博士論文

Weighting Methods for Information Retrieval Models and Video Retrieval Experiments

（情報検索モデルにおける重み付け法と映像検索データを使用した実証実験）

A dissertation submitted for the degree Doctor of Philosophy

in information science and technology

The University of Tokyo

by

Masaya Murata

村田　眞哉

June 2017

Weighting Methods for Information Retrieval Models and Video Retrieval Experiments

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my supervisor, Prof. Shin’ichi Satoh, during the past three years. I would not have been able to complete this dissertation without his guidance and support.

I gratefully appreciate my thesis committee at The University of Tokyo, Prof. Kiyoharu Aizawa, Prof. Yoichi Sato, Prof. Toshihiko Yamasaki and Prof. Masashi Toyoda for their valuable comments.

I would also like to thank my former bosses at NTT Communication Science Labs, Dr. Hidehisa Nagano, Dr. Kaoru Hiramatsu and Dr. Kunio Kashino, for their kind support. My appreciation also goes out to all the members of Shin’ichi Satoh Lab and of Media Recognition Group at NTT Communication Science Labs.

Last but not least, I want to express my sincere gratitude to my parents and my wife, for their continuous encouragement and support in my life.

Masaya Murata

ABSTRACT

Weighting Methods for Information Retrieval Models and Video Retrieval Experiments

by

Masaya Murata

In this dissertation, new information retrieval (IR) models and the video retrieval experiments are addressed. My first two contributions are about the IR model called the Best Match (BM) 25, which is one of the representative and widely-used IR models, and about its application to the video retrieval task called the instance search. The instance search is a challenging task that has been attracting attentions from video retrieval researchers. For this task, given a specific object shown in image queries, a system is developed to rank videos in which the specific objects are actually shown. The search results are the list of videos ranked in the decreasing order of their relevance degrees to the specific object. I first experimentally demonstrate that the BM25 with my proposed modification is effective in this task and significantly enhances the video retrieval accuracy. Such a modification is performed on the discriminative power called the BM25 inverse document frequency (IDF) and I found that enhancing these powers by my methodology significantly improves the instance search accuracy. The new weight is called the exponential IDF (EIDF).

I next show that the EIDF can be theoretically interpreted in the Bayesian framework and some problems regarding the EIDF are successfully resolved. In this framework, the setting of informative prior knowledge on retrieval features leads to enhance the discriminative power and the new weight resembling the EIDF can be deduced. Compared with the EIDF, since this formulation is theoretically consistent, the new weight called the Bayesian EIDF (BEIDF) does not retain mathematical problems that the EIDF suffers from. The high retrieval accuracy is also confirmed through the instance search experiments.

The third contribution is about the latest IR models called the information-based model (IM) and the divergence from independence (DFI) which both retain model simplicity and retrieval effectiveness. Therefore, the research objective is to develop a simple and effective IR model. The term weight for the IM is designed as the extent that the normalized, within-document term frequency diverges from the standard value. The standard value is calculated by the so-called information model and I show that the model based on the generalized Pareto distribution (GPD), which is the main asymptotic distribution in the extreme value statistics (EVS), results in extending the DFI. Together with the novel parameter estimation method for the GPD, the proposed model becomes data-driven, that is, the model parameters can be estimated and specified by data to be searched, and the retrieval effectiveness is also verified using the instance search dataset.

As for the theoretical results, since the GPD includes the log-logistic distribution (LLD) as a special case, some existing knowledge on IR models relying on the LLD assumption can be also interpreted from the GPD viewpoint. Since the LLD has been often assumed as the underlying distribution for constructing IR models, its extension, that is, GPD, is also expected to become another basic principle. Exploring this research direction is promising and intriguing.

To summarize, the new IR models are addressed and three novel weighting methods are proposed in this dissertation. The first two methods are for achieving the state-of-the-art retrieval accuracy and the third one is also for the model simplicity. Their effectiveness was experimentally verified using the instance search dataset and I expect that the findings of this dissertation contribute for the further exploration and development of a new family of IR models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND SYMBOLS

IR	Information retrieval
TF	Term frequency
IDF	Inverse document frequency
VSM	Vector space model
PRP	Probabilistic ranking principle
BM25	Best match 25
LM	Language model-based IR
DFR	Divergence from randomness
AA	Axiomatic approach
IM	Information-based model
PM	Percentile-based model
EVS	Extreme value statistics
DFI	Divergence from independence
LMJM	LM with Jelineck-Mercer smoothing
LMDS	LM with Dirichlet Smoothing
LLD	Log-logistic distribution
CDF	Cumulative distribution function
EIDF	Exponential IDF
BIDF	Bayesian IDF
BEIDF	Bayesian EIDF
GPD	Generalized Pareto distribution

CBVR	Content-based video retrieval
RSJ weight	Robertson-Sparck-Jones weight
TREC	Text REtrieval Conference
SIFT	Scale-invariant feature transform
CSIFT	Color SIFT
TRECVID	TREC video retrieval evaluation
BOVW	Bag-of-visual words
EBM25	BM25 with EIDF
ROI	Region-of-interest
RW	ROI weighting
RR	ROI re-ranking
NTF	Normalized TF
MBD	Maximum block data
TED	Threshold excess data
EVI	Extreme value index
MEF	Mean excess function

Rel Relevanceness taking binary values of rel in the case of relevance and $\overline{\mathrm{rel}}$ in the case of irrelevance

v Video vector whose elements are within-video keypoint frequencies indexed by BOVW

v_i Within-video frequency of the ith keypoint

q	Image query vector whose elements are within-image-query keypoint frequencies indexed by BOVW
n_B	Dimension of BOVW
E_i	Eliteness to keypoint i taking binary events of $\bar{e}_i$ in the case of non-eliteness and e_i in the case of eliteness
r_i	Number of judged relevant videos containing keypoint i
n_i	Number of videos containing keypoint i
R	Number of videos judged relevant
N	Number of videos in the database
w_i^{IDF}	BM25 IDF
ξ'	Parameter of EIDF
w_i^{EIDF}	EIDF
m_i	Probability of the ith keypoint becoming elite
γ	Parameter of BEIDF
λ	Parameter of ROI weight
q_{ROI}	Vector of query keypoints inside the ROI
τ	Parameter of ROI re-ranking
q'	Keyword-query vector whose elements are within-query TFs
d	Document vector whose elements are within-document TFs
q_i'	Within-query frequency of the ith term
d_i	Within-document frequency of the ith term
l_i	Total frequency of the ith term in the document collection
λ'	Parameter of LMJM

μ'	Parameter of LMDS
e_i'	Expected TF of the ith term in the document
u	Threshold for TED
v	Parameter of MEF
$\mathrm{e}_u(v)$	MEF with a threshold u
$\hat{\mathrm{e}}_\mu(v)$	Estimated MEF with a threshold μ
(ϕ, σ, μ)	Parameters of GPD
$(\hat{\phi}, \hat{\sigma}, \mu)$	Estimated values of GPD parameters (μ is a tuning parameter)

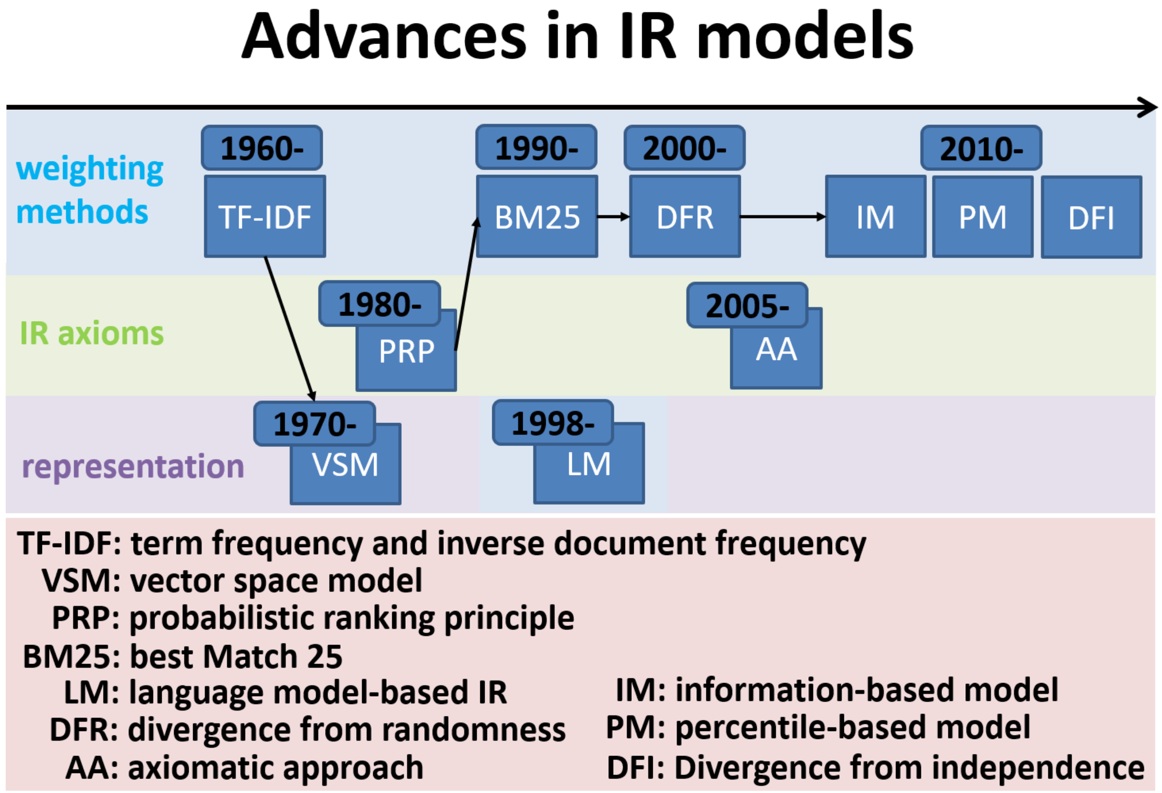
Chapter 1. Introduction

1. Background and Motivation

The information retrieval (IR) is a research field in which an exploration of methods on representing, searching, and ranking large collections of electronic text and other human-language data is conducted [4]. As well as the other engineering fields, the IR has a long research history. It was roughly started around 1960, and the researchers from all over the world have been intensively involved in this field [4]. Some of the research achievements of IR are already realized as actual services such as Web search engine, article search, library search, patent search, etc. It is now inconvenient to spend a day without using these services and consequently, from such an application viewpoint, it is not too much to state that the IR has already met the practical level.

Thus, the natural question arising from this background is that what are important and essential points remained unexplored in the IR research field? I think one of the answers for this question is the underling theory for the search and ranking processes. Of course, such theoretical results have been also intensively pursued by many researchers, however, compared with the application achievements mentioned above, there is still a room left for this issue. Such a theoretical research is called the development of IR models [4] and it composes the main topic of this dissertation. I next summarize the development history of the representative IR models.

2. Development of IR Models


Advances in IR models

TF-IDF: term frequency and inverse document frequency
VSM: vector space model
PRP: probabilistic ranking principle
BM25: best Match 25
LM: language model-based IR
DFR: divergence from randomness
AA: axiomatic approach
IM: information-based model
PM: percentile-based model
DFI: Divergence from independence

Figure 1: Development history of representative IR models.

To start with, I briefly describe the existing IR models. Figure 1 illustrates the development history of the representative IR Models. The TF-IDF developed in 1960s [4] provides the weighting method for terms that compose a document. The weight is calculated by the within-document term frequency (TF) multiplied by the inverse document frequency (IDF). The IDF is calculated based on the number of documents containing the term in the document database (corpus) and it is multiplied with the TF to suppress the weights for often-appearing terms in documents.

The TF-IDF is utilized for the VSM [4] and the VSM was designed for the similarity calculation between two documents. In the VSM, the document is expressed by a term vector whose elements are TF-IDF weights. Then, the similarity score between two documents are calculated by the closeness between the two vectors such as using the cosine similarity measure. The idea of expressing a document by the term vector greatly influences the development of the subsequent IR models and this approach is often called the Bag-of-words (BoW) expression.

The PRP [4][5] opened up a new line of research on IR models and provided an axiom for formulating the search and ranking processes in IR. The PRP states that the ideal ranked list (search result) is obtained by ranking documents in the decreasing order of their relevance degrees to the keyword query. This statement is formulated as probabilistic, not deterministic such as TF-IDF and VSM, and the relevance probability is expressed by the conditional relevance probability of a document given a keyword query. After the proposal of PRP, the BM25 [6] was designed to realize this axiom with additional assumptions in 1990s. The effectiveness of the BM25 has been verified by many researchers for a series of keyword-query document retrieval experiments. The BM25 is now one of the most-often used IR models and indeed, it is often chosen as a baseline model to compare with the subsequent IR models.

The LM [7] is also probabilistic but was developed with a different theoretical basis from the PRP. It expresses query and document by term-occurrence probability distributions and estimates the closeness between these distributions using the Kullback-Leibler divergence [4]. Therefore, the LM can be regarded by the probabilistic extension of the VSM. The essential technique is the smoothing method applied to the document probability distribution. It is applied to avoid the mathematical problem, that is, the problem of division by zero, in the LM equation and the different smoothing methods result in the different LM equations. The representative LMs are LM with Jelineck-Mercer smoothing (LMJM) and LM with Dirichlet Smoothing (LMDS) [4]. As well as the BM25, the effectiveness of these models has been also confirmed by a series of document retrieval experiments.

The DFR [8] focuses on the TF and estimates the weight by measuring its divergence degree from the one predicted by the standard model. Such a standard model is called the randomness model and the divergence degree is expressed by a probability. By assuming the adequate randomness model, the probability of the actual TF observed in a document is calculated. Then, the terms that having small probabilities, that is, those having high information, are rewarded high weights. The actual calculation of the DFR relies on using two different randomness models to avoid the extremely high information (called the divergence problem). The DFR with specific randomness models was proved to yield the BM25-like model [8].

The approach of the AA [9] is similar to that of the PRP. The AA also provides axioms for constructing IR models but the essential difference from the PRP is that these axioms are based on experimental or heuristic findings. Therefore, the axioms used in the AA are not as obvious as that for the PRP. However, since these axioms have been experimentally verified by a series of document retrieval experiments, the models following these axioms are expected to provide the satisfactory retrieval accuracy. This approach also made it possible to analytically diagnose the effectiveness of the existing IR models by investigating whether these models satisfy the AA axioms or not. The AA approach made a huge impact on developing the subsequent IR models.

The IM was proposed to simplify the DFR [10] [11]. Using the normalized TF, not the raw TF, the IM was designed to avoid the divergence problem in the DFR. With this change, only one randomness model called the information model is required. This is a clear advantage over the DFR since adequate settings of the two randomness models have been troublesome. The LMJM was derived by choosing the log-logistic distribution (LLD) for the information model [10] and this fact somehow supported the reason the LM has been successful for keyword-query document retrieval experiments.

The PM was proposed to define the TF weight by using the cumulative distribution function (CDF) [12]. The CDF is called the percentile-based model and by multiplying the derived TF weight with the IDF, it is shown that the BM25-like model can be derived from the PM [12]. The selection problem of the adequate percentile-based model was recently addressed using the knowledge of the EVS [13].

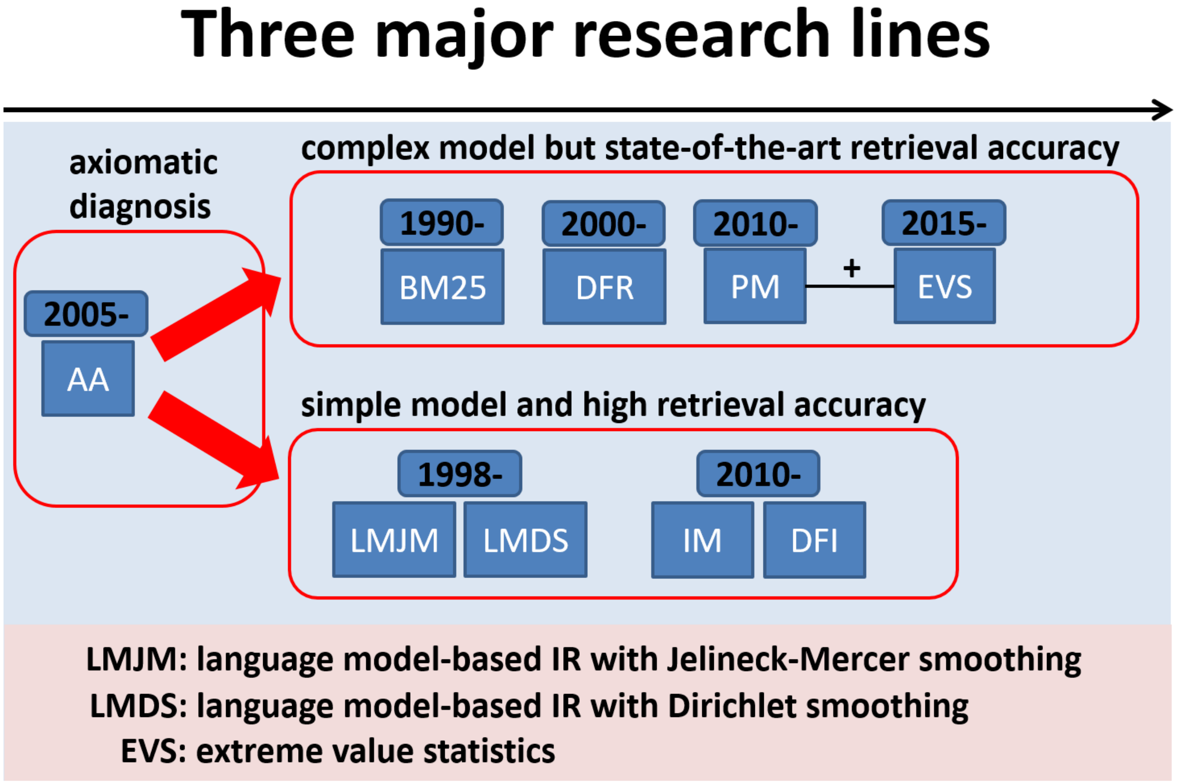
The DFI [14] is similar to the DFR and IM. However, the model estimates the term weights by observing the extent the actual TF diverges from the expected TF within the document. The expected TF is calculated under the assumption that the term occurs independently within the document. Therefore, the setting of two randomness models or the information model is no longer necessary in the DFI. Unlike the DFR and IM, since the distribution assumption is not necessary, the model can be regarded as distribution-free and parameter-free.

3. Connections Between IR Models

Figure 2 depicts the connections between the representative IR models. There are three primary lines of research on IR models: BM25, LMJM/LMDS, and AA. Roughly speaking, the BM25 line of research often achieves the state-of-the-art retrieval accuracy, however, the

LMJM/LMDS line provides simpler but effective IR models with less number of parameters that should be tuned. The DFR and PM include BM25-like models as special cases, on the other hands, the IM and DFI include the LMJM as special cases. The model that includes the LMDS as a special case is not discovered yet and the further research on this issue is highly desired. The AA makes it possible to investigate the effectiveness of these models by checking whether they unconditionally or conditionally satisfy the empirical axioms. Every time a new model is proposed, we can perform the AA to analytically check the potential effectiveness.

The proposed models in this dissertation are regarding the first two lines, BM25 and LMJM/LMDS lines, and the contributions are summarized in the next section.



Three major research lines

axiomatic
diagnosis
2005-
AA
complex model but state-of-the-art retrieval accuracy
1990-
BM25
2000-
DFR
2010-
PM
+
2015-
EVS
simple model and high retrieval accuracy
1998-
LMJM
LMDS
2010-
IM
DFI

LMJM: language model-based IR with Jelineck-Mercer smoothing
LMDS: language model-based IR with Dirichlet smoothing
EVS: extreme value statistics

Figure 2: Connections between representative IR models.

4. Contributions and Organization

In this dissertation, I extended the existing two representative IR models, BM25 and DFI. Such extensions were designed to make the models well suit to the data to be searched and using the video retrieval experiments called the instance search task, the improvement in the retrieval accuracy was confirmed. The instance search task has been attracting attentions from researchers and for this task, given a specific object shown in image queries, a retrieval system is pursued in which the videos are ranked according to their relevance degrees to the given specific object.

