score of the cumulative distribution is given by:

$$C_{mono-prod} = C_{mono}(U) \prod_{i=1}^{k} u_i \tag{18}$$

5.3 Statistical Testing

We test the statistical significances by using a Wilcoxon signed-rank test at two significance levels-0.01 and 0.05.

The three major observations are: **(1)** The models using a copula perform significantly better than the models using linear combination, which has shown high performance so far. **(2)** The models that consider dependency perform significantly better than the models that ignore dependency. **(3)** The models that use a mixture copula perform significantly better than the models that use a single copula.

In order to clearly demonstrate these three observations, we compare **(1)** LIN with C_{indep}, C_{mono}, $C_{mono-prod}$, C_{mix}, $C_{mix-prod}$, and $C_{mix-prod}$, and **(2)** C_{indep} with C_{mono}, $C_{mono-prod}$, C_{mix}, and $C_{mix-prod}$, and **(3)** C_{mono}, $C_{mono-prod}$ with C_{mix}, $C_{mix-prod}$.

5.4 Results

In order to determine the best combination of a marginal distribution, a copula, and the number of clusters, we compared the performance of the proposed models for each combination of parameters in Table 1.

We conducted preliminary experiments to determine the best marginal distribution models, copulas, and cluster sizes. Due to the page limitation, we only show the summary of the best combinations in Table 2.

Next, we compare our models with baselines. Table 3 shows the results of the performance. In Table 3, The symbols $*$, †, ‡, and § indicate statistically significant improvements over LIN, C_{indep}, C_{mono}, and $C_{mono-prod}$, respectively. A single symbol indicates statistically significant improvements at the 0.01-level and a double symbol indicates statistically significant improvements at the 0.05-level. A cumulative function of a joint distribution estimated by a mixture copula C_{mix} gains of 10% and 15% over linear combination with respect to P@5 and nDCG@5, respectively. In particular, in terms of nDCG@5, C_{mix} shows a statistically significant improvement over LIN, C_{indep}, C_{mono}, and $C_{mono-prod}$. Among C_{indep}, C_{mono}, and C_{mix}, the performance of C_{mix} is the best, and the performance of C_{mono} exceeds that of C_{indep}. This result indicates that a multidimensional cumulative distribution function can retrieve more relevant documents in the top-5 results when considering the dependency between marginal distributions.

Table 2: Best Combination for Methods with a Copula

Model	Marginal Distribution	Copulas	Number of Clusters
C_{mono}	Empirical Distribution	Empirical Copulas	-
$C_{mono-prod}$	Empirical Distribution	Empirical Copulas	-
C_{mix}	Gaussian Distribution	Clayton Copulas	3
$C_{mix-prod}$	Gaussian Distribution	Clayton Copulas	6

However, in terms of P@$k(\geq 10)$, we do not observe a tendency that the performance of C_{mix} surpasses that of C_{mono}. C_{mix} is effective for the top-5 results, whereas $C_{mix-prod}$ is relatively effective when retrieving 20% of relevant documents. In terms of IP@$i(= 0.1, 0.2)$, $C_{mix-prod}$ outperforms the other models and shows a 5% improvement over LIN, C_{indep}, and $C_{mono-prod}$, C_{mix} is the worst model. From these discussions, we conclude that **(1)** the performance of C_{mix} tends to deteriorate when retrieving 10 or more documents, whereas it

Table 3: Evaluation Results

	LIN	HM	C_{indep}	C_{mono}	$C_{mono-prod}$	C_{mix}	$C_{mix-prod}$
IP@0.0	0.4326	0.4228	0.4392	0.4367	0.4388	0.4603	**0.4609**
IP@0.1	0.2779	0.198	0.2799	0.2756	0.2797	0.2725	**0.294**
IP@0.2	0.2032	0.1196	0.2049	0.198	0.2019	0.1708	**0.2207**
IP@0.3	**0.0846**	0.056	0.0852	0.0874	0.089	0.0471	0.0814
IP@0.4	0.0366	0.0216	0.0382	**0.0404**	0.0398	0.0186	0.0306
IP@0.5	0.0141	0.0068	0.0146	0.0149	**0.0156**	0.0077	0.0111
MAIP	0.0963	0.0751	0.0974	0.0966	0.0976	0.0892	**0.1003**
P@5	0.236	0.236	0.212	0.228	0.228	**0.26**†	0.24†
P@10	**0.232**	0.2	0.206	0.226†	0.226†	0.218	0.224†
P@15	0.2187	0.1893	0.2027	0.2147†	0.2133	0.204	**0.2227**††
P@20	0.219	0.18	0.205	**0.22**†	0.219	0.196	**0.22**
nDCG@5	0.1616	0.1595	0.1529	0.1613	0.1615	**0.1873***†‡‡§§	0.1685
nDCG@10	0.161	0.1472	0.1537	0.1613	0.1604	**0.166**	0.1644†
nDCG@15	0.1574	0.1442	0.1505	0.1576	0.1561	0.1595†	**0.1634**††
nDCG@20	0.1638	0.1426	0.1583	0.1661†	0.1663	0.1612	**0.1711**†