To be more specific, the main contributions of this dissertation are summarized as follows:

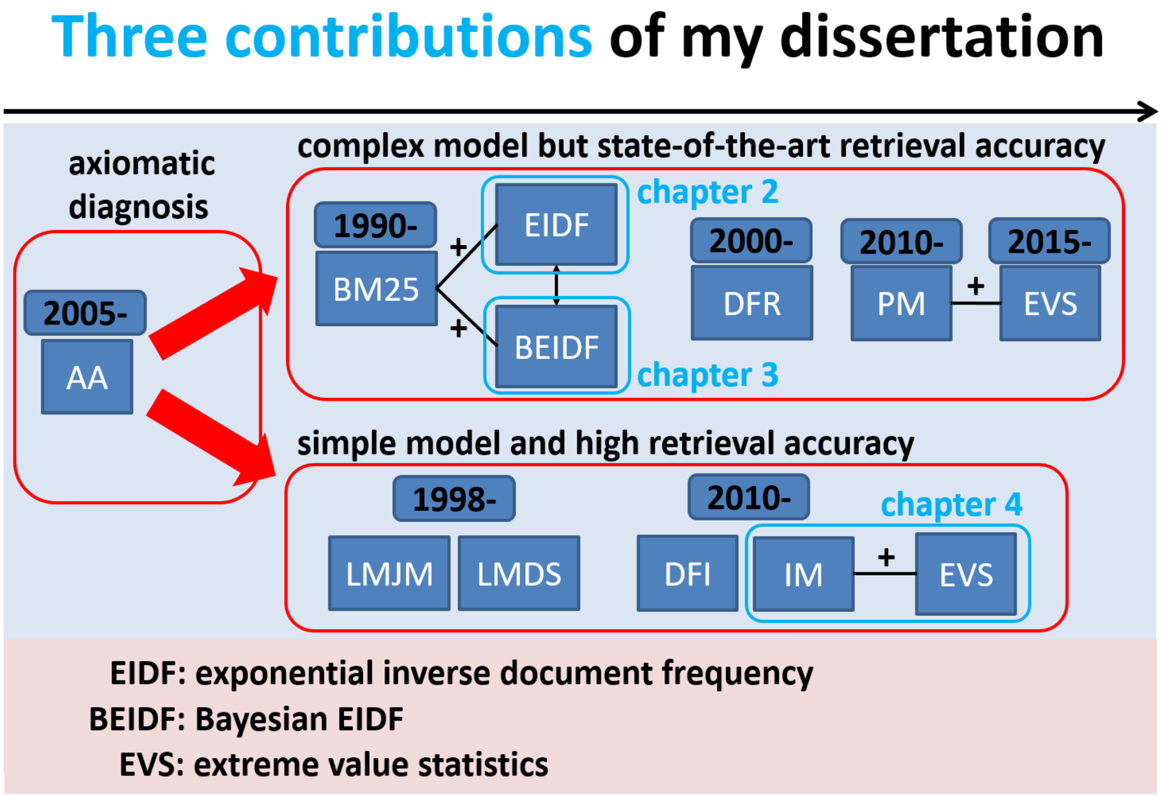
(1) Chapter 2: I first extend the BM25 to enhance the discriminative power of the BM25 IDF. The new power is called the exponential IDF (EIDF) and the effectiveness was confirmed for the instance search experiment. The EIDF contributes for suppressing the weights for non-essential key-points in image queries whose weights could not be suppressed sufficiently by using the standard BM25 IDF [1].

(2) Chapter 3: I next show that the EIDF-like weight can be formally derived in the Bayesian framework. Within this framework, I show that the use of informative priors modeled by Beta distributions for retrieval features leads to the EIDF-like weight called the Bayesian EIDF (BEIDF). The use of non-informative priors is proved to derive the BM25 IDF-like weight. Therefore, this approach can be regarded as the formal extension of the BM25 and it does not suffer from the mathematical problems that the EIDF involves. Adequately setting the parameters of the Beta priors is found to significantly improve the instance search accuracy [2].

(3) Chapter 4: I finally propose the IM realized by using the knowledge of the EVS. I especially focus on the use of the Generalized Pareto distribution (GPD) to set the information model and provide the parameter estimation method. Since the GPD is an extension of the LLD, the existing IR models based on the LLD assumption (LMJM and DFI) can be regarded as the special cases of this model. Since the GPD parameters are estimated according to the data to be searched by the proposed methodology, the retrieval accuracy is naturally improved, which is also confirmed using the instance search dataset [3].

Figure 3 illustrates these three contributions. IR models that are superior in terms of the search accuracy follow the BM25 line. With this in mind, among the proposed models, the highest retrieval accuracy was confirmed by the first two contributions (BM25+EIDF and BM25+BEIDF). As shown by the double-headed arrow in Fig. 3, there is a theoretical relationship between these two models, that is, between EIDF and BEIDF. On the other hand, the advantage of the LMJM/LMDS line is the model simplicity. Since the number of tuning parameters is smaller compared with those for the models in BM25 line, it can be easily applied to different kinds of data. The third contribution (IM+EVS) is on this line and provides the data-driven IR model with the confirmed retrieval accuracy.

Finally, Chapter 5 concludes this dissertation by summarizing the basic findings. I expect that these findings contribute for the further exploration of a new family of IR models. Some ideas on the future research directions are also provided in this chapter.



Three contributions of my dissertation

EIDF: exponential inverse document frequency
BEIDF: Bayesian EIDF
EVS: extreme value statistics

Figure 3: Illustration of three contributions.

Chapter 2. BM25 With Exponential IDF for Instance Search

1. Overview

This chapter deals with a novel concept of an EIDF in the BM25 formulation and compares the search accuracy with that of the BM25 IDF in a content-based video retrieval (CBVR) task. The video retrieval method is based on a bag-of-keypoints (bag-of-local-visual-features) and the EIDF estimates the keypoint importance weights more accurately than the BM25 IDF. The EIDF is capable of suppressing the keypoints from frequently occurring background objects in videos, and I found that this effect is essential for achieving improved search accuracy in CBVR. This proposed method is especially designed to tackle instance video search, one of the CBVR tasks, and I demonstrate its effectiveness in significantly enhancing the instance search accuracy using the TRECVID2012 video retrieval dataset.

2. Background

The probabilistic information retrieval model was originally proposed by Robertson and Jones for keyword query document retrieval in 1977 [5], [15]. The documents are expressed by sets of term frequencies and each document has a binary relevance property (relevant or irrelevant) for the keyword query. The relationships among document, query, and relevance are expressed by conditional probabilities and investigating the conditional relevance probability given a document and a query was the primary research objective. Under a so-called binary independence model assumption that either the keyword is supposed to be present or absent in documents (i.e., the documents are expressed by sets of binary keyword frequencies), the conditional relevance probability was expressed by a sum of the keyword weightings called Robertson-Sparck-Jones (RSJ) weights [15]. Furthermore, by adding a reasonable assumption that usually there is no relevance information available prior to searching, the RSJ weights were simply approximated by IDF-like weights (IDF weights were defined by Jones in different research in 1972 [16]). Under the binary assumption, we found that the conditional relevance probability relates to the sum of the within document keyword IDF-like weights.

Robertson and Walker proceeded with research without the use of the binary assumption and introduced a new binary variable called the keyword eliteness [6]. The meaning of the eliteness can be interpreted as aboutness and if the keyword is elite in a document, in some sense the document is about a concept denoted by the keyword and vice versa. They posited that the within document keyword frequency depended on this keyword eliteness and further assumed that the frequency follows a Poisson distribution with larger expectation for the elite keyword than that for the non-elite keyword. By applying this assumption, which is called a two-Poisson assumption, a state-of-the-art probabilistic information retrieval model called the BM25 document-ranking function was formulated in 1990s [17], [18]. The two-Poisson assumption gave us the knowledge that as a within document keyword frequency becomes sufficiently large, its contribution to the conditional relevance probability approaches the IDF-like weight called the BM25 IDF. Even though there is a slight gap in the subsequent reasoning, by following this asymptotic behavior, the BM25 was again defined using the within document keyword BM25 IDF. This is a brief overview of the BM25 development.

Since the BM25 is a mathematical model for information retrieval tasks in general and its effectiveness has been demonstrated in a series of widely known TREC (Text REtrieval Conference) experiments, many researchers, especially in the text retrieval field, now use

BM25 as the baseline ranking function [19]. However, the great success of BM25 in text search was implicitly supported by the use of the BM25 IDF. Recall that the BM25 is based on the two-Poisson assumption and it requires the larger expectation when the keyword is elite than when the keyword is non-elite. It is obvious that not every keyword has such a clear property and these keywords are non-conceptual ones like articles or conjunctions. Since the two-Poisson assumption does not hold for these kinds of keywords, the BM25 seems to readily produce a wrong estimate of the conditional relevance probability. However, the use of BM25 IDF in the BM25 formulation considerably alleviates this drawback because for such non-conceptual keywords, the BM25 IDF values become sufficiently small. These keywords' inadequate contributions to the conditional relevance probability are thus sufficiently suppressed by their small BM25 IDF values and the resulting BM25 score does not deviate very much from the theoretical one. This favorable BM25 IDF effect supports the robust feature of BM25 for document search.

I have interests in applying BM25 to the other media searches, and this paper deals with image query content-based video retrieval (CBVR). In the BM25 formulation for CBVR, keywords are replaced with keypoints (local visual features) described by the scale-invariant feature transform (SIFT) [20] and color SIFT [21], and videos are expressed by sets of the keypoint frequencies. Unfortunately, we found that the aforementioned effects of BM25 IDF could not be sufficiently obtained for video retrieval tasks. At this time, keypoints that violate the two-Poisson assumption mainly originate from background regions such as grass, trees, and sky. They often appear in videos as background objects even though they are not elite (that is, even though they are not the topics of the videos) and thus work against the two-Poisson assumption that forms the essential part of BM25. Following this viewpoint, I propose an EIDF concept to enhance the BM25 IDF’s effect and eliminate such unwanted keypoints. Using the new IDF weights instead of the original BM25 IDF significantly improves the BM25 search accuracy for CBVR. I demonstrate my video retrieval performance, especially for the instance video search that has been actively discussed in the TRECVID (TREC Video Retrieval Evaluation) community since 2010. The instance video retrieval system searches for videos in which the instance (specific person/object/place) really appears, given the query images with the region-of-interest images showing the instance to be searched. I summarize the main contributions of this paper as follows.

1) The problem of applying the conventional BM25 to CBVR is clarified.

2) An EIDF concept is proposed to significantly enhance the BM25 search accuracy in CBVR.

3) My approach is verified using TRECVID2012 instance search datasets.

The rest of this chapter is organized as follows. I introduce related work in Section 3. I next explain the formulation of the BM25 for CBVR and describe the EIDF concept in Section 4. In Section 5, I present the instance search experiments, including the datasets used, the results obtained, and the evaluations determined. Finally, in Section 6, I summarize the main features of the proposed BM25 with EIDF and the basic findings of this chapter.

3. Related Work

Generally speaking, video retrieval approaches are classified into the following two categories: approaches that use textual metadata and those that don’t. I call the former metadata-based retrieval and latter content-based retrieval. I first introduce previous work on metadata-based retrieval and then introduce work in the latter category. I finally present some recent developments of the BM25 model in the information retrieval field.

A. Metadata-Based Video Retrieval

For metadata-based video retrieval, BM25 models are mostly used as the baseline text retrieval methods and the search results obtained by using BM25 are then further re-ranked by using link analysis methods like PageRank [22] to enhance the retrieval precision. Yan *et al.* [23] investigated the retrieval performance of BM25 on various text sources obtained by performing automatic speech transcripts and video optical character recognition for the videos and also on production metadata such as titles and descriptions of the videos to be searched. Their suggestion was to combine two or more different text sources because these sources contain complementary information and the combinational use often leads to improvement of the video retrieval accuracy. Liu *et al.* [24] used a modified BM25 as the text-based search method to obtain the initial ranked list and then created virtual hyperlinks based on the similarity of the visual information between the videos. Then the initial ranked list was re-ranked based on the relevance scores propagated by applying a modified PageRank algorithm to the constructed graph structure. Hoi *et al.* [25] represented a video retrieval task by using graphs whose links are based on the similarities of the visual and textual metadata information between the videos. Then the search results obtained by using BM25 were further re-ranked on the basis of the probabilities of being hit by the starting query node in a random walk viewpoint. Liu *et al.* [26] used BM25 to obtain the initial ranked list and mined the graph structure using the bag-of-visual words (BOVW) representation within the list. The BOVW was constructed on the basis of the visual information in the videos and then the PageRank algorithm was applied to calculate the salient pattern, which indicates the importance of each visual word, and the concurrent pattern, which expresses the interdependent relations among the visual words. The initial search results were re-ranked using these two criteria, aiming to improve the search accuracy.

Jeon *et al.* [27], [28] proposed a statistical generative model for learning the semantics of images and an automatic approach to annotating and retrieving images based on a training set of images. They showed that the probabilistic model allowed to predict the probability of generating a word given blobs in an image. Their method may be used to automatically retrieve images given a word as a query in the framework of metadata-based video retrieval.

B. Content-Based Video Retrieval

The content-based retrieval approaches mostly represent images or videos as vectors whose elements are frequencies of certain image features. Analogous to the vector space model in the text retrieval field, the aim is to compress high-dimensional image information into a lower dimensional vector space. The similarity between the two vectors can be seen as a distance measure in the space. and the scoring function most commonly used is based on cosine correlation, which itself is based on angles between vectors. Squire *et al.* [29] reported the application of text retrieval techniques to content-based image retrieval. Their search system employed more than 80,000 simple color and spatial frequency features, both local and global, extracted at several scales, and demonstrated its effectiveness using 10 queries on a test database of 500 images. Vries *et al.* [30], [31] discussed the relationship between the text information retrieval and multimedia retrieval, and introduced the retrieval with Bayesian networks from the conventional text retrieval approaches. Vries *et al.* [32] later applied the language modeling retrieval approach to the problem of image searching.

Sivic. *et al.* [33] represented objects and scenes in videos by a set of viewpoint invariant 128-dimensional SIFT descriptors [20]. The vector quantization was carried out by K-means clustering and with the constructed BOVW, object and scene retrievals were performed by using the similarity basis in the vector space. The BOVW framework proved to be effective in video retrieval task and was later widely followed by many researchers [34], [35], [36], [37], [38].

Zhu *et al.* [38] represented videos by a set of 192-dimensional color SIFT [21] and adopted hierarchical K-means (HKM) clustering [34] to accelerate building a large-scale BOVW and performing online searches. Recently, they also adopted approximate K-means clustering (AKM) [35] and found that AKM is superior to HKM in terms of search accuracy in a series of instance video retrieval experiments [39]. Zhao *et al.* [40] used various image features such as color-based, texture-based, and shape-based features, and linearly combined each query-video distance score to enhance the final instance video retrieval performance. Peng *et al.* [41] also used various image features, such as Color Moment Grid, Local Binary Pattern, SIFT, color SIFT, and opponent SIFT [42], to search for instances in a video collection. To improve the instance search task, Zhang *et al.* [43] proposed a retrieval method that exploits the spatial information of the local image features extracted from the instance regions, and ranked the videos using the keypoint proximity information.

As for the concept-based video retrieval, Aly *et al.* [44] proposed the general ranking framework to define effective and robust ranking functions. Their framework used the probability of relevance given concept occurrences as a ranking function, which was derived from the probability of relevance ranking function originally proposed in the text retrieval field. While they dealt with the high-level semantic concepts, they reported that the effect on the retrieval model might well be the same as that using low-level visual word features. Their proposed method improved the search accuracy in the shot retrieval and the segment retrieval tasks over several baselines in TRECVID test collections.

C. Recent Developments of BM25

It is widely known that the BM25 model was extended to the BM25F model for the retrieval of structured documents by Robertson *et al.* [45] in 2004. The problem in applying BM25F was the setting of the various tuning parameters, and Svore *et al.* [46] proposed a machine learning approach to effectively optimize them.

Fang *et al.* [9] defined a set of basic desirable constraints that any reasonable retrieval model should satisfy and investigated what extent the BM25 retains such favorable properties. They mentioned that for most of the cases when queries are input as a few number of keywords, the BM25 satisfied the constraints, but also pointed out that the BM25 IDF term in BM25 might violate some of them. To the authors, such violation occurs when the BM25 IDF term becomes negative, which often happens when the query keywords are verbose, and in that case, they suggested to replace the BM25 IDF term with the IDF term in the pivoted normalization retrieval formula [47] in the vector space model. Their modified BM25 approach experimentally outperformed the standard BM25 for the verbose queries. Lv *et al.* [48] revealed that BM25 overly penalized very long documents and presented a BM25 extension to boost the ranking scores of such documents. Recently, Blanco *et al.* [49] proposed an extension of BM25 and BM25F by determining virtual regions of documents using multiple query operators.

My new concept of EIDF seeks to eliminate unwanted keypoints that violate the two-Poisson assumption behind the BM25 formulation. The use of EIDF is aimed at restoring and enhancing the overall BM25 search performance for CBVR, and I demonstrate its effectiveness, especially for image query instance video search.

4. BM25 For Image Query CBVR

In this section, I first explain the preprocessings of image queries and videos in a database required for the calculation of the BM25. I next describe the BM25 for image query CBVR by analogy to the BM25 for keyword query document retrieval in the text retrieval field. I then propose the EIDF concept to enhance the BM25 search performance for image query CBVR tasks.

A. Preprocessings

Figure 4 shows an overview of my preliminary processings. I suppose that the queries input to CBVR systems are composed of a few images and the systems rank the large numbers of videos stored in the database according to their relevance degrees to the image query. The details of each preliminary processing are as follows.

1) *Keyframe and Keypoint Extraction:* I first extract frame images from the videos at a certain rate, such as one frame per second. I next extract the keypoints showing the prominent local visual features from the query images and from the frame images using

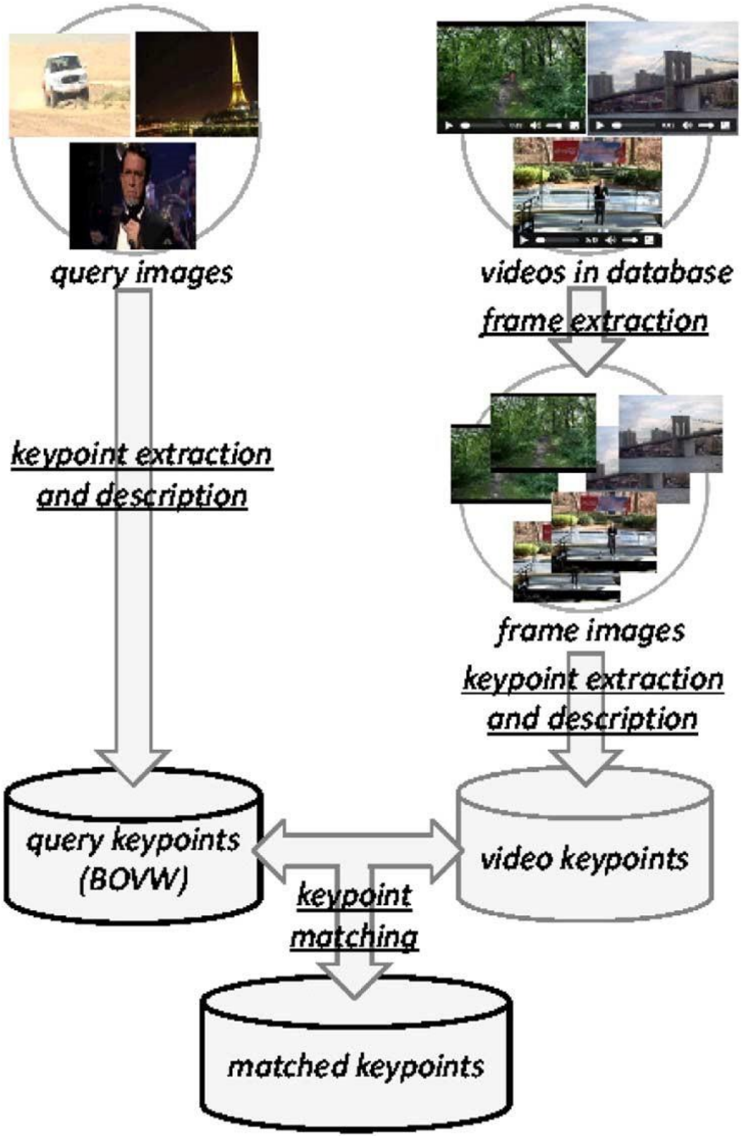


Figure 4: Overview of preliminary processings of image queries and videos in a database. (Copyright©2014 IEEE, [R1] Fig. 1)

detectors, such as Harris-Laplace [21] or Maximally Stable Extremal Regions [50]. The description methods for the keypoints are also optional; however, recent research often adopts two major description methods such as SIFT [20] (128-dimensional) and color SIFT [21] (192-dimensional) vectors. I thus use these two methods to generate two feature vectors for each keypoint extracted from image queries and from videos in a database.

2) *Matching Keypoints:* The keypoints extracted from image queries compose the BOVW for the subsequent video retrieval. Here, keypoint clustering is not necessarily required but methods that could be used are K-means, HKM [34], or AKM [35]. Then, the keypoints extracted from the frame images are matched with each visual word (each keypoint extracted from image queries) on the basis of the cosine similarity value between the two feature vectors. The keypoint pair showing the highest cosine value larger than a certain threshold among the other visual words is considered as matched. I usually set the threshold to 0.95 or higher, and these settings are to tolerate subtle differences in local visual features of the same object, which sometimes occur in taking images or videos. The matching procedure is performed for every pair between the visual words and the keypoints extracted from the frame images. After keypoints have been alternately matched on the basis of the SIFT and the color SIFT feature vectors, videos in a database are expressed by sets of two local visual feature frequencies (SIFT and color SIFT), similar to how documents are expressed by sets of keyword frequencies in the text search BM25.

B. Probabilistic Information Retrieval

The retrieval system ranks the videos by the relevance probabilities given image queries and the videos to be searched. The conditional relevance probability is expressed by the following equation:

$$P(\mathrm{Rel} = \mathrm{rel}|v, q) \propto_q \log\left(\frac{P(\mathrm{Rel} = \mathrm{rel}|v, q)}{P\left(\mathrm{Rel} = \overline{\mathrm{rel}}|v, q\right)}\right) \tag{1}$$

Here, Rel is a relevanceness taking binary values of rel in the case of relevance and $\overline{\mathrm{rel}}$ in the case of irrelevance. The right side of Eq. (1) is the log-odds of relevance. I will use the short-hand notations $P(\mathrm{rel}|v, q)$ to denote $P(\mathrm{Rel} = \mathrm{rel}|v, q)$ and $P\left(\overline{\mathrm{rel}}|v, q\right)$ to denote $P\left(\mathrm{Rel} = \overline{\mathrm{rel}}|v, q\right)$. $\propto_q$ indicates the equivalence of rank order, and the left and the right sides of Eq. (1) thus yield the same rank order of videos in a database.

v and q are the vectors of a video and an image query, whose elements are the within-video and within-query-image frequencies of the keypoints indexed by the constructed BOVW, respectively. The two random vectors are denoted as follows:

$$v \coloneqq \left(KF_1, KF_2, \cdots, KF_{n_B}\right) \tag{2}$$
$$q \coloneqq \left(qKF_1, qKF_2, \cdots, qKF_{n_B}\right) \tag{3}$$

Here, n_B is the dimension of the BOVW, and it specifies the dimensions of the two random vectors. KF_1, KF_2, and KF_{n_B} denote the keypoint frequency counts, respectively, and for the element of q, I usually adopt the binary presence or absence feature within the image query, having only the values zero (absence) or one (presence).

Then, following the mathematical transformations for the development of the BM25 for document search shown in [18], Eq. (1) is expressed by the following equations:

$$
\begin{align}
P(\mathrm{rel}|v,q) &\propto_q \log\left(\frac{P(\mathrm{rel}|v,q)}{P\left(\overline{\mathrm{rel}}|v,q\right)}\right) \nonumber\\
&= \log\left(\frac{P(v|\mathrm{rel},q)}{P\left(v|\overline{\mathrm{rel}},q\right)}\frac{P(\mathrm{rel}|q)}{P\left(\overline{\mathrm{rel}}|q\right)}\right) \tag{4}\\
&\propto_q \log\left(\frac{P(v|\mathrm{rel},q)}{P\left(v|\overline{\mathrm{rel}},q\right)}\right) \tag{5}\\
&\approx \log\prod_{i\in V}\frac{P(KF_i = kf_i|\mathrm{rel},q)}{P\left(KF_i = kf_i|\overline{\mathrm{rel}},q\right)} \tag{6}\\
&\approx \log\prod_{i\in q}\frac{P(KF_i = kf_i|\mathrm{rel},q)}{P\left(KF_i = kf_i|\overline{\mathrm{rel}},q\right)} \tag{7}\\
&\approx \log\prod_{i\in V}\frac{P(KF_i = kf_i|\mathrm{rel},q_i)}{P\left(KF_i = kf_i|\overline{\mathrm{rel}},q_i\right)} \tag{8}\\
&= \log\prod_{i\in V}\frac{P(KF_i = kf_i|\mathrm{rel})}{P\left(KF_i = kf_i|\overline{\mathrm{rel}}\right)} \tag{9}\\
&= \sum_{q}\log\left(\frac{P(KF_i = kf_i|\mathrm{rel})}{P\left(KF_i = kf_i|\overline{\mathrm{rel}}\right)}\right) \tag{10}\\
&= \sum_{q,kf_i>0}\log\left(\frac{P(kf_i|\mathrm{rel})}{P\left(kf_i|\overline{\mathrm{rel}}\right)}\right) \nonumber\\
&\qquad + \sum_{q,kf_i=0}\log\left(\frac{P(0|\mathrm{rel})}{P\left(0|\overline{\mathrm{rel}}\right)}\right) \nonumber\\
&\qquad - \sum_{q,kf_i>0}\log\left(\frac{P(0|\mathrm{rel})}{P\left(0|\overline{\mathrm{rel}}\right)}\right) \nonumber\\
&\qquad + \sum_{q,kf_i>0}\log\left(\frac{P(0|\mathrm{rel})}{P\left(0|\overline{\mathrm{rel}}\right)}\right) \tag{11}\\
&\propto_q \sum_{q,kf_i>0}\log\left(\frac{P(kf_i|\mathrm{rel})}{P\left(kf_i|\overline{\mathrm{rel}}\right)}\right) \nonumber\\
&\qquad - \sum_{q,kf_i>0}\log\left(\frac{P(0|\mathrm{rel})}{P\left(0|\overline{\mathrm{rel}}\right)}\right) \tag{12}\\
&\qquad = \sum_{q,kf_i>0} w_i \tag{13}
\end{align}
$$

where

$$w_i = \log\left(\frac{P(KF_i = kf_i|\text{rel})P\left(KF_i = 0|\overline{\text{rel}}\right)}{P\left(KF_i = kf_i|\overline{\text{rel}}\right)P(KF_i = 0|\text{rel})}\right) \tag{14}$$

Here, the first transformation for Eq. (4) is based on Bayes' rule and the second one is obtained by discarding the second component in Eq. (4), which is independent of the video. Then, each probability in Eq. (5) is expressed by a product over the visual words of the BOVW under the assumption of independence among the KF_i. Equation (7) is derived from the assumption that for any non-query keypoints (keypoints that are absent in the query vector), the probabilities are independent of relevance properties. Thus, the product of Eq. (6) is restricted to the keypoints of query images (keypoints that are present in the query vector). Equation (8) follows the assumption that KF_i only depends on q_i and, since q_i is always 1 under the setting of the binary presence or absence feature, q_i is omitted in Eq. (9). Note here that I implicitly assumed that q_i is a binary attribute and thus, Eq. (9) does not insist that KF_i is independent of q_i. Equation (9) is transformed to Eq. (10) and it is further divided into four terms as shown in Eq. (11). Here, the third and fourth terms are cancelled each other and the Eq. (10) is separated into the summation over the query keypoints that are present in the video (first term) and that over the keypoints that are absent in the video (second term). Then, since the sum of the second and the forth terms is independent of the video, in other words, since the sum of these terms only depend on the query, Eq. (12) that discards these terms does not change the video ranking order. And finally, as expressed in Eq. (13), the sum of w_i over the query keypoints present in the video preserves the same rank order as that of $P(\text{rel}|v, q)$. This is the mathematical deduction of the probabilistic information retrieval models for CBVR.

5. BM25 Formulation

w_i in Eq. (14) is expressed by the following equations using the eliteness assumption. With the use of a two-Poisson distribution assumption, the asymptotic behavior of w_i is further investigated as follows:

$$
\begin{aligned}
w_i &= \log\left(\frac{P(kf_i|e_i)P(e_i|\mathrm{rel}) + P(kf_i|\bar{e}_i)P(\bar{e}_i|\mathrm{rel})}{P(kf_i|e_i)P\left(e_i|\overline{\mathrm{rel}}\right) + P(kf_i|\bar{e}_i)P\left(\bar{e}_i|\overline{\mathrm{rel}}\right)} \right. \\
&\qquad \left. \times \frac{P(0|e_i)P(e_i|\mathrm{rel}) + P(0|\bar{e}_i)P\left(\bar{e}_i|\overline{\mathrm{rel}}\right)}{P(0|e_i)P(e_i|\mathrm{rel}) + P(0|\bar{e}_i)P(\bar{e}_i|\mathrm{rel})}\right) && (15) \\
&= \log\left(\frac{P(e_i|\mathrm{rel}) + \dfrac{P(kf_i|\bar{e}_i)}{P(kf_i|e_i)}P(\bar{e}_i|\mathrm{rel})}{P\left(e_i|\overline{\mathrm{rel}}\right) + \dfrac{P(kf_i|\bar{e}_i)}{P(kf_i|e_i)}P\left(\bar{e}_i|\overline{\mathrm{rel}}\right)} \right. \\
&\qquad \left. \times \frac{\dfrac{P(0|e_i)}{P(0|\bar{e}_i)}P\left(e_i|\overline{\mathrm{rel}}\right) + P\left(\bar{e}_i|\overline{\mathrm{rel}}\right)}{\dfrac{P(0|e_i)}{P(0|\bar{e}_i)}P(e_i|\mathrm{rel}) + P(\bar{e}_i|\mathrm{rel})}\right) && (16) \\
&\to \log\left(\frac{P(e_i|\mathrm{rel})\left(1 - P\left(e_i|\overline{\mathrm{rel}}\right)\right)}{P\left(e_i|\overline{\mathrm{rel}}\right)\left(1 - P(e_i|\mathrm{rel})\right)}\right), \text{as } kf_i \to \infty && (17)
\end{aligned}
$$

The assumption of the eliteness model is that for any video-keypoint pair, there is a hidden property which I refer to as eliteness (aboutness) and, instead of the relevance, this eliteness affects the actual occurrences of the keypoint in the video. The eliteness to keypoint i is a binary event E_i, having either $\bar{e}_i$ (not elite) or e_i (elite). Under the additional assumptions that the two probability distributions of KF_i conditioned on the eliteness and the non-eliteness follow two Poisson distributions with larger expectation for the elite keypoint than for the non-elite keypoint, w_i in Eq. (14) approaches its upper bound value expressed in Eq. (17) as kf_i sufficiently increases since $\frac{P(kf_i|\bar{e}_i)}{P(kf_i|e_i)} \to 0$ and $\frac{P(0|e_i)}{P(0|\bar{e}_i)} \approx 0$.

This asymptotic behavior of w_i is simply approximated as below:

$$w_i \approx \frac{kf_i}{kf_i + h}\log\left(\frac{P(e_i|\text{rel})\left(1 - P\left(e_i|\overline{\text{rel}}\right)\right)}{P\left(e_i|\overline{\text{rel}}\right)\left(1 - P(e_i|\text{rel})\right)}\right) \tag{18}$$

$$\approx \frac{kf_i}{kf_i + h}\log\left(\frac{\frac{r_i + b}{R + a}}{\left(\frac{n_i - (r_i + b)}{N - (R + a)}\right)}\frac{\left(1 - \frac{n_i - (r_i + b)}{N - (R + a)}\right)}{\left(1 - \frac{r_i + b}{R + a}\right)}\right) \tag{19}$$

Here, h is a positive integer parameter and I usually set $h = 2$. To estimate the logarithm value, I use reasonable approximations that $P(e_i|\text{rel}) \approx \frac{r_i+b}{R+a}$ and $P\left(e_i|\overline{\text{rel}}\right) \approx \frac{n_i-(r_i+b)}{N-(R+a)}$ where r_i, n_i, R, and N are the number of judged relevant videos containing keypoint i (visual word i in BOVW), the number of videos containing keypoint i, the number of videos judged relevant, and the number of videos in the database, respectively. Then, by adding an assumption that there is no relevance information available prior to searching ($R = r_i = 0$) and by setting $a = -1, b = -0.5$, the BM25 expressed by using the BM25 IDF [6] is deduced as follows:

$$w_i \approx \frac{kf_i}{kf_i + h}\log\left(\frac{N - n_i + 0.5}{n_i + 0.5}\right)$$
$$\approx \frac{kf_i}{kf_i + h}w_i^{\text{IDF}} \tag{20}$$

This is the derivation of the conventional BM25 model for image query CBVR.

6. Exponential IDF (EIDF)

As I mentioned in Section 2, the problem is that the BM25 is totally based on the two-Poisson assumption with larger expectation when the keypoint is elite than when it is non-elite. As in the case for keywords, not every keypoint follows such a property. From my consideration, keypoints from frequently occurring background objects like trees and a sky violate the two-Poisson assumption since these objects often appear when taking videos outside (but do not necessarily often appear in the videos taken inside). As for the videos taken inside, such keypoints that might violate the two-Poisson assumption could be from interior walls or house lighting. It means that the keypoints concerning them tend to easily occur even though they are not elite (though they are not the topics) in the videos. What makes matters worse, contrary to non-conceptual keywords in text search, I found that these keypoints are not sufficiently eliminated when the original BM25 IDF weights are used. The reason is described as follows.

Documents are written by following well-defined forms and thus non-conceptual keywords tend to repeatedly occur because they retain functions to relate or connect conceptual keywords. This is why the corresponding BM25 IDF weights become sufficiently lowered. However, videos are not well-formed; they are just produced by taking scenes or objects at outside or inside home, so that the keypoints of the problem tend not to repeatedly occur. Therefore, the original BM25 IDF values calculated by Eq. (20) do not get sufficiently suppressed and the keypoints that violate the the-Poisson assumption thus bring unwanted contributions to the BM25 estimate for the conditional video relevance probability. Consequently, the use of the conventional BM25 for CBVR needs careful treatment and its simple application often results in unsatisfactory video search accuracy. The problem is due to the mild settings of $a = -1$ and $b = -0.5$ that have successfully worked for the cases of text (document) search tasks.

From this viewpoint, enhancing the BM25 IDF's effect to eliminate such keypoints is greatly desired. I achieve that by setting $a = e^{n_i/\xi'}$ and $b = e^{-n_i/\xi'}$ in Eq. (19). This drastically lowers $P(e_i|\mathrm{rel})$ when the keypoints tend to occur in different kinds of videos, corresponding to suppression of the contributions of the aforementioned troublesome keypoints to the final BM25 ranking score. In other words, the new BM25 IDF weight (EIDF) is designed to suppress the keypoints extracted from frequently occurring background objects in videos in a database. ξ' is a parameter and I empirically set $\xi' = 10 \sim 20$. Then the BM25 incorporating this EIDF is expressed as follows:

$$
\begin{aligned}
w_i &\approx \frac{kf_i}{kf_i + h} \times \\
&\log\left(\frac{\left(\dfrac{r_i + e^{-n_i/\xi'}}{R + e^{n_i/\xi'}}\right)}{\left(\dfrac{n_i - \left(r_i + e^{-n_i/\xi'}\right)}{N - \left(R + e^{n_i/\xi'}\right)}\right)} \frac{\left(1 - \left(\dfrac{n_i - \left(r_i + e^{-n_i/\xi'}\right)}{N - \left(R + e^{n_i/\xi'}\right)}\right)\right)}{\left(1 - \left(\dfrac{r_i + e^{-n_i/\xi'}}{R + e^{n_i/\xi'}}\right)\right)} \right) \\
&\approx \frac{kf_i}{kf_i + h} \log\left(\frac{e^{-n_i/\xi'}}{\left(n_i - e^{-n_i/\xi'}\right)} \frac{\left(N - e^{n_i/\xi'} - n_i + e^{-n_i/\xi'}\right)}{\left(e^{n_i/\xi'} - e^{-n_i/\xi'}\right)} \right) \\
&\approx \frac{kf_i}{kf_i + h} w_i^{\mathrm{EIDF}}
\end{aligned}
\tag{21}
$$

Here, Eq. (21) is obtained by setting $R = r_i = 0$. With the analogical use of the standard normalization of the keyword frequency by the document length in text retrieval BM25, the keypoint frequency is normalized as follows:

$$
\begin{aligned}
w_i' &\approx \frac{kf_i'}{h + kf_i'} w_i^{\mathrm{EIDF}}, \\
&where\ \ kf_i' = \frac{kf_i}{(1 - b') + b' \cdot vl/avvl}
\end{aligned}
\tag{22}
$$

Here, vl denotes the video length (total keypoint frequencies within the video) $vl \coloneqq \sum_{i \in V} kf_i$ and $avvl$ is the average video length over the videos in the database. b' is a normalization design parameter ranging from 0 to 1 and 0 means there is no video length normalization and

1 means the full normalization. Since the keypoint frequency highly depends on the video length, such normalization procedure is crucial and indeed is one of the most essential ranking components as well as the use of the IDF weight in the information retrieval.

Finally, the BM25 video-scoring function is obtained by summing these keypoint-weights over the set of query keypoints that are present in the video as follows:

$$P(\mathrm{rel}|v,q) \propto_q \sum_{q,kf_i>0} w_i' \tag{23}$$

The videos in the database are ranked according to Eq. (23), and I denote the BM25 with EIDF model simply as EBM25 in the following experiment section.

7. Experiments

In this section, I present the experimental evaluation of the instance search performance of the BM25 with EIDF weights (EBM25). I first describe the dataset I used and the preprocessings I performed prior to the experiments. I next explain the baseline retrieval methods along with the various settings of the proposed methods. Finally, I present the evaluation results, showing actual examples of instance video search results to clarify the effectiveness of the EBM25.

A. Dataset and Preliminary Processings

I utilized the TRECVID2012 instance video search dataset. As already mentioned, given query images with the region-of-interest images showing the instance (specific person/object/place) to be searched, the instance retrieval system searches for the videos in which the instance really appears. For this dataset, there are 21 instance topics and each topic is composed of 5 query images and region-of-interest images on average. I show some examples of query, region-of-interest, and masked query images of the 21 instance topics below.

In Fig. 5, the white regions in the region-of-interest images indicate the instance topics shown in the query images. Some instance topics have entirely white region-of-interest images, indicating that the query images themselves are the instances to be searched. I generated the masked query images by superimposing the region-of-interest images on the query images.

The number of videos in the dataset is 76751, and I extracted frame images from each video by the rate of one frame per second and obtained about a total of 740,000 frame images. From them, I extracted the keypoints using the Harris-Laplace detector [21] and featured the keypoints on the basis of the 128-dimensional SIFT [20] and the 192-dimensional color SIFT [21], in which the first 128 dimensions in the descriptor correspond to the normal luminance SIFT and the latter 64 dimensions contain chrominance information. The keypoints extracted from the query images compose the BOVW and as explained in Section 4-A, the keypoint matching procedure was alternately performed using the SIFT and the color SIFT feature vectors by setting each cosine similarity threshold to 0.95. Then the videos in the database are expressed by sets of two local visual feature (SIFT and color SIFT) frequencies and the videos are ranked according to the EBM25.

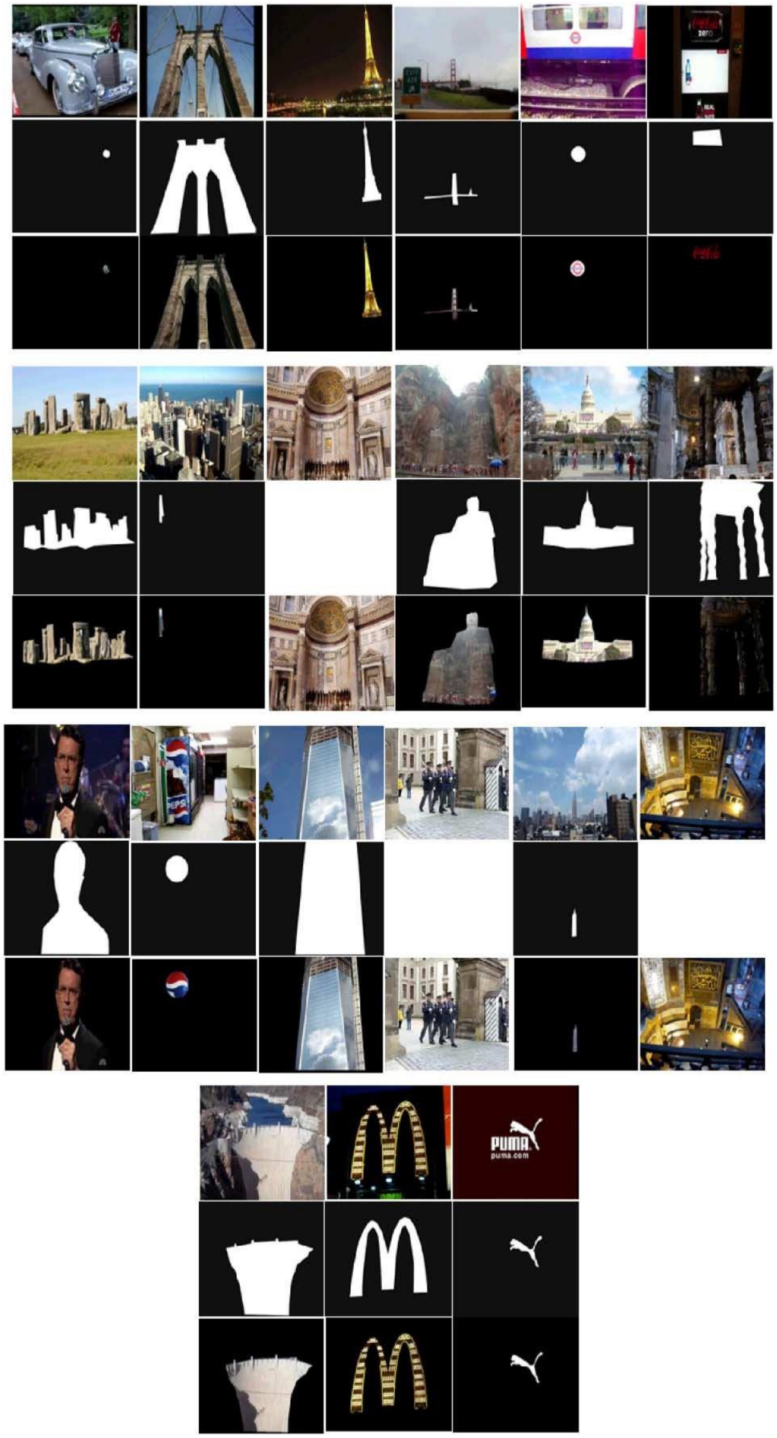


Figure 5: Example images of 21 instance topics. Upper, middle, and lower images are query, region-of-interest, and masked query images, respectively. (Copyright©2014 IEEE, [R1] Fig. 2)

The relevance judgment data was binary, so a video was judged either correct or wrong for each instance topic. Some videos were judged neither correct nor wrong for an instance topic and in measuring the retrieval accuracy of the search result ranking, I simply discarded the non-judged videos and raised the lower ranked videos one rank higher.

B. Evaluation Results

I compared the video retrieval accuracy of EBM25 with those of three baseline retrieval models. I also evaluated the proposed method with a different ξ' setting using the masked query images to investigate the effectiveness of masked query images in improving the instance video search accuracy. I summarize a total of five retrieval methods that were evaluated as follows:

1) IDF: A method that ranks the videos in decreasing order of the total within-video keypoint IDF weights of the query images,

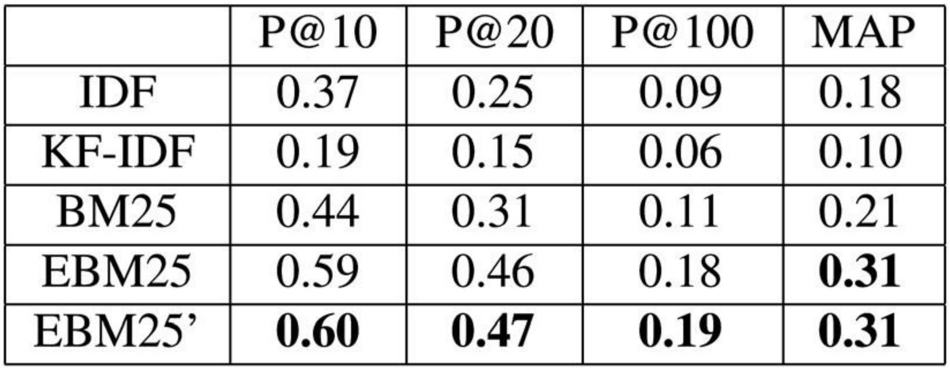
2) KF-IDF: A method that ranks the videos in decreasing order of the total within-video keypoint frequencies of the query images multiplied by keypoint IDF weights,

3) BM25: BM25 video-ranking model with the settings $h = 2.0$ and $b' = 0.75$,

4) EBM25: EBM25 with the settings $h = 2.0$, $b' = 0.75$, $\xi' = 10$,

5) EBM25': EBM25 with the settings $h = 2.0, b' = 0.75$, and either $\xi' = 20$ (emphasizing keypoints when they appear in masked query images) or $\xi' = 10$ (not emphasizing keypoints when they do not appear in masked query images).

Table 1: Evaluation results of three baselines (IDF, KF-IDF, BM25) and two proposed (EBM25, EBM25') methods on instance search. Bold numbers denote the highest values. (Copyright©2014 IEEE, [R1] Table 1)

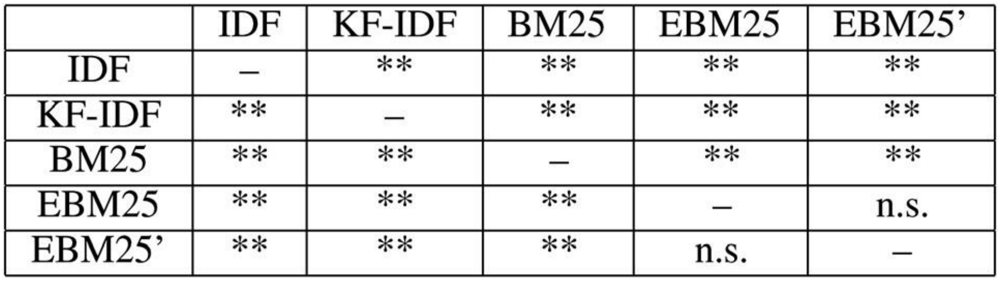


	P@10	P@20	P@100	MAP
IDF	0.37	0.25	0.09	0.18
KF-IDF	0.19	0.15	0.06	0.10
BM25	0.44	0.31	0.11	0.21
EBM25	0.59	0.46	0.18	**0.31**
EBM25'	**0.60**	**0.47**	**0.19**	**0.31**

I measured the retrieval accuracy of the top 1000 ranked video lists using Precision@10, 20, 100, and MAP (Mean Average Precision) measures. Table I shows the evaluation results. From Table 1, I can see that KF-IDF is much worse than the other models, suggesting that the within video keypoint frequency must follow the saturation behavior as shown in the BM25 equation (that is, the form of $kf_i/(kf_i + h)$). The BM25 improves the IDF and KF-IDF methods; however, the result indicates that the original BM25 IDF is not good enough to eliminate unwanted keypoints in searching instance videos since the improvement is not so large. Its retrieval accuracy is further enhanced by using the EIDF weights as shown at the EBM25 row in the table. I also found that using masked query images slightly raises the search accuracy.

I further investigated the statistically significant differences among the MAP measures of IDF, KF-IDF, BM25, EBM25, and EBM25'. Table 2 shows the significance test results using the Wilcoxon signed-rank test. From Table 2, I see that the probabilistic models are significantly better than the IDF and KF-IDF methods and that among the methods, EBM25 or EBM25’ is the most promising retrieval approach for the instance video search task. Since I can see that EBM25 significantly outperforms BM25, it is clear that the use of EIDF weights is essential for tackling CBVR tasks. The use of region-of-interest images and of the generated masked query slightly improves the instance video retrieval accuracy but the improvement is not significant, indicating the effect of masked query images is limited. I also note that the MAP values of EBM25 and EBM25’ are higher than the highest MAP scored among all of the teams that participated in the TRECVID2012 instance search task last year.

Table 2: Statistically significant differences in MAP values among IDF, KF-IDF, BM25, and EBM25’. (, *, and n.s. denote $p < 0.01$, $0.01 < p < 0.05$, and no significant difference, respectively.) (Copyright©2014 IEEE, [R1] Table 2)**



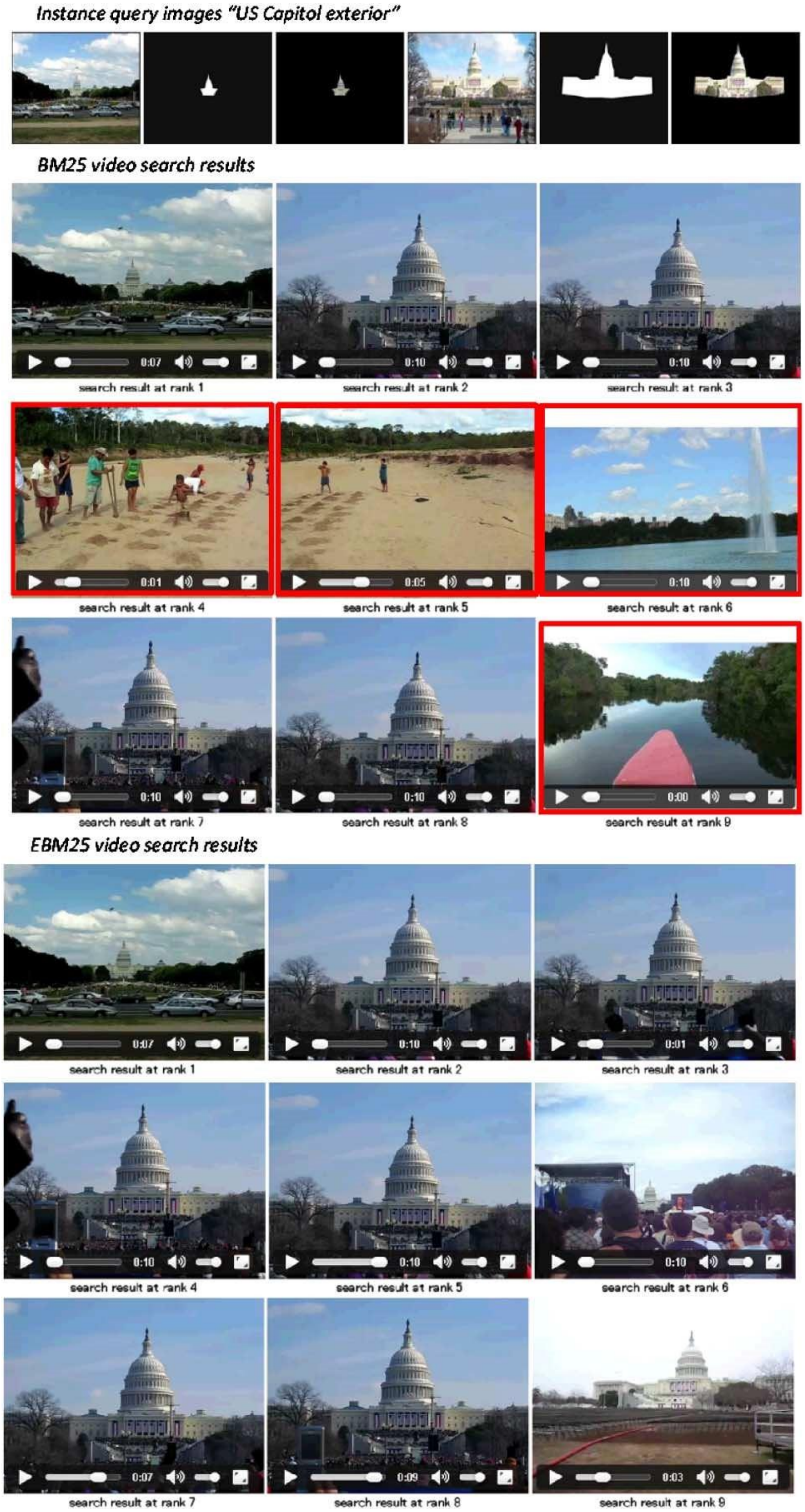
	IDF	KF-IDF	BM25	EBM25	EBM25’
IDF	–	**	**	**	**
KF-IDF	**	–	**	**	**
BM25	**	**	–	**	**
EBM25	**	**	**	–	n.s.
EBM25’	**	**	**	n.s.	–

The appropriate ξ' value depends on the threshold of the keypoints cosine similarity matching. For example, if I set the threshold to 0.9 instead of 0.95, I empirically confirmed that the appropriate ξ' was between $100 \sim 200$. When a soft keypoint mapping is employed instead of my hard mapping, the optimal value could be also changed. Exploration of the optimal EIDF parameter would be one of my future works. The overall results demonstrate the vital role of EIDF weights and the capability of the proposed video retrieval approach.

C. Actual Search Result Examples

I present actual examples of video search results for BM25 and EBM25. Figure 6 shows the search results for the instance query images of "US Capitol exterior". The retrieval systems search for videos in which the "US Capitol exterior" really appears and rank the videos according to the relevance degrees. The number of query images provided for this instance search was 2, and in Fig. 6, I also list the region-of-interest and the generated masked query images. The middle set of videos are the top 9 search results obtained by using the conventional BM25 model. The videos surrounded by the red lines were irrelevant to the query. The top 9 search results for the EBM25 model are listed below and all of them were relevant to the instance query images. The MAP value was improved from 0.10 to 0.38.

Since the original BM25 IDF weights of BM25 are not good enough to suppress the frequently occurring background keypoints of instance query images, the irrelevant videos showing trees, sand, and sky are ranked higher at the search results. Note that only using the masked query images as the query images for the BM25 calculation could suppress these troublesome effects, however, since it was already known that the instance search with complete ignorance of the instance background regions would greatly hurt the search accuracy mainly due to the rather small instance region, I did not take that approach. My approach both utilized the original instance query images and the masked query images. Figure. 6 clearly shows that the original BM25 IDF could not sufficiently suppress the effects of the unnecessary keypoints of trees, sand, and sky and this failure dramatically lowered the MAP values since the measure puts high values on the higher search result rankings.



Instance query images "US Capitol exterior"

BM25 video search results

EBM25 video search results

Figure 6: Video search result examples for BM25 and EBM25 methods on instance query "US Capitol exterior." Upper, middle, and lower images show the set of query images, search result examples of BM25, and those of EBM25, respectively. Videos enclosed by red lines are irrelevant to the instance query. (Copyright©2014 IEEE, [R1] Fig. 3)

To the contrary, EBM25 successfully retrieved relevant videos at the top search result ranks, demonstrating the effectiveness of the EIDF weights in eliminating such troublesome keypoints. Since EBM25 automatically reduces the weights of the frequently occurring background keypoints, in other words, since it automatically decides whether the keypoints are necessary or not for the instance searching from the statistical point of view, the search accuracy is boosted compared with that of the standard BM25 approach. Such keypoint selection is similar to the manual stop-words removal often performed in the text retrieval and indeed, the result of the EIDF might be employed for the automatic construction of the "stop-keypoints" list for the successful image/video retrieval.

8. Conclusion

In this paper, I proposed the EIDF in the BM25 formulation for CBVR and compared the search accuracy with that of BM25 with the original BM25 IDF. My video retrieval method is bag-of-keypoints (local visual features) based and the EIDF is designed to suppress the frequently occurring background keypoints that violate the two-Poisson assumption forming the essential part of the BM25 approach. Using the TRECVID2012 dataset, I verified my video retrieval approach, especially for the instance video search task, and demonstrated that the use of the new weight significantly improves the search accuracy compared with the BM25 with BM25 IDF weights.

The EIDF idea can be also applied to the other well-known information retrieval approaches such as DFR [8] and language model based document ranking methods [7]. It is thus my future work to pursue the wide application of the proposed EIDF idea, not only to the other image/video search tasks, but also to the textual document retrieval researches. I also intend to apply the proposed approach to other media search tasks such as music retrieval to investigate the effectiveness of the EIDF weightings for the bag-of-audio words representation.

Chapter 3. Bayesian EIDF and ROI Effect for Enhancing Instance Search Accuracy

1. Overview

In this chapter, I first analyze the discriminative power in the BM 25 formula and provide its calculation method from the Bayesian point of view. The resulting, derived discriminative power is quite similar to the EIDF that I have previously proposed [1] but retains more preferable theoretical advantages. In the previous paper [1], I proposed the EIDF in the framework of the probabilistic IR method BM25 to address the instance search task, which is a specific object search for videos using an image query. Although the effectiveness of the EIDF was experimentally demonstrated, I did not consider its theoretical justification and interpretation. I also did not describe the use of region-of-interest (ROI) information, which is supposed to be input to the instance search system together with the original image query showing the instance. Therefore, here, I justify the EIDF by calculating the discriminative power in the BM25 from the Bayesian viewpoint. I also investigate the effect of the ROI information for improving the instance search accuracy and propose two search methods incorporating the ROI effect into the BM25 video ranking function. I validated the proposed methods through a series of experiments using the TREC Video Retrieval Evaluation instance search task dataset.

2. Background

In this section, I first describe the instance search task and the BM25 retrieval model using the standard BM25 inverse document frequency (simply abbreviated as BM25 IDF to differentiate it from the original IDF in [16]). I next explain the EIDF I previously proposed [1] to enhance the discriminative power of the BM25 IDF and to improve the effectiveness of the probabilistic IR method for addressing the instance search task. I then clarify the research questions tackled in this chapter and summarize the main contributions. The organization of this chapter is also provided in this section.

A. Instance Search Task

A video retrieval task called instance search has been rigorously discussed in the TREC Video Retrieval Evaluation (TRECVID) community since 2010. In this task, a system is required to search for and rank videos showing a specific object/person/place given in the image queries. Such a specific object/person/place is called an instance or instance topic. The image queries are composed of original images showing the instance to be searched and region-of-interest (ROI) images, which specify the instance region within the original images.

Figure 7 shows some query examples. Each white region specified in the ROI image shows the instance to be searched within the original image. In the instance search task, given these original and ROI images, I need to design a search system that automatically retrieves and ranks videos stored in the database according to their degrees of relevance to the instance topic.

To address this issue, I previously proposed the video ranking method based on the probabilistic IR method, which is briefly described in the next section.

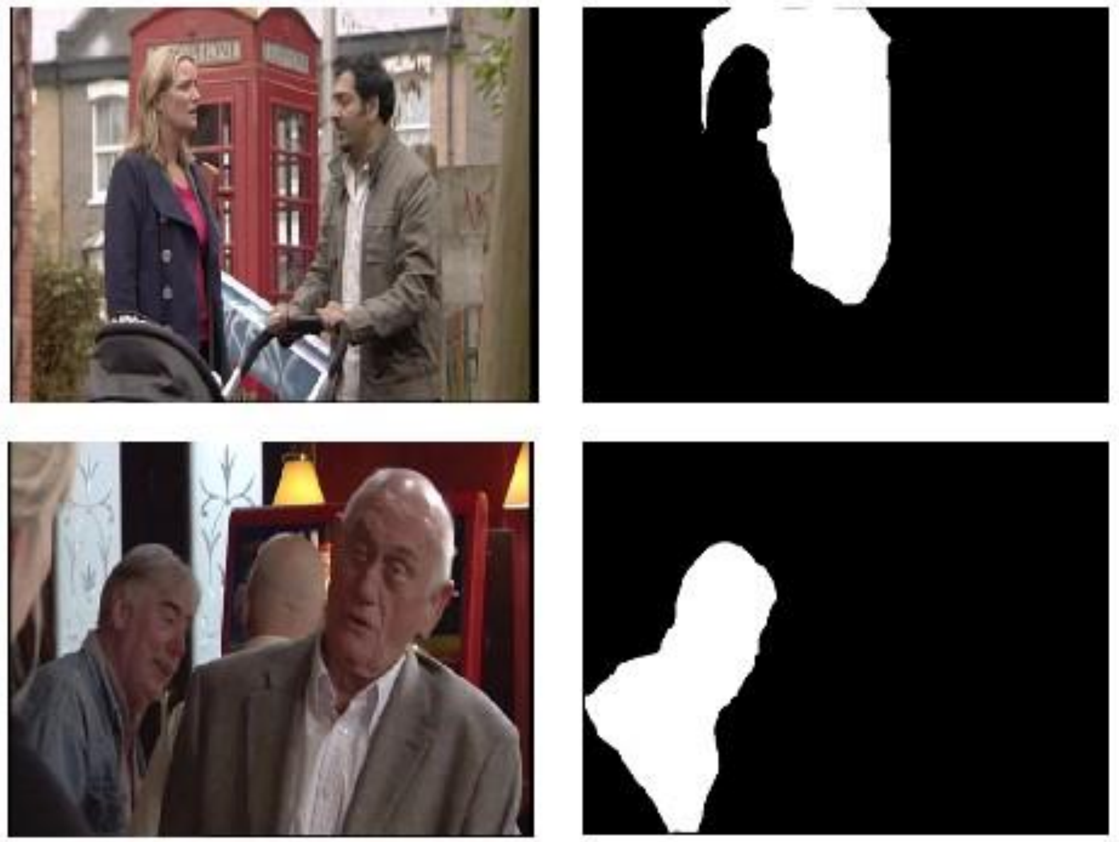


Figure 7: Original and ROI images for “this public phone booth” and “this man”. (Copyright©2016 IEICE, [R2] Fig. 1)

B. BM25 using IDF

The BM25 method was originally proposed for addressing the document retrieval task, and it is now regarded as one of the state-of-the-art probabilistic IR methods [4]. The BM25 method is designed to order documents ranked by their relevance probabilities to the input keyword query. Mathematically, BM25 ranks the documents according to the following conditional probabilities:

$$P(\text{rel}|q', d) \tag{24}$$

Here, rel is an event indicating relevance to the document. q' and d are vectors whose elements are within-query and within-document keyword frequencies. These elements denoted as q_i' (i = 1,2, · · ·, M) and d_i (i = 1,2, · · · , M), where, M is the total number of unique words, are discrete random variables. Imposing the reasonable assumptions for text/document search task, Eq. (24) becomes the well-known BM25 document ranking function as shown below.

$$P(\text{rel}|q', d) \propto_{q'} \sum_{q_i'} \frac{d_i}{d_i + \kappa} \log\left(\frac{N' - n_i' + 0.5}{n_i' + 0.5}\right) \tag{25}$$

Here, $\propto_{q'}$ indicates that the document ranking results from the left side of the equation and the right side of the equation are equal (that is, $\propto_{q'}$ denotes the ranking equivalence sign). $\sum_{q_i'}$ means the summation over keywords within q_i' that are also present in d (summation over the common keywords), κ is a parameter for which $\kappa = 2$ is often used for a text/document retrieval task, and N' and n_i' are the total number of documents in the database to be searched and the number of documents that contain the ith query keyword (called document frequency).

On the right side of Eq. (25), the log(·) is called the BM25 IDF, which corresponds to the importance weight of the *i*th query keyword and is often interpreted as the discriminative power for searching relevant documents to the keyword query issued. The discriminative powers become smaller for query keywords that often appear in many kinds of documents since such keywords are not supposed to enable the discrimination between correct and incorrect documents to the query. To the contrary, query keywords only appearing in specific documents are statistically evaluated as having high discriminative powers. This is the key insight observed on the right side of Eq. (25), along with the fact that the effect of within-document keyword frequency (first term) only approaches 1; therefore, the discriminative power (second term) is theoretically more emphasized in the BM25 document ranking function.

C. BM25 using EIDF

In the previous paper [1], I first applied Eq. (25) for the instance search task and found that the IDF was not discriminative enough to retrieve relevant videos. The BM25 method was applied to the video retrieval task as follows:

$$P(\mathrm{rel}|q, v) \tag{26}$$

Here, q and v now represent keypoints from the image query and those from the video keyframes, both represented by keypoint vectors whose elements are within-image and within-video-keyframes keypoint frequencies. Keypoints can be detected using any keypoint detectors such as the Harris-Laplace detector [21]. Then, Eq. (26) becomes the following

BM25 video ranking function since the same assumptions for the text/document search task are also applicable to the image/video search task.

$$P(\mathrm{rel}|q,v) \propto_q \sum_{q_i} \frac{v_i}{v_i+\kappa} \log\left(\frac{N-n_i+0.5}{n_i+0.5}\right) \tag{27}$$

Here, $\sum_{q_i}$ means the summation over keypoints within q that are also present in v, $\kappa = 2$, which is the same value adopted for a text/document retrieval task, and N and n_i are now the total number of videos in the video database to be searched and the number of videos that contain the ith query keypoint (called video frequency). Note that to use Eq. (27), I need to determine the keypoint correspondence (matching) among keypoints detected in an image query and video keyframes. For the text/document retrieval task, there are no such problems since I can simply regard the same keyword pair as matched. For the image/video retrieval task, since the keypoints are usually described by high-dimensional feature vectors such as 128-dimensional vectors for the SIFT features [20], the keypoint correspondence is evaluated based on the cosine similarity between the two keypoint vectors. That is, the keypoint pair whose cosine similarity is over a pre-defined threshold value, such as 0.9 are considered as matched, and v_i and n_i in Eq. (27) can be determined in this matching procedure. Interestingly, Iwamura et al. mentioned that this matching process is known as the bichromatic reverse nearest neighbor search problem, which has been extensively studied [51] [52].

Using Eq. (27) for the instance search task did not show satisfactory results and search accuracy. The main reason was the low n_i and resulting relatively high discriminative powers for almost all query keypoints. In contrast to the text/document retrieval task, since the query keypoints are described by high-dimensional vectors, similar keypoint pairs do not often appear in the video database, leading to low n_i Therefore, I alternatively used the following new discriminative power:

$$\log\left(\frac{e^{-n_i/\xi'}}{n_i - e^{-n_i/\xi'}} \frac{\left(N - e^{n_i/\xi'} - n_i + e^{-n_i/\xi'}\right)}{\left(e^{n_i/\xi'} - e^{-n_i/\xi'}\right)}\right) \tag{28}$$

Here, ξ' is a newly introduced parameter depending on the database size. Since the new weight rapidly becomes small as n_i becomes relatively large, it significantly suppresses the discriminative powers for query keypoints showing the frequently appearing tendency. I called the new weight in Eq. (28) the EIDF and I used the BM25 video ranking function using the EIDF for the TRECVID instance search task dataset and confirmed the significant effectiveness.

D. Research Questions and Contributions

Following the above sections, there are two issues left unaddressed. The first research question concerns the theoretical interpretation for the EIDF in Eq. (28). In the previous chapter, I just proposed to use the new weight in the framework of BM25 to improve the instance search task; thus, I need to conduct its theoretical analysis to understand the meaning much more clearly. The second research question is regarding the use of the ROI images shown in Figure 7. Since the instance search system is supposed to accept both original and ROI images and the latter information is certainly helpful to clarify the region of instance within the original image, in this chapter, I devise two ranking methods that incorporate the

ROI information into my BM25 framework. To summarize, the main contributions of this chapter are listed as follows:

(1) Theoretical derivation and interpretation of the EIDF are provided.

(2) ROI information is incorporated into the BM25 ranking strategy to further enhance instance search accuracy.

(3) The proposed methods are verified using a series of TRECVID datasets comprising the TRECVID2012, 2013, and 2014 instance search tasks.

The rest of this chapter is organized as follows. In Section 3, I describe the Bayesian derivation of the discriminative power and interpret the EIDF from this viewpoint. I next discuss the ROI information for the instance search task and propose two ranking methods incorporating the ROI effect in Section 4. In Section 5, I present the instance search experiments, including the datasets used, results obtained, and evaluations determined. Finally, in Sections 6 and 7, I summarize related work and basic findings of this paper.

3. Bayesian Discriminative Power

To address the first research question described in the previous section, I first explain the concept of eliteness affecting the discriminative power. I next describe the mathematical interpretation for the BM25 IDF from the Bayesian point of view and derive the Bayesian counterpart. I then insist that choosing an informative probability density function as the prior distribution for the keypoint becoming elite leads to a new discriminative power. The new discriminative power is called the Bayesian EIDF (BEIDF), which is quite similar to the EIDF. The theoretical comparison between the BEIDF and EIDF is also provided.

A. Eliteness Describing Discriminative Power

Equation (27) is originally based on the following equation [4]:

$$P(\text{rel}|q, v) \propto_q \sum_{q_i} \frac{v_i}{v_i + \kappa} \log \left(\frac{P(e_i|\text{rel})P(\bar{e}_i|\text{irrel})}{P(e_i|\text{irrel})P(\bar{e}_i|\text{rel})} \right) \tag{29}$$

Here, $P(e_i|\text{rel})$ and $P(e_i|\text{irrel})$ are the probabilities that q_i becomes elite (aboutness) e_i in videos relevant to q and that q_i becomes elite e_i in the irrelevant videos, respectively. The terms $P(e_i|\text{rel})$ and $P(e_i|\text{irrel})$ are the probabilities that q_i becomes non-elite (non-aboutness) e_i in relevant videos and that it becomes non-elite $\bar{e}_i$ in irrelevant videos. Note that $P(\bar{e}_i|\text{rel}) = 1 - P(e_i|\text{rel})$ and $P(\bar{e}_i|\text{irrel}) = 1 - P(e_i|\text{irrel})$. The eliteness or non-eliteness is a property assigned for each keypoint and it is regarded as affecting the actual frequency within a video.

When I assume that $P(e_i|\text{rel}) \approx \frac{r_i+b}{R+a}$, $P(e_i|\text{irrel}) \approx \frac{n_i-r_i-b}{N-R-a}$ where R is the number of all relevant videos to q in the database and r_i is the number of relevant videos containing q_i, and setting $R = r_i = 0, a = -1, b = -0.5$ yields the BM25 IDF shown in Eq. (27). Note that the information of relevant videos is usually unobtainable prior to searching, and I set $R = r_i = 0$ in the above derivation. Setting $R = r_i = 0, a = e^{n_i/\xi'}, b = e^{-n_i/\xi'}$ leads to the EIDF in Eq. (28).

B. Bayesian IDF

In this section, I discuss the estimation of $P(e_i|\text{rel})$ in Eq. (29) based on the Bayes' theorem. I use m_i to denote $P(e_i)$, which is the random variable of the probability of q_i

becoming elite. Therefore, the domain of m_i is [0,1]. I then assume that $E[m_i|r_i, R]$ is a good approximation for $P(e_i|\mathrm{rel})$, which is the expectation of m_i given relevant information to q. Here, r_i and R are realizations, that is, the actual number of relevant videos containing the *i*th keypoint and the total number of relevant videos. Then, the following Bayes' theorem holds:

$$p(m_i|r_i, R) = \frac{p(r_i|R_i, m_i)p(m_i)}{p(r_i|R)} \tag{30}$$

Here, $p(m_i) = p(m_i|R)$. By assuming $p(m_i) \sim U(0,1)$ (m_i is assumed to follow a continuous uniform distribution with the support of [0,1]) and that $p(r_i|R_i, m_i)$ follows a binomial distribution $Binomial(R, p)$, the posterior distribution of m_i theoretically becomes $Beta(r_i + 1, R - r_i + 1)$. In other words, when I assume that the prior distribution of m_i follows the uniform distribution with the expectation of 0.5, that is, a non-informative prior distribution often adopted when no information is available, the posterior distribution $p(m_i|r_i, R)$ becomes $Beta(r_i + 1, R - r_i + 1)$. Then, the posterior expectation is shown as follows.

$$E[m_i|r_i, R] = \frac{r_i + 1}{R + 2} \tag{31}$$

I thus obtain $P(e_i|\mathrm{rel}) = (r_i + 1)/(R + 2)$. As mentioned, $P(\bar{e}_i|\mathrm{rel})$ is calculated by $1 - P(e_i|\mathrm{rel}) = 1 - (r_i + 1)/(R + 2)$. Comparing the assumption $P(e_i|\mathrm{rel}) \approx \frac{r_i+b}{R+a}$ used for deriving the BM25 IDF, the aforementioned Bayesian estimation approach using the uniform prior distribution theoretically determines the unknown parameters a and b.

In the same way, $P(e_i|\mathrm{irrel})$ is estimated with the following procedure:

$$p(m_i|n_i - r_i, N - R) = \frac{p(n_i - r_i|N - R_i, m_i)p(m_i)}{p(n_i - r_i|N - R)} \tag{32}$$

Then, the posterior expectation becomes

$$E[m_i|n_i - r_i, N - R] = \frac{n_i - r_i + 1}{N - R + 2} \tag{33}$$

Note that $P(\bar{e}_i|\text{irrel})$ becomes $1 - P(e_i|\text{irrel}) = 1 - (n_i - r_i + 1)/(N - R + 2)$.

Substituting these results into Eq. (29) yields the following computable BM25 video ranking function using the Bayesian IDF (BIDF) under the same setting of $R = r_i = 0$.

$$\begin{aligned} P(\text{rel}|q, v) &\propto_q \sum_{q_i} \frac{v_i}{v_i + \kappa} \log\left(\frac{(r_i + 1)(N - R - n_i + r_i + 1)}{(n_i - r_i + 1)(R - r_i + 1)}\right) \\ &\approx \sum_{q_i} \frac{v'_i}{v'_i + \kappa} \log\left(\frac{N - n_i + 1}{n_i + 1}\right) \end{aligned} \tag{34}$$

Here, $\log(\cdot)$ is called the BIDF for q_i, showing the quite similar form as the BM25 IDF weight. In the case of $\log(\cdot) < 0$, I usually set $\log(\cdot) = 0$. The term v'_i is v_i normalized by the video length vl of v. Following a text/document retrieval task, I regard vl as the total number of keypoints extracted from all of the video keyframes. Although there are various normalization methods, I use the widely used pivoted length normalization $v_i' = v_i/\big(1 - c + c(vl/avdl)\big)$, where c = 0.75, vl is the total number of keypoint frequencies in v, as mentioned earlier, and $avdl$ is the average vl in the video database.

Equation (34) clearly shows that the larger v'_i does not greatly affect the video ranking score since that factor just approaches 1 at maximum. The important factor is the BIDF weight of the keypoint, and according to the equation, videos including high BIDF query keypoints tend to be more likely those that the user is searching for. This is the formulation of the BM25 video ranking function using the BIDF.

C. Bayesian Exponential IDF

I next explain the derivation of the BEIDF. In so doing, I adopt a different prior distribution from the uniform distribution $Beta(1,1)$, which was used to derive the BIDF. Before explaining this approach, I consider the following, general prior distribution.

$$p(m_i) \sim Beta(\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} m_i^{\alpha-1}(1-m_i)^{\beta-1}$$

Then, since $p(r_i|R,m_i) \sim Binomial(R,m_i)$, $p(r_i|R)$ in Eq. (30) becomes

$$\begin{aligned} p(r_i|R) &= \frac{\Gamma(R+1)}{\Gamma(r_i+1)\Gamma(R-r_i+1)\beta(\alpha,\beta)} \\ &\quad \times \int_0^1 m_i^{r_i+\alpha-1}(1-m_i)^{R-r_i+\beta-1} dm_i \\ &= \frac{\Gamma(R+1)}{\Gamma(r_i+1)\Gamma(R-r_i+1)\beta(\alpha,\beta)} \\ &\quad \times \frac{\Gamma(r_i+\alpha)\Gamma(R-r_i+\beta)}{\Gamma(r_i+\alpha+\beta)} \end{aligned}$$

Thus, the posterior distribution of $p(m_i|r_i,R)$ (the left side of Eq. (30)) becomes

$$\begin{aligned} p(m_i|r_i,R) &= \frac{\Gamma(r_i+\alpha+\beta)}{\Gamma(r_i+\alpha)\Gamma(N-r_i+\beta)} \\ &\quad \times m_i^{r_i+\alpha-1}(1-m_i)^{R-r_i+\beta-1} \\ &\sim Beta(r_i+\alpha, R-r_i+\beta) \end{aligned}$$

This implies that the probability distribution of interest changes from $Beta(\alpha, \beta)$ to $Beta(r_i + \alpha, R - r_i + \beta)$ after observing the relevance information r_i and R. Its expectation becomes

$$E[m_i|r_i, R] = \frac{r_i + \alpha}{R + \alpha + \beta}$$

Then, I adopt $p(m_i) \sim Beta\left(e^{-n_i/\gamma}, e^{n_i/\gamma} - e^{-n_i/\gamma} + 1\right)$. This prior distribution indicates that when q_i often appears in many videos (that is, for the query keypoint with large n_i), in other words, for a frequently appearing query keypoint, its probability of becoming elite is considerably lowered. Therefore, it corresponds to the informative prior distribution, implying that the keypoints often appearing in many kinds of videos do not become elite in videos and reflecting the idea of the EIDF. This fact can be confirmed by taking the expectation and variance of the prior distribution $p(m_i)$.

$$E[m_i] = \frac{e^{-n_i/\gamma}}{e^{n_i/\gamma} + 1} \ , V[m_i] = \frac{e^{-n_i/\gamma}\left(e^{n_i/\gamma} - e^{-n_i/\gamma} + 1\right)}{(e^{n_i/\gamma} + 1)^2(e^{n_i/\gamma} + 2)}$$

The above equations both rapidly approach 0 as n_i becomes large; thus, for that case, q_i is regarded never to become elite. Figure 8 plots the prior distributions when $n_i = 100, \gamma = 1000$ and $n_i = 500, \gamma = 1000$. Since the query keypoint q_j often appears in videos, its probability of becoming elite is sufficiently lowered compared to that for q_i. The γ depends on the number of database videos used to calculate n_i and is a parameter for this newly enhanced discriminative power. This prior design aims to reduce search result errors from such often appearing, less discriminative query keypoints. Generally speaking, since such keypoints are often regarding the background information in the original image query, keypoints representing the instance often tend to acquire more discriminative powers by this prior distribution design using an informative distribution. Since my objective is to search for the instance, not for the exact or similar videos to the instance image queries, instance search accuracy is expected to improve by using this enhanced discriminative power.

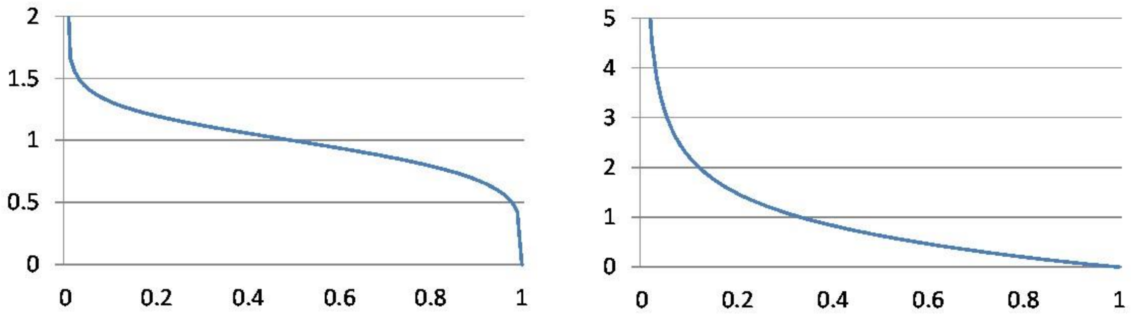


Figure 8: (Left) $n_i = 100, \gamma = 1000$**. (Right)** $n_i = 500, \gamma = 1000$. **X and Y axes are** $m_{i\ or\ j}$ **and** $p(m_{i\ or\ j})$**, respectively. (Copyright©2016 IEICE, [R2] Fig. 2)**

The corresponding expectations for the posterior distributions become

$$P(e_i|\text{rel}) \approx E[m_i|r_i, R] = \frac{r_i + e^{-n_i/\gamma}}{R + e^{n_i/\gamma} + 1},$$

$$P(e_i|\text{irrel}) \approx E[m_i|n_i - r_i, N - R]$$

$$= \frac{n_i - r_i + e^{-n_i/\gamma}}{N - R + e^{n_i/\gamma} + 1}$$

Substituting these and the results for $P(\bar{e}_i|\text{rel}) = 1 - P(e_i|\text{rel})$ and $P(\bar{e}_i|\text{irrel}) = 1 - P(e_i|\text{irrel})$ into Eq. (29) yields the following new BM25 video ranking function using the BEIDF.

$$P(\text{rel}|q, v) \propto_{\text{q}} \sum_{q_i} \frac{v_i'}{v_i' + \kappa}$$

$$\times \log\left(\frac{e^{-n_i/\gamma}}{(n_i - e^{-n_i/\gamma})} \frac{\left(N - e^{\frac{n_i}{\gamma}} - n_i + e^{\frac{-n_i}{\gamma}} + 1\right)}{(e^{n_i/\gamma} - e^{-n_i/\gamma} + 1)}\right) \quad (35)$$

As shown from the above $\log(\cdot)$, the discriminative power of the BEIDF is enhanced compared to the BIDF. As a result, unnecessary query keypoints for instance searching whose weights are not sufficiently lowered by using the standard/Bayesian IDF can be adequately suppressed using this enhanced weight.

D. Comparison with EIDF

Recalling the EIDF shown in Eq. (28), the BEIDF in Eq. (35) retains the following theoretical advantages:

1. The value inside the logarithmic function does not tend to become negative as n_i becomes large.
2. It asymptotically approaches the BIDF in Eq. (34) as $\gamma \rightarrow \infty$.
3. It becomes the maximum value of $\log(N+1)$ for the query keypoint whose n_i is equal to zero.

The first feature has been desirable since the logarithmic function is not defined for the negative domain. The second feature clarifies the relationship between the enhanced weight and standard weight. This fact is obvious since for the case of $\gamma \rightarrow \infty$, the prior distribution for deriving the BEIDF approaches $Beta(1,1)$, which corresponds to the uniform prior distribution used for deriving the BIDF. The third feature is also preferable since the EIDF could not be evaluated for the query keypoint with $n_i = 0$ (see Eq. (28)).

In this section, I proposed a method to actually calculate the discriminative power in the BM25 formula and provided the Bayesian discriminative powers (BIDF and BEIDF). In so doing, I assumed that the $m_i = p(e_i)$ followed a Beta distribution $Beta(\alpha, \beta)$ and the conditional expectation $E[m_i | r_i, R]$ was a good approximation for $p(e_i | \text{rel})$. These assumptions generally hold since the Beta distribution is suitable as a prior distribution for a parameter whose domain is [0,1]. In addition, the conditional expectation is proved to be the unbiased estimate. Therefore, $E[m_i|r_i, R] \approx p(e_i|\mathrm{rel}) = \frac{r_i+\alpha}{R+\alpha+\beta}$. It also applies to the $E[m_i|n_i - r_i, N - R]$ as an approximation for $p(e_i|\mathrm{irrel})$.

As explained in Section 3-C, the prior distribution design for the BEIDF was based on the basic idea of the EIDF such that the keypoints often appearing in many kinds of videos do not become elite in videos. Therefore, the resulting BEIDF retains the same property and makes it possible to interpret the EIDF from the Bayesian viewpoint.

My basic video ranking method for the instance search task is based on Eq. (35). However, the ROI information is not incorporated into the ranking function yet. The next section describes the use of ROI and explains BM25 using the ROI effect to further improve instance search accuracy.

4. Instance Search using ROI

As explained and illustrated in Section B and Fig. 6, the instance search system is supposed to accept both original and ROI image queries. Since the ROI is supplied to specify the instance to be searched within the original image, the ROI information should be effective for further improving search accuracy. In this section, I propose two methods called ROI weighting and ROI re-ranking that incorporate the ROI effect into their video ranking functions.

A. ROI Weighting

I first explain the method that puts larger weights on the query keypoints detected within the ROI. As mentioned above, the query keypoints within the ROI are supposed to be emphasized more since the ROI clearly specifies the instance to be searched in the original image query. To incorporate this effect into the BM25 video ranking function, I modify Eq. (35) as follows:

$$P(\mathrm{rel}|q,v) \propto_{\mathrm{q}} \sum_{q_i} \frac{v_i'}{v_i'+\kappa} \times \mathrm{ROI_i} \log\left(\frac{e^{-n_i/\gamma}}{(n_i - e^{-n_i/\gamma})} \frac{\left(N - e^{\frac{n_i}{\gamma}} - n_i + e^{\frac{-n_i}{\gamma}} + 1\right)}{(e^{n_i/\gamma} - e^{-n_i/\gamma} + 1)}\right) \tag{36}$$

Here, $\mathrm{ROI_i}$ is set to

$$\mathrm{ROI_i} = \begin{cases} \lambda \;\; (\text{if } q_i \text{ is within the ROI}) \\ \quad 1 \;\; (\text{otherwise}) \end{cases}$$

and $\lambda \geq 1$ is a parameter and specifies to what extent I emphasize the query keypoints within the ROI. Its optimal value might depend on the instance to be searched since for some instances the background information becomes quite useful because there are some instances that frequently appear in the same environment. For example, a logo of a shop keeps co-occurring with the same shop. In that case, the background information becomes useful and contributes to improving the instance search accuracy further. On the other hand, when instances do not relate to the environment, the use of the background hurts the retrieval accuracy. However, I set the constant value for all of the instance topics since I have no clue for the λ value prior to searching. I call this method ROI weighting.

B. ROI Re-ranking

One problem with ROI weighting is a possible topic drift that may occur during the search process. As shown in Eq. (36), the query keypoints are now weighted by the BEIDF multiplied by the ROI effect. Although it depends on the λ value, some query keypoints outside the ROI may gain higher weights than the keypoints inside the ROI since the BEIDF can dominate the ROI effect. As a result, even though the user specifies the instance to be searched using the ROI information, the resulting search results may become relevant to the objects outside the ROI and may not be relevant to the instance itself. I call this critical issue a (instance) topic drift.

I can alleviate this problem by simply enlarging λ, but this simultaneously means that the background information outside the ROI is neglected more. Some instances are very much related to the background and for these cases, enlarging λ results in decreasing overall instance search accuracy. In short, λ is a tuning parameter that highly depends on the instance topic and, generally speaking, its adaptive setting is quite difficult. I, therefore, need to adopt a constant value for every instance topic, and there is always a possibility of experiencing a severe topic drift. Following this observation, I describe a method that is especially designed to prevent or alleviate such topic drift from occurring in the instance search result ranking.

The method is based on a three-stage search result reranking using the ROI information. The procedure is described as follows:

(1) First Step

I first generate the initial search results by using only the query keypoints inside the ROI. That is, I first rank the videos in the database according to $P(\mathrm{Rel} = \mathrm{rel} | q_{\mathrm{ROI}}, v)$, where q_{ROI}

is the vector of query keypoints inside the ROI. The $P(\text{Rel} = \text{rel}|q_{\text{ROI}}, v)$ is evaluated by replacing q on the right side of Eq. (35) with q_{ROI}. Theoretically speaking, with this first procedure, the database videos are ordered by their relevance probabilities to the instance itself.

(2) Second Step

Then, for the top K search results, I next estimate the relevance probabilities to the background information in the original query image. That is, I calculate $P(\text{Rel} = \text{rel}|q', v)$, where q' is the vector of query keypoints outside the ROI. Again, $P(\text{Rel} = \text{rel}|q', v)$ is evaluated by replacing q on the right side of Eq. (35) with q''.

(3) Third Step

Finally, the initial video result ranking is re-ranked according to the following combination scores:

$$score(q, v) = P(\text{rel}|q_{\text{ROI}}, v) + \tau P(\text{rel}|q', v) \qquad (37)$$

Here, τ is a parameter and $\tau = 0$ corresponds to the no re-ranking strategy. Generally speaking, since the query keypoints outside the ROI are supposed to be less important than the keypoints inside the ROI, τ is smaller than 1. The scores for videos ranked higher than K are set to $P(\text{Rel} = \text{rel}|q_{\text{ROI}}, v)$.

Since this three-stage re-ranking method first retrieves and determines videos that have relatively high relevance probabilities to the instance and only evaluates the overall relevance probabilities to the original image query (instance and background) for the top-K ranked videos, the aforementioned topic drift is expected to be alleviated more than by simply using ROI weighting in Eq. (36). The instance search methods using ROI weighting and ROI re-ranking are both evaluated in the next section using the TRECVID instance search task dataset.

C. When the ROI is not available

This section assumed that the ROI information was simultaneously input together with the original images and I called such video retrieval task instance search. Although the primary objective of this paper is to improve the instance search accuracy, the proposed method based on the BIDF or BEIDF can also be applied to the video retrieval task without such a ROI information. Indeed, to be shown in the next section of experiments, the proposed method without the use of ROI also achieves the high retrieval accuracy. Although the deterioration of the search accuracy depends on the dataset to be searched, the proposed method without the use of ROI also provides satisfactory results. This is due to the enhanced discriminative power estimated by the BEIDF and the proposal of such Bayesian discriminative powers is the primary contribution of this paper.

5. Experiments

I evaluated the proposed methods using the instance search task dataset at TRECVID2012, 2013, and 2014. In this section, I first describe the dataset and give query examples. I next confirm that, as well as the EIDF, the BEIDF is also superior over the BIDF in the framework of BM25; therefore, it is effective for enhancing the instance search accuracy. I then discuss the search accuracy of the two proposed instance search methods incorporating ROI effects and give examples for the actual search results.

A. Instance Search Task Dataset

I first describe the dataset used for the experiments. There are 21 instances (15 objects, 1 person, 5 places) for the TRECVID2012 dataset and five original and ROI images are provided on average for each instance. The database videos are about 77,000 clip movies uploaded to Flickr.com (the average duration is about 10 seconds). For the TRECVID2013 dataset, 30 instances (26 objects, 4 people, 0 places) with four original and ROI images for each instance are provided. The database videos are about 470,000 shot videos from a BBC drama (the average duration is between 1 ~ 3 seconds). The query examples are shown in Fig. 7 in Section 2. The TRECVID2014 dataset is composed of 27 instances (21 objects, 5 people, 1 place) with four original and ROI images for each instance. The database videos are the same as those for the TRECVID2013 dataset. Figure 9 shows the query examples for this dataset.

The database videos are judged as relevant or irrelevant to each instance topic and there are some videos that are not judged. In evaluating the search accuracy, I simply remove these un-judged videos and put the lower ranked videos higher in the search result ranking. The measure for the search accuracy is MAP, which has been widely used to evaluate the ranking accuracy of a search system using the binary relevance data. For this experiment, the MAP values were evaluated for the top 1000 search results for each instance topic.

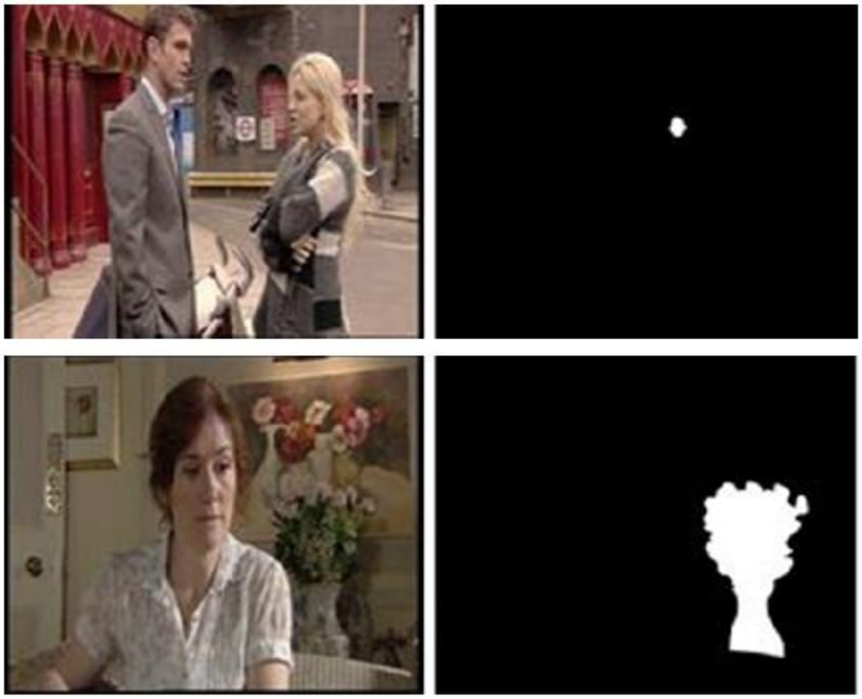


Figure 9: Original and ROI images for "London underground logo" and "Vase with this flower". (Copyright©2016 IEICE, [R2] Fig. 3)

B. Investigating BEIDF

In my instance search methodology, keyframe images were taken by a 1 frame/sec ratio from database videos. Then, keypoints were extracted using the Harris-Laplace detector [21] from the query images and keyframe images, and those extracted from the query images were aggregated as the query keypoints. The keypoints extracted from the keyframe images were aggregated as the video keypoints. All the keypoints were featured by 128-dimensional SIFT vectors [20] and 192-dimensional color SIFT (CSIFT) vectors [21]. The threshold for the keypoint matching was set to 0.9. The search result rankings were generated according to Eqs. (34) and (35), that is, they were generated using BM25 using BIDF and BEIDF. Note that two

BM25 ranking scores were calculated for using SIFT and CSIFT features, respectively, and these scores were added to generate the instance search results for each instance topic.

The evaluation results are listed in Table 3. The γ value in Eq. (35) for the BEIDF was set to 100. As explained in subsection 3-C, the γ depends on the number of database video used to calculate n_i (video frequency). I varied the γ and adopted the value that scored the high retrieval accuracy. In Table 3, I can confirm the significant improvement in the MAP value for the BEIDF. Comparing with the other methods, the MAP of 0.37 is quite high. The method of Zhu et al. used the query-adaptive asymmetrical dissimilarities and was the improved method that won the TRECVID2012 instance search task. The second comparison method utilized a spatial verification method to improve the instance search accuracy further. Table 3 clearly shows the significant improvement in the search accuracy by applying the proposed BEIDF.

Table 3: Evaluation results for instance search accuracy using TRECVID2012 dataset (abbreviated as TV12). BIDF and BEIDF stand for BM25 using Bayesian IDF and Bayesian exponential IDF. The comparison methods are the improved method of the winner at the TRECVID2012 instance search task (denoted as Zhu)[53] and that used the topological spatial verification method (denoted as Zhang)[54]. Bold number indicates maximum value. (Copyright©2016 IEICE, [R2] Table 1)

	BIDF	BEIDF	Zhu[53]	Zhang[54]
MAP (TV12)	0.25	0.37	0.27	0.22

C. Search Accuracy Enhancement Using ROI

I next confirm the impact of ROI information on the improvement in instance search accuracy. The MAP results are shown below. For the TRECVID2012 dataset, the λ values in Eq. (36) for BM25 using the BIDF with ROI weighting (BIDF+RW) and BM25 using the BEIDF with ROI weighting (BEIDF+RW) were both set to 2. Note that replacing the BEIDF in Eq. (36) with the BIDF yields the method of BIDF+RW. For the TRECVID2013 and 2014 datasets, the γ values were set to 1000 and 100, respectively, and the λ values were both set to 10. Again, these γ and λ values are the ones that showed the high search accuracy and I chose them by simply varying the parameter values.

Table 4: Evaluation results for instance search accuracy using TRECVID2012, 2013, and 2014 datasets (abbreviated as TV12, TV13, and TV14). BIDF+RW and BEIDF+RW stand for BIDF with ROI weighting and BEIDF with ROI weighting, respectively. Bold numbers indicate maximum values. (Copyright©2016 IEICE, [R2] Table 2)

	BIDF+RW	BEIDF+RW
MAP (TV12)	0.28	0.38
MAP (TV13)	0.26	0.31
MAP (TV14)	0.16	0.27

From Table 4, I first observed that the use of ROI as the additional weighting led to higher search accuracy (compared with the results in Table 3). I next confirmed that using both enhanced discriminative power and ROI effect significantly improved instance search accuracy. The MAP of 0.38 of BEIDF+RW was much higher than the official highest MAP of 0.27 scored by the aforementioned method of Zhu et al. [53] during the TREVID2012 instance search task.

D. Actual Instance Search Results

I describe the effectiveness of the proposed methods by illustrating some actual instance search results. I first show the result examples for the instance topic “Eiffel tower” in Fig. 10. The left result ranking was obtained using BM25 using the BIDF and the right result ranking was produced with BM25 using the BEIDF. I can see that the BIDF could not suppress the weights for query keypoints at the background in the original image query. As a result, the videos showing similar background to the original image query were ranked high, leading to significantly lowered instance search accuracy. On the other hand, the BEIDF successfully assigned lower weights for such noisy keypoints, in other words, it automatically interpreted that the keypoints detected from the background region were less discriminative compared to the keypoints regarding the Eiffel tower, resulting in much better instance search result ranking.

I next discuss an example for the instance topic “This dog” in Fig. 11. The left result ranking was obtained using the BEIDF and the right result ranking was for the BEIDF+RW. The results shown in Fig. 11 clearly illustrate the effectiveness of the ROI information. Although the top 3 ranked videos were both correct to the instance topic, the BEIDF could not retrieve the relevant ones from the 4th to 8th rankings.

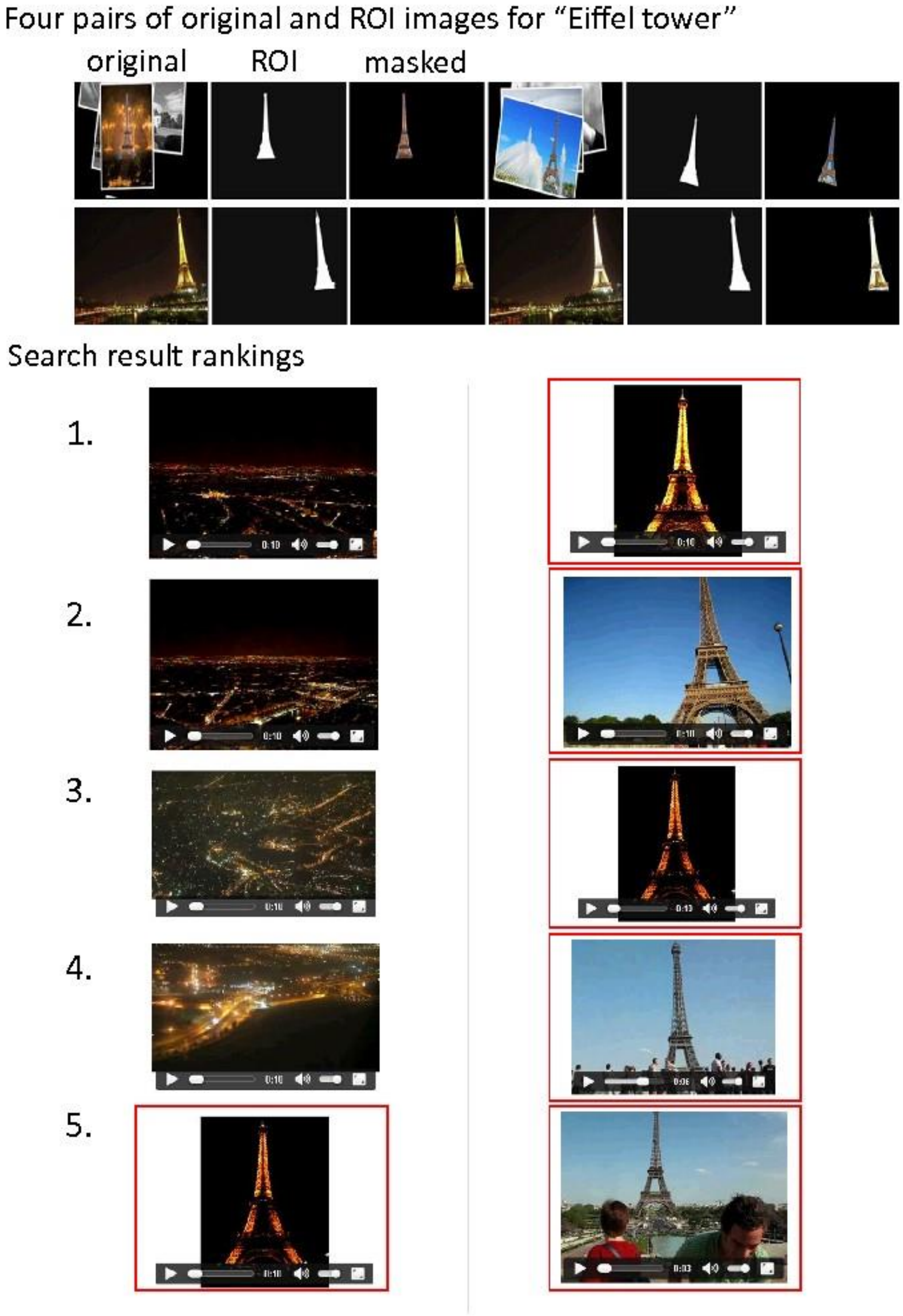


Figure 10: Instance search result rankings for “Eiffel tower”. Left ranking is for BIDF and right ranking is for BEIDF. Videos with red borders are correct videos. (Copyright©2016 IEICE, [R2] Fig. 4)

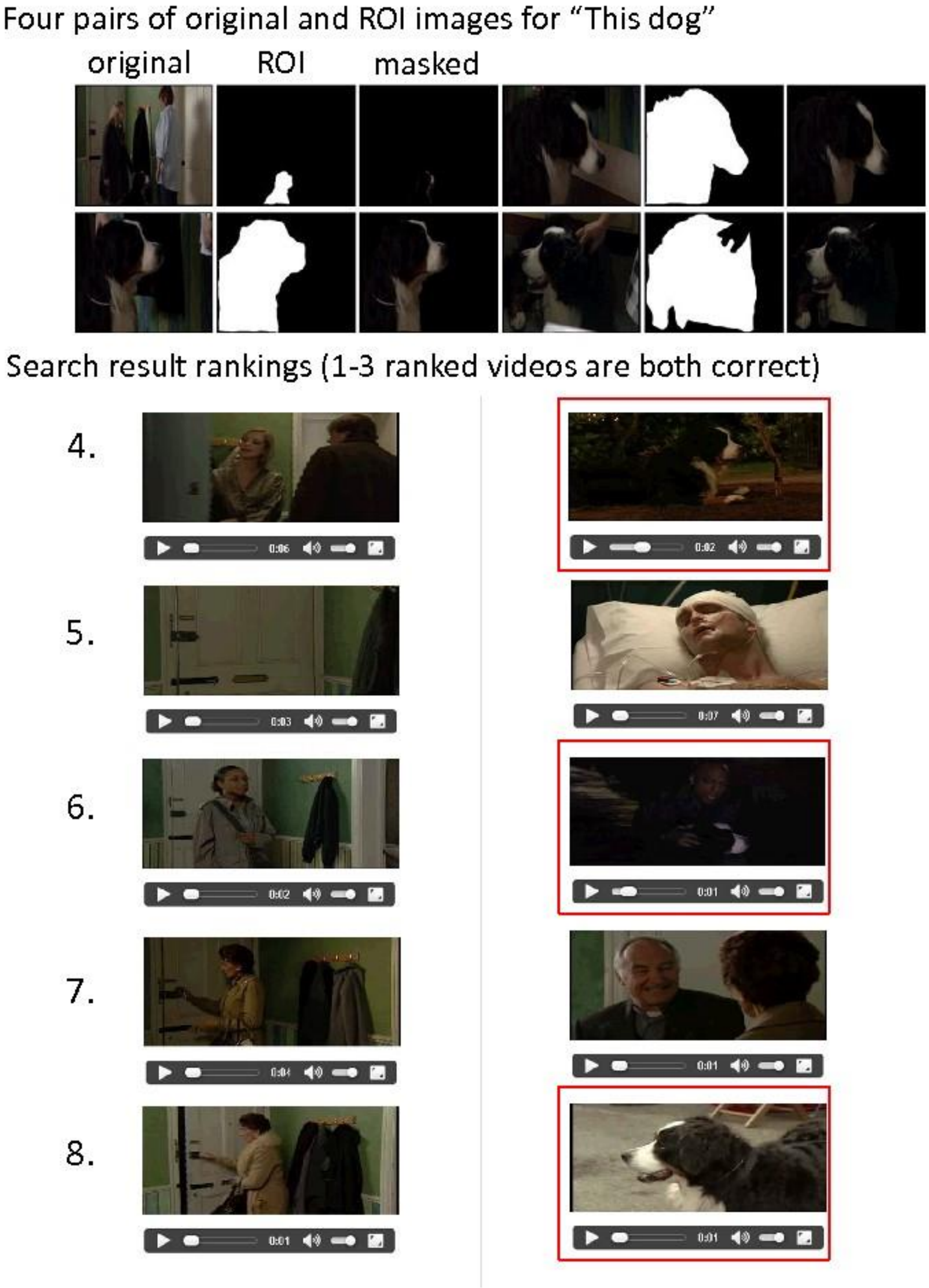

Four pairs of original and ROI images for "This dog"
original
ROI
masked
Search result rankings (1-3 ranked videos are both correct)
4.
0:06
0:02
5.
0:03
0:07
6.
0:02
0:01
7.
0:04
0:01
8.
0:01
0:01

Figure 11: Instance search result rankings for "This dog". Left ranking is for BEIDF and right ranking is for BEIDF+RW. Videos with red borders are correct videos. (Copyright©2016 IEICE, [R2] Fig. 5)

The videos ranked 4th to 8th were relevant to the background of the original image query, and it is obvious that the BEIDF assigned relatively large weights for the query keypoints extracted from the background compared to the keypoints inside the ROI. Using the ROI information and weighting the query keypoints inside the ROI alleviated this problem more and the resulting instance search accuracy significantly improved.

E. Evaluating ROI Re-ranking Method

From the result examples shown in Figs. 10 and 11, I can say that using both enhanced discriminative power BEIDF and ROI weighting are effective in improving instance search accuracy. However, as mentioned in section 4-B, one problem is that the ROI weighting method is always affected by the possible topic drift occurring in the search result ranking. I next confirm how much accuracy can be improved by using the three-stage re-ranking method in Eq. (37).

The evaluation results are listed in Table 5. The τ in Eq. (37) was set to 1/10 since the query keypoints outside the ROI are generally supposed to be less important than the keypoints inside the ROI. In this experiment, re-ranking was done for the top 30 initial search results. From Table 5, I can confirm that additional improvement in instance search accuracy is expected using the proposed ROI re-ranking method. Indeed, the MAP value of 0.28 scored with the BEIDF with ROI re-ranking (BEIDF+RR) was high compared to those scored with the other participating teams in the TRECVID2014 instance search task (the MAP of 0.28 was ranked within the top 3 search accuracies achieved). The highest MAP officially scored at the TRECVID2014 instance search task was 0.32 [55] and the winning method combined many techniques, including bag-of-visual-words (BOVW), geometric verification (called the RANSAC), and object detectors established for the instance search.

Table 5: Evaluation results for instance search accuracy using TRECVID2014 dataset (abbreviated as TV14). BEIDF+RR stands for BEIDF with ROI re-ranking. The bold number indicates maximum value. (Copyright©2016 IEICE, [R2] Table 3)

	BEIDF	BEIDF+RW	BEIDF+RR
MAP (TV14)	0.24	0.27	0.28

The overall results discussed in this section support the effectiveness of the proposed methods. The BEIDF is theoretically more rigorous than the EIDF, and using the enhanced discriminative power, I confirmed that significant improvement in the instance search results was achieved in the BM25 framework. Using the ROI information was expected to prevent the topic drift that may occur in the search result rankings and the proposed ROI weighting and re-ranking methods were both found to be effective.

6. Related Work

In the previous sections, I described the theoretical derivation and interpretation for the EIDF previously proposed and the use of ROI information to improve instance search accuracy. Generally speaking, research related to the first topic has been actively discussed in the text/document retrieval community and a recent study showing the effectiveness of the BM25 theory is discussed in the following section. For the second topic, I discuss studies regarding the instance search task in general. Note that approaches taken in these papers are essentially different from the proposed approach. The proposed approaches are motivated by the application and extension of the probabilistic IR methods originally developed in the document (text) retrieval community to content-based image/video retrieval tasks including the instance search task.

A. BM25 and the Probabilistic IR Models

Wilkie et al. conducted a comprehensive empirical evaluation for widely known probabilistic IR methods such as BM25, language model based IR (LM) [4], divergence from randomness (DFR) [8], and divergence from independence (DFI) [14] models using the TREC test collections [56]. The results were quite impressive. Although BM25 was the oldest probabilistic IR model among the models compared, they experimentally showed that it generally exhibited the least bias on the collection and produced competitive retrieval performance. Their results clearly support the fact that BM25 has been attracting much attention from IR researchers and that it is still the representative state-of-the-art probabilistic IR model. This fact supports the use of the BM25 model even for multimedia search tasks and, indeed, directed us to employ the BM25 video ranking function for addressing the instance search task.

B. Recent Developments in Instance Search

Zhu et al. compared the BOVW framework [33] with the Approximate Nearest-Neighbor (ANN) [35] based system for the instance search task [57]. Since the ANN system is quantization-free, the performance loss caused by the quantization error in the BOW framework can be estimated. Their experiments showed that the vector quantization was the bottleneck of the BOW framework and that using a reference dataset as the prior distribution of local features improved the retrieval performance of the ANN voting system. Zhu et al. also experimentally demonstrated that comparing the query and keyframe images based on the standard symmetrical measures decreased instance search accuracy and proposed to use the asymmetric measure adapted to the BOW framework [53].

Zhang et al. proposed an elastic spatial verification method for the instance search task [54] [58]. Their method was designed to elastically verify the topological spatial consistency, which is invariant to various spatial transformations, with a triangulated graph and verified its effectiveness using three years of TRECVID instance search task datasets. Zhou et al. investigated the effectiveness of the spatial re-ranking method for the instance search task [59]. Since frame-by-frame spatial verification is too prohibitive, they sped up the algorithm by selecting the most representative keyframe of videos and proposed an efficient RANSAC algorithm. They also proposed an ROI-originated RANSAC method in which the geometrical transformation matrix is computed based on the known ROI information, leading to further improvement in re-ranking performance.

Tao et al. tackled the problem of object search by including locality at all steps of their retrieval method [60]. They focused on the locality in an image and that in the feature space. With their method, many candidate locality boxes are generated in the database images, enabling the local search in the feature space constructed using the Fisher vector [61] and VLAD [62]. The effectiveness of their method was experimentally validated using three datasets including the Oxford buildings [35], and the method also successfully delivered reliable localization results. Meng et al. presented an approach to localize a specific object in videos using a single image query [63]. They formulated the problem as a spatio-temporal search of the optimal object trajectories in videos. The experimental results showed the effectiveness of their approach by comparing them with the results from current methods with which each keyframe is treated independently.

Yang et al. proposed to use “self-taught” image features, not “hand-designed” features such as SIFT and RGB to address the instance search task [64]. Their methods are based on the independent component analysis, and their effectiveness was confirmed using the TRECVID2011 instance search task dataset. Zhu et al. investigated the method of aggregating multiple query images for the instance search task and suggested that selecting the average pooling method was the best in terms of both accuracy and calculation cost among the five multi-image aggregation methods compared such as the maximum pooling and average similarity score [65]. Apostolidis et al. proposed a method for fast and accurate object detection in video and to use the GPU-based processing for the object detection part, new structure-based keyframe sampling technique, and SURF descriptors [66] robustified to scale variations [67]. Araujo et al. addressed video search using image query and studied solutions to reduce storage requirements of the video database [68]. They proposed a compression algorithm and reported that the search quality was also improved by the storage reduction strategy

C. Comparison with the Other Video Search Tasks

Zhu et al. compared the content-based image retrieval, duplicate detection, and instance search tasks and classified them according to the query type, searching criteria, and difficulty [38]. They also proposed a large-scale BOW framework to address the instance search task and argued that its performance was mainly due to similar scene retrieval and that the use of ROI information was the key factor to further enhance instance search accuracy. Zhu et al. analyzed the two highly cited object retrieval datasets, Oxford dataset and TRECVID instance search dataset, and classified them into either specific object search or similar image search. They also found that in a qualified object retrieval dataset, the labeled ROIs in queries should be much more important than the background information in the original image query [69].

Kaavya et al. reviewed recent work and the challenges with the multimedia indexing and retrieval tasks including the instance search task [70]. In their paper, my method based on the EIDF was mentioned and the essential points were summarized. Iwamura et al. proposed a method to make my BM25-based instance search method faster in terms of the computational burden [52]. They mentioned that BM25 video ranking score was obtained by solving the bichromatic reverse nearest neighbor search problem [51] and proposed an approximate method using the ANN search method using bucket distance hashing [71]. The experiment using the TRECVID2012 instance search task dataset showed the reduced calculation cost of my BM25-based method while slightly sacrificing search accuracy.

7. Conclusion

I first conducted an analysis of the discriminative power in the BM25 formula. My analysis was based on the Bayesian viewpoint, and I derived the BEIDF, which is the Bayesian counterpart of the EIDF that I previously proposed [1]. The form of the BEIDF is quite similar to the EIDF but retains some theoretical advantages over the EIDF as summarized in Section 3-D. I thus revealed that the EIDF can be derived by setting the informative prior distribution for the Bayesian calculation of the discriminative power. To address the instance search task, I also incorporated the ROI information into the BM25 video ranking model using the BEIDF. The effectiveness of the proposed methods was verified through a series of experiments using the TRECVID instance search task dataset.

For the application and extension of other state-of-the-art probabilistic IR models to content-based image/video retrieval tasks, LM, DFR, and DFI are strong candidates. However, as I mentioned in this and the previous papers, simply applying the text/document IR methods to the image/video retrieval task does not necessarily provide satisfactory results, primarily due to the essential difference in the text and image feature distributions. However, I expect that the lessons learnt from the proposed methodology are also helpful when applying LM, DFR and DFI models to content-based image/video retrieval tasks, contributing to bridging the gap in different information modals.

Chapter 4. IR Model using GPD and Its Application to Instance Search

1. Overview

In this chapter, I adopt the GPD [72] for the IM [11] and show that the parameters can be estimated based on the mean excess function applied to the data to be searched. The proposed IR model corresponds to the extension of the DFI [14] and is designed to be data-driven. The proposed model is then applied to the specific object search called the instance search and the effectiveness is experimentally confirmed.

2. Background

Effective IR models have been actively studied roughly since 1960s in the IR community [4]. BM25 [6], LM [7], and DFR [8] are state-of-the-art. The axiomatic approach is also proposed [9]. Recently, IM [11] and PM [12] are proposed. These models are simple compared to the DFR, since only one distribution is necessary (two distributions are required for the DFR). The problem then becomes the adequate setting of the distribution.

There are interesting relations such that adopting the LLD for IM and PM yields LMJM and sub-linear normalized term frequency (NTF) term in the BM25, respectively [10] [12]. Although these facts somehow support the effectiveness of IM and PM, the essential problem is whether the data to be searched follow the LLD or not. Setting the suitable distribution to the objective data is important for the successful retrieval.

Recently, the distribution is estimated using the EVS [13]. The distribution that the maximum NTF follows according to the EVS is adopted in the PM. To my knowledge, it is the first successful application of the EVS to the development of the IR model. In this paper, I propose the distribution estimation for the IM according to the knowledge of the EVS and this is the main contribution of this chapter.

I also show that the proposed model corresponds to the extension of the DFI [14] which is a parameter-free IR model. Since the parameters of the proposed model are estimated according to the data, the retrieval accuracy is expected to be improved. I demonstrate it for the specific object search called the instance search [1] using image-query video retrieval dataset.

The rest of this chapter is organized as follows. The Existing IR Models are explained in Section 3. I next describe the proposed model and explain the relation to the DFI model in Section 4. I also clarify the difference from the method in [13]. Section 5 shows the experimental justification using the instance search dataset and the detailed discussion is provided. Finally, Section 6 concludes this chapter.

3. Existing IR Models

The IR research field has progressed by focusing on actual document (text) retrieval tasks. However, retrieval models have been studied mathematically and their application is not necessarily restricted to the document retrieval task. Digital data that can be searched are within the scope of the formal IR models, and these models are now being applied to image, video, and even music and speech retrieval tasks [1] [73].

Various research topics have been addressed in the IR field [4] and efforts have been directed towards theoretical research on retrieval models [74] [75] [76]. The retrieval models are intended to mathematically model document retrieval and ranking processes based on an input query. Their realistic counterparts are, for example, Web search engines. Web search systems accept a keyword query and return the search results in a ranked list of Web pages uploaded to the World Wide Web. Since such search systems are now essential in our daily lives, research pursuing more natural and effective IR models is becoming much more important.

If IR models that are heuristic (not theoretically rigorous) are included, the number of models already proposed will become enormous. However, if I only focus on practical models whose theoretical basis and derivation are sufficiently mathematically sound, the number will become much smaller. I can consider the following six representative IR models, which are listed in chronological order.

1. BM25 [4] [6]
2. LM [4] [7]
3. DFR [4] [8]
4. IM [10] [11]
5. PM [12] [77]
6. DFI [14]

The BM25 and LM were extensively studied and experimentally verified by IR researchers in the 1990s. They are now the most-often used standard IR models. The DFR, IM, and DFI, proposed in 2002, 2008, 2012 and 2014, respectively, are rather new. These models were developed to ameliorate or generalize the former IR models. It is worth mentioning that these six representative IR models (BM25, LM, DFR, IM, PM, and DFI) are all probabilistic.

As explained in the next section, the BM25, LM, DFR, IM, PM, and DFI models are derived from different theoretical bases. However, the resulting models are all reduced to the means to quantify the importance degrees (weights) of words (terms) within documents. Therefore, improving the formal IR models corresponds to pursuing more natural and effective formulas that provide the weights of within-document terms. This fact is in accordance with the findings by Salton et al. in the late 1980s [78].

In this section, I summarize the BM25, LM, DFR, IM, PM, and DFI. In short, the BM25 evaluates the relevance probability of a document to a query. The LM quantifies the closeness between a document and query based on the divergence between their word occurrence probability distributions. The DFR and IM evaluate the weight of TF by measuring its divergence from basic randomness models. The PM is designed to model the TF weight and the ranking formula is defined by the multiplication with the IDF. The DFI model also measures the divergence by comparing the TF with the expected TF calculated under the independence occurrence assumption.

A. BM25

In the BM25, the following conditional probability is estimated.

$$P(\mathrm{Rel} = \mathrm{rel} | q', d) \tag{38}$$

Here, q' and d represent query and document, and they are vectors whose elements are within-query and within-document TFs. If I assume a total of M unique terms in the document collection and that index numbers are assigned to each word, q and d are *M*-dimensional vectors whose *i*th element is the within-query and within-document *i*th TF. The term Rel is the binary event taking either rel (relevance) or irrel (irrelevance). Equation (38) denotes the relevance probability between a query and document, and the BM25 model is designed to rank documents in decreasing order of their relevance probabilities to the query.

The ranking equation for the BM25 model is as follows:

$$P(\mathrm{Rel} = \mathrm{rel}|q', d) \propto_q \sum_{d_i \neq 0} \frac{q_i' d_i}{d_i + \kappa'} dp_i \qquad (39)$$

Here, dp_i expresses the discriminative power of the *i*th term. One of the problems in the BM25 is that the parameters should be well specified, especially for the actual calculation of dp_i. The dp_i is often approximated by

$$dp_i = \log\left(\frac{a'(N' - n_i' + b')}{(n_i' + a')b'}\right) \qquad (40)$$

Here, N' is the total number of documents in the database to be searched and n_i' is the number of documents that contain the *i*th word. The a' and b' highly affect the overall retrieval accuracy, and they should be well specified to obtain satisfactory results. $a' = b^{'} = 0.5\ or\ 1$ are often selected. As for the κ', $\kappa' = 2$ is often used from the experimental perspective.

B. LM

In this model, the following Kullback-Leibler divergence is analyzed:

$$\sum_{i \in q'} M_{q'}(i) \log \left(\frac{M_{q'}(i)}{M_d(i)} \right) \tag{41}$$

Here, $\sum_{i \in q'}$ means the summation over the query terms, $M_{q'}(i)$ is the query language model, and $M_{q'}(i) = \frac{q_i'}{ql}$. Here, ql is $\sum_{i \in q'} q_i'$ and represents the query length. In the same way, $M_d(i)$ is the document language model and $M_d(i) = \frac{d_i}{dl}$, where $dl = \sum_{i \in q} d_i$. The small value of Eq. (41) indicates that the language models for a query and document are close; thus, it is possible to rank documents in the increasing order of Eq. (41). This is the theoretical basis of the LM model. Equation (41) is equivalent to the following equation;

$$\sum_{i \in q'} \left(M_{q'}(i) \log M_{q'}(i) - M_{q'}(i) \log M_d(i) \right) \tag{42}$$

Therefore, I can see that Eq. (41) becomes small as the second term in Eq. (42) becomes large. Consequently, instead of Eq. (41), I can consider ranking documents in the decreasing order of $\sum_{i \in q'} M_{q'}(i) \log M_d(i)$.

It is further decomposed as follows:

$$\sum_{i \in q'} M_{q'}(i) \log M_d(i) = \sum_{i \in q' \cap d} M_{q'}(i) \log M_d(i) + \sum_{i \in q' \setminus d} M_{q'}(i) \log M_d(i) \tag{43}$$

Here, $\sum_{i\in q'\cap d}$ denotes the summation over the common terms in query and document and $\sum_{i\in q'\setminus d}$ denotes the summation over the terms that are only present in the query. To circumvent $\log(0)$ in Eq. (43), the following smoothing is performed for the document language model:

$$M_d(i) := M_d(i) + \alpha_d M_c(i) \tag{44}$$

Here, $M_c(i)$ is the collection language model, $M_c(i) = \frac{l_i}{cl}$ where l_i is the total frequency of the ith term in the document collection and $cl = \sum_i l_i$, and α_d depends on the smoothing method selected. Representative smoothing methods are $\alpha_d = \lambda'$ for the LMJM and $\alpha_d = \frac{\mu'}{(dl+\mu')}$ for the LMDS, where λ' $(0 < \lambda' < 1)$ and μ' (virtual document length added) are both parameters. Substituting Eq. (44) into Eq. (43) yields

$$\sum_{i\in q'} M_{q'}(i)\log M_d(i) \propto_{q'} \sum_{i\in q'\cap d} q_i' \log \frac{M_d(i)}{\alpha_d M_c(i)} + ql \cdot \log\alpha_d \tag{45}$$

Then, the ranking equation for the LMJM [4] becomes as follows:

$$\sum_{i\in q'} M_{q'}(i)\log M_d(i) \propto_{q'} \sum_{d_i\neq 0} q_i' \log\left(\frac{1-\lambda'}{\lambda'}\frac{\frac{d_i}{dl}}{\frac{l_i}{cl}} + 1\right) \tag{46}$$

The ranking equation for the LMDS [4] with parameter μ' is as follows:

$$\sum_{i \in q'} M_{q'}(i) \log M_d(i) \propto_{q'} \sum_{d_i \neq 0} q'_i \log\left(\frac{\frac{d_i}{\mu'}}{\frac{l_i}{cl}} + 1\right) - ql \cdot \log\left(1 + \frac{dl}{\mu'}\right) \quad (47)$$

Setting $\mu' = avdl$ makes Eq. (47) as follows:

$$\sum_{i \in q'} M_{q'}(i) \log M_d(i) \propto_{q'} \sum_{d_i \neq 0} q'_i \log\left(d_i \frac{N}{l_i} + 1\right) - ql \times \log\left(1 + \frac{dl}{avdl}\right) \quad (48)$$

The LMJM and LMDS become equivalent when all documents in the collection have the same length and by setting $\mu' = dl \frac{\lambda'}{1-\lambda'}$. In other words, the LMDS retains an effect of explicitly penalizing longer documents and this fact characterizes the LMDS compared with the LMJM.

C. DFR

The BM25 and LMJM can be summarized as follows:

$$\sum_{d_i \neq 0} q'_i g(d_i, d) \quad (49)$$

Here, $g(d_i, d)$ expresses the importance degree of d_i within the document d. As already mentioned in the previous sections, the BM25 and LMJM stand on different theoretical bases but are reduced to calculation methods for the importance degree of d_i. This fact implies that for the formal IR models, pursuing the natural form of $g(d_i, d)$ is essential, and the DFR model explained in this section further prompts this research direction.

In the general DFR model, $g(d_i, d)$ is defined as

$$g(d_i, d) = \left(1 - p_2(d_i|d_i > 0, d)\right)\left(-\log p_1(d_i|d)\right) \tag{50}$$

Here, $p_1(d_i|d)$ is called the basic randomness model, and it provides the probability of observing d_i in a document d; $p_2(d_i|d_i > 0, d)$ is called the normalization model, and it expresses the probability of the *i*th word showing d_i in an elite document to the *i*th word. The elite document is defined as a document in which the *i*th word appears at least once, that is, $d_i > 0$. Thus, Eq. (50) states that the term whose $p_1(d_i|d)$ and $p_2(d_i|d_i > 0, d)$ are small, that is, the term deviating from both the basic randomness and normalization models, gains a large importance weight. Deviating from each randomness model indicates that the word frequency is special, and, in the DFR framework, it is such a word for which a large importance weight is assigned.

The BM25-like model is derived within this framework. Setting $p_2(d_i|d_i > 0, d) = \frac{d_i}{d_i+\kappa}$ based on the Laplace's law of succession model and $p_1(d_i|d) = \left(\frac{n_i+0.5}{N+1}\right)^{d_i}$ yields

$$\sum_{d_i \neq 0} q_i \frac{d_i}{d_i + b} \log\left(\frac{N' + 1}{n_i' + 0.5}\right) \tag{51}$$

Although the $\log(\cdot)$ in Eq. (51) has a form similar to the dp_i in Eq. (40) with $a' = b' = 0.5\ or\ 1$, it never takes a negative value. Indeed, the $\log(\cdot)$ is the original IDF and the dp_i in Eq. (40) is the BM25 IDF.

D. IM

For the DFR, the difficulty lies in appropriately setting the two randomness models $p_1(d_i|d)$ and $p_2(d_i|d_i > 0, d)$. Naturally, setting these two probability models is not obvious, and a different setting yields a different DFR model. The reason for introducing the second randomness model is that since $p_1(d_i|d)$ is generally a decreasing function of d_i, $p_1(d_i|d)$ rapidly becomes small as d_i becomes large, and the information is readily diverged. To avoid this situation, the second normalization model $p_2(d_i|d_i > 0, d)$ is introduced by contrarily setting it as an increasing function of d_i. Therefore, $1 - p_2(d_i|d_i > 0, d)$ can be regarded as an adjustment factor that prevents $g(d_i, d)$ from becoming too large as d_i increases.

However, if I deliberately choose $p_1(d_i|d)$, which does not become too small as d_i increases, I can omit the second normalization model. Then the DFR model becomes

$$g(d_i, d) = -\log p_1(d_i|d) \tag{52}$$

and $p_1(d_i|d)$ is now called the information model. This model is called the IM. Setting it using the CDF of the LLD, to be more precise, by setting $p_1(d_i|d) = P_{LLD}(x \geq d_i|d)$, Eq. (52) becomes

$$g(d_i, d) = \log\left(\frac{\alpha'^{\beta'} + d_i^{\beta'}}{\alpha'^{\beta'}}\right) \tag{53}$$

Here, α' and β' are LLD parameters. Setting $\alpha' = \frac{l_i}{N}$, $\beta' = 1$ and replacing d_i with $\frac{c' \times d_i \times avdl}{dl}$, where c' is a parameter, makes the IM as follows:

$$\sum_{d_i \neq 0} q_i' \log\left(c' \frac{d_i}{dl} \frac{cl}{l_i} + 1\right) \tag{54}$$

This model is equivalent to the LMJM in Eq. (46) by regarding $c' = \frac{1-\lambda'}{\lambda'}$.

E. PM

The PM model is defined as follows:

$$g(d_i, d) = P(x \leq d_i | d) dp_i \tag{55}$$

The $P(x \leq d_i|d)$ is for modeling the TF term such as that for the BM25. Indeed, setting it using the LLD, $P(x \leq d_i|d) \coloneqq \frac{d_i^{\beta'}}{\alpha'^{\beta'} + d_i^{\beta'}}$, makes Eq. (55) as follows:

$$g(d_i, d) = \frac{d_i^{\beta'}}{\alpha'^{\beta'} + d_i^{\beta'}} dp_i \tag{56}$$

Further setting $\alpha' = \kappa', \beta' = 1$ and replacing d_i with the normalized one make the PM equivalent to the BM25 model (Eqs. (39) and (40)). Another setting of the CDF results in the different PM model and recently, it was determined by the knowledge of the EVS [13].

Paik set the CDF by the weighted combination of the two EVS distributions, Gumbel and Frechet distributions, that are both asymptotic distributions that the maximum NTF follow under the EVS basic assumptions. Then, Paik focuses on two different NTFs, relative intra-document frequency [79] and length normalized frequency, and defined the PM. Note that this approach was not designed to model the CDFs for these NTFs and the implicit assumption is that these NTFs follow the asymptotic distributions for the maximum NTF in documents. This assumption makes the weights for the NTFs larger when they approach the right most tail of the asymptotic distributions and in that case, the weighting method is reasonable. However, replacing the CDF for the NTF with the asymptotic distribution for the maximum NTF is not theoretically consistent, although the keyword-query document retrieval experiments showed significant improvements in the retrieval accuracy over the baseline IR models.

F. DFI

The DFI model is similar to the DFR and IM in terms of measuring the divergence of TF from a standard basic randomness model. However, for the DFI, such a standard model is constructed under the so-called independence assumption. Suppose $P(i,j)$, which is the joint probability of the *i*th term occurring in the *j*th document, and define the following value:

$$\frac{P(i,j)}{P(i)P(j)} = \frac{P(i|j)P(j)}{P(i)P(j)} = \frac{P(i|j)}{P(i)} \tag{57}$$

Here, $P(i)$ and $P(j)$ denote the probabilities of the *i*th term and *j*th document's occurring, respectively. The term $P(i|j)$ denotes the occurrence probability of the *i*th term given the *j*th document (within the *j*th document). Then, under the assumption that events *i* and *j* are independent, Eq. (57) reduces to 1 since $P(i|j) = P(i)$ in that case. However, if such an independence assumption does not hold, Eq. (57) takes a value other than 1. Therefore, the DFI model quantifies the word importance by measuring the deviation of the occurrence probability of the term from the standard model based on the independence assumption.

The $P(i|j)$ and $P(i)$ in Eq. (57) are estimated by $\frac{d_i}{dl}$ and $\frac{l_i}{cl}$; thus, Eq. (57) becomes

$$\frac{P(i|j)}{P(i)} = \frac{\frac{d_i}{dl}}{\frac{l_i}{cl}} = \frac{d_i}{e'_i} = w_i^{\mathrm{DFI}} \tag{58}$$

Here, $e'_i = \frac{l_i dl}{cl}$, which denotes the expected TF of the *i*th word. Equation (58) measures the deviation of d_i from its expected number e'_i based on the aforementioned independence assumption, and the higher the w_i^{DFI} becomes, the more the word is regarded as important.

Since important terms should be those whose TFs are higher than the expected values, the following value is also used as a measure for divergence from independence:

$$w_i^{\mathrm{DFI}} = \frac{d_i}{e'_i} - 1 = \frac{(d_i - e'_i)}{e'_i} \tag{59}$$

Note that for this model, when the right side of Eq. (59) becomes negative, w_i^{DFI} is replaced with zero.

According to Eq. (58) or (59), the following $g(d_i, d)$ is defined in the DFI model:

$$g(d_i, d) = w_i^{\mathrm{DFI}} \tag{60}$$

The following function is also often used:

$$g(d_i, d) = \log\left(w_i^{\mathrm{DFI}} + 1\right) \tag{61}$$

Since there are no parameters in Eqs. (58) to (61), the DFI model is non-parametric (parameter-free) compared with the other IR models explained before. In addition, since the effect of dl is taken into account in the calculation of e'_i, the document length normalization is naturally incorporated. The theoretical basis regarding the aforementioned independence assumption is quite natural. The experimental results also indicate the competitive, or even better, search accuracy against the current representative IR models mentioned in this section [80] [81] [82].

In this section, I summarized the six classical and current representative IR models (BM25, LM, DFR, IM, PM, and DFI). I saw that these IR models were derived from different theoretically sound bases and that all were reduced to the calculation of the importance weight for the TF.

4. EVS AND ITS APPLICATION TO IM

A. Brief Introduction of EVS

The EVS provides us what distributions maximum block data (MBD) and threshold excess data (TED) asymptotically follow, respectively [72]. In this chapter, I focus on the TED, which is the data larger than a pre-specified threshold. I first review some theoretical results for the TED in the EVS [72].

Suppose a random variable x whose CDF is F. Then, $F_u(y) = P(x - u < y|x > u)$ is considered. Here, u is the threshold. Then, under the EVS basic assumption, as u approaches the upper limit of F, $F_u(y) \approx H_\phi\left(\frac{y}{\sigma}\right)$. Here, $H_\phi\left(\frac{y}{\sigma}\right)$ is the GPD as follows:

$$H_\phi\left(\frac{y}{\sigma}\right) = 1 - \left(1 + \frac{\phi y}{\sigma}\right)_+^{\left(-1/\phi\right)} = \mathrm{GPD}(\sigma, \phi) \tag{62}$$

Here, $\sigma > 0$ and $-\infty < \phi < \infty$. Here, $(\zeta)_+ = \max\{\zeta, 0\}$.

When $\phi < 0$, $\mathrm{GPD}(\sigma, \phi)$ becomes the Beta distribution with the upper limit $0 \leq y \leq -\sigma/\phi$. When $\phi = 0$, $\lim_{\phi \to 0} \mathrm{GPD}(\sigma, \phi) = 1 - e^{-y/\sigma}$, that is, $\mathrm{GPD}(\sigma, \phi)$ becomes the exponential distribution and $0 \leq y \leq \infty$. For $\phi > 0$, $\mathrm{GPD}(\sigma, \phi)$ is the Pareto distribution and $0 \leq y \leq \infty$. As for F, as u approaches the upper limit of F and when F belongs to the domain of attraction for the Weibull distribution $\left(G_\phi, \phi < 0\right)$, $F_u(y)$ can be approximated by the Beta distribution. When F belongs to the domain of attraction for the Gumbel distribution (G_0), $F_u(y)$ is approximated by the exponential distribution. On the other hands, when F belongs to the domain of attraction for the Frechet distribution $\left(G_\phi, \phi > 0\right)$, $F_u(y)$ is approximated by the fat tail Pareto distribution. Here, G_ϕ is the following generalized extreme value distribution (GEVS):

$$G_\phi(x) = \exp\left[-(1+\phi x)_+^{-1/\phi}\right], -\infty < \phi < \infty \tag{63}$$

$$G_0(x) = \lim_{\phi \to 0} G_\phi(x) = \exp(-e^{-x}) \tag{64}$$

The GEVD is the asymptotic distribution for the MBD as the block size becomes sufficiently large. Therefore, the EVS basic assumption can be re-expressed for the following three cases:

(1) F belongs to the domain of attraction for the Weibull distribution $\left(G_\phi, \phi < 0\right)$,

(2) F belongs to the domain of attraction for the Gumbel distribution (G_0),

(3) F belongs to the domain of attraction for the Frechet distribution $\left(G_\phi, \phi > 0\right)$.

Under these EVS basic assumptions, as explained, it is mathematically shown that as the threshold increases, the TED asymptotically follows the GPD(σ, ϕ) and (ϕ, σ) are the GPD parameters. Note that the GPD is heavy-tailed for $\phi > 0$. In other words, under the EVS basic assumptions, $F_u(y)$, that is, the right tail of F larger than u can be approximated by the GPD as u becomes sufficiently large.

The ϕ is called the extreme value index (EVI) for F. When $\phi < 0$, F has a finite upper limit. When $\phi = 0$, F does not have a finite upper limit and there is a risk that the large realizations of x happen. When $\phi > 0$, F does not have a finite upper limit and there is a huge risk that the large realizations of x happen. Note that besides the domain of attraction mentioned above, in the EVS for TED, I can not state the shape of F lower than the threshold that is taken as sufficiently large. To the contrary, the EVS for MBD, the asymptotic distribution that the normalized data follows is stated as the block size becomes sufficiently large, although the choice of the normalization parameters is difficult. Then, the entire shape of the distribution is determined, although the distribution that the original data follow is difficult to be determined.

The mean excess function (MEF) is defined as follows:

$$\mathrm{e}_u(v) = \mathrm{E}[y - v | y > v] \tag{63}$$

When $y \sim \mathrm{GPD}(\sigma, \phi)$, $y - v | y > v \sim \mathrm{GPD}(\sigma + \phi v, \phi)$. Then, the MEF for the GPD exists for $\phi < 1$ and becomes

$$\mathrm{e}_u(v) = \frac{\sigma}{1 - \phi} + \frac{\phi}{1 - \phi} v \tag{64}$$

It is proved that the MEF is linear in v only for the GPD and I exploit this nature for designing a IR model in the next section.

B. IM using GPD

IM is defined as follows:

$$\mathrm{score}(q', d) = \sum_{i, q_i' = 1, d_i > 0} -\log P(X \geq d_i' | d) \tag{65}$$

Here, (q', d) are query and document vectors, and d_i' is the NTF for the ith term. $\mathrm{P}(X \geq d_i' | d)$ is called the information model and it should be heavy-tailed to prevent the document ranking score from becoming diverged.

I then replace the information model with that for the GPD. Using Eq. (62), equation (65) becomes

$$\begin{aligned}\mathrm{score}(q',d) &= \sum_{i,q_i'=1,d_i>0} \frac{1}{\phi}\log\left(1+\frac{\phi y}{\sigma}\right)\\ &\propto_q \sum_{i,q_i'=1,d_i>0} \log\left(1+\frac{\phi\left({}^{d_i}\!/_{e_i'}-\mu\right)_+}{\sigma}\right) \qquad (66)\end{aligned}$$

Here, d_i is the within-document term frequency for the ith term. I set $y = d_i' = {}^{d_i}\!/_{e_i'} - \mu$. $\left({}^{d_i}\!/_{e_i'} - \mu\right)_+$ is not zero only for ${}^{d_i}\!/_{e_i'} - \mu > 0$ (it is zero when ${}^{d_i}\!/_{e_i'} - \mu \leq 0$). Note that $\phi > 0$ is assumed since the information model in Eq. (65) should be heavy-tailed.

For the execution of Eq. (66), (ϕ, σ, μ) should be specified according to the data to be searched. For a certain (fixed) μ, I estimate the MEF as follows:

$$\hat{\mathrm{e}}_\mu(v) = \frac{1}{\#\{j|d_j' > v\}}\sum_j \left(d_j' - v\right)_+ \qquad (67)$$

Here, $\#\{j|d_j' > v\}$ is