is effective in the top-5 results. **(2)** $C_{mix-prod}$ is effective when retrieving 20% of relevant documents.

6 Conclusion

In this paper, we proposed a score fusion method that uses a mixture copula. Copulas, a family of robust statistical methods, can unify multidimensional relevance scores into a single score, capturing the non-linear dependency among relevance scores. In addition, copulas can provide an interpretable reason for the final result by decomposing a joint distribution into individual marginal distributions and their dependency structure. In the existing score fusion methods that use a copula, it is difficult to capture complex dependencies because these methods employ a unimodal copula, which is expected to be used for a unimodal joint distribution. In contrast, our proposed method can capture complex dependencies by using a mixture copula, which can accurately model a multimodal distribution.

We used more than 44 million documents in *ClueWeb'09*, to compare our method with linear combination and existing score fusion methods that use a copula. For nDCG with the top-5 documents, the proposed method showed a 15% improvement in effectiveness when compared with linear combination.

In future work, the following challenges must be addressed: **(1)** In order to construct a mixture copula automatically, we must determine a method to find the appropriate number of copulas. For example, an information criterion such as AIC can be adopted; and **(2)** We must determine a method to choose the best family of copulas that precisely fits the documents by using certain criteria such as tail dependence correlation and rank correlation coefficient.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers 15H02701 and 15K20990.

References

1. Aslam, J.A., Montague, M.: Bayes optimal metasearch: a probabilistic model for combining the results of multiple retrieval systems. In: Proc. of SIGIR. pp. 379–381 (2000)
2. Bordogna, G., Pasi, G.: A model for a SOft fusion of information accesses on the web. Fuzzy Sets and Systems 148(1), 105–118 (2004)
3. Borlund, P.: The concept of relevance in IR. Journal of the American Society for information Science and Technology 54(10), 913–925 (2003)
4. Bouchaud, J.P., Potters, M.: Theory of financial risk and derivative pricing: from statistical physics to risk management. Cambridge university press (2003)
5. Breymann, W., Dias, A., Embrechts, P.: Dependence structures for multivariate high-frequency data in finance (2003)
6. Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., Hullender, G.: Learning to rank using gradient descent. In: Proc. of ICML. pp. 89–96 (2005)
7. Chen, K., Lu, R., Wong, C., Sun, G., Heck, L., Tseng, B.: Trada: tree based ranking function adaptation. In: Proc. of CIKM. pp. 1143–1152 (2008)
8. Choroś, B., Ibragimov, R., Permiakova, E.: Copula estimation. In: Copula theory and its applications, pp. 77–91 (2010)
9. da Costa Pereira, C., Dragoni, M., Pasi, G.: Multidimensional relevance: A new aggregation criterion. In: Advances in information retrieval, pp. 264–275 (2009)
10. Craswell, N., Robertson, S., Zaragoza, H., Taylor, M.: Relevance weighting for query independent evidence. In: Proc. of SIGIR. pp. 416–423 (2005)
11. Cummins, R.: Measuring the ability of score distributions to model relevance. In: Information Retrieval Technology, pp. 25–36 (2011)
12. Diday, E., Schroeder, A., Ok, Y.: The dynamic clusters method in pattern recognition. In: IFIP Congress. pp. 691–697 (1974)
13. Eickhoff, C., Serdyukov, P., De Vries, A.P.: A combined topical/non-topical approach to identifying web sites for children. In: Proc. of WSDM. pp. 505–514 (2011)
14. Eickhoff, C., de Vries, A.P.: modelling complex relevance spaces with copulas. In: Proc. of CIKM. pp. 1831–1834 (2014)
15. Eickhoff, C., de Vries, A.P., Collins-Thompson, K.: Copulas for information retrieval. In: Proc. of SIGIR. pp. 663–672 (2013)
16. Eickhoff, C., de Vries, A.P., Hofmann, T.: Modelling term dependence with copulas. In: Proc. of SIGIR. pp. 783–786 (2015)
17. Embrechts, P., Lindskog, F., McNeil, A.: Modelling dependence with copulas and applications to risk management. Handbook of heavy tailed distributions in finance 8(1), 329–384 (2003)
18. Fox, E.A., Shaw, J.A.: Combination of multiple searches. NIST SPECIAL PUBLICATION SP pp. 243–243 (1994)
19. Gerani, S., Zhai, C., Crestani, F.: Score transformation in linear combination for multi-criteria relevance ranking. In: Advances in Information Retrieval, pp. 256–267 (2012)

20. Kanoulas, E., Dai, K., Pavlu, V., Aslam, J.A.: Score distribution models: assumptions, intuition, and robustness to score manipulation. In: Proc. of SIGIR. pp. 242–249 (2010)
21. Liu, T.Y.: Learning to rank for information retrieval. Foundations and Trends in Information Retrieval 3(3), 225–331 (2009)
22. Mizzaro, S.: Relevance: The whole history. Journal of the American society for information science 48(9), 810–832 (1997)
23. Montague, M., Aslam, J.A.: Relevance score normalization for metasearch. In: Proc. of CIKM. pp. 427–433 (2001)
24. Montague, M., Aslam, J.A.: Condorcet fusion for improved retrieval. In: Proc. of CIKM. pp. 538–548 (2002)
25. Nelsen, R.B.: An introduction to copulas, vol. 139. Springer Science & Business Media (2013)
26. Onken, A., Grünewälder, S., Munk, M.H., Obermayer, K.: Analyzing short-term noise dependencies of spike-counts in macaque prefrontal cortex using copulas and the flashlight transformation. PLoS computational biology 5(11), e1000577 (2009)
27. Ponte, J.M., Croft, W.B.: A language modeling approach to information retrieval. In: Proc. of SIGIR. pp. 275–281 (1998)
28. Radlinski, F., Joachims, T.: Query chains: learning to rank from implicit feedback. In: Proc. of SIGKDD. pp. 239–248 (2005)
29. Renard, B., Lang, M.: Use of a gaussian copula for multivariate extreme value analysis: some case studies in hydrology. Advances in Water Resources 30(4), 897–912 (2007)
30. Rijsbergen, C.J.V.: Information Retrieval. 2nd edn. (1979)
31. Robertson, S., Zaragoza, H., Taylor, M.: Simple BM25 extension to multiple weighted fields. In: Proc. of CIKM. pp. 42–49 (2004)
32. Robertson, S.E., Jones, K.S.: Relevance weighting of search terms. Journal of the American Society for Information science 27(3), 129–146 (1976)
33. Robertson, S.E., Walker, S., Jones, S., Hancock-Beaulieu, M.M., Gatford, M., et al.: Okapi at TREC-3. NIST SPECIAL PUBLICATION SP pp. 109–109 (1995)
34. Salton, G.: Automatic text processing: The transformation, analysis, and retrieval of. Reading: Addison-Wesley (1989)
35. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Communications of the ACM 18(11), 613–620 (1975)
36. Saracevic, T.: The concept of relevance in information science: A historical review. Introduction to information science pp. 111–151 (1970)
37. Saracevic, T.: Relevance reconsidered. In: Proc. of CoLIS 2. pp. 201–218 (1996)
38. Schamber, L., Eisenberg, M.B., Nilan, M.S.: A re-examination of relevance: toward a dynamic, situational definition. Information processing & management 26(6), 755–776 (1990)
39. Schoelzel, C., Friederichs, P., et al.: Multivariate non-normally distributed random variables in climate research–introduction to the copula approach. Nonlin. Processes Geophys. 15(5), 761–772 (2008)
40. Scott, A.J., Symons, M.J.: Clustering methods based on likelihood ratio criteria. Biometrics pp. 387–397 (1971)
41. Vogt, C.C., Cottrell, G.W.: Fusion via a linear combination of scores. Information Retrieval 1(3), 151–173 (1999)
42. Vrac, M., Billard, L., Diday, E., Chédin, A.: Copula analysis of mixture models. Computational Statistics 27(3), 427–457 (2012)
43. Wu, S., Crestani, F.: Data fusion with estimated weights. In: Proc. of CIKM. pp. 648–651 (2002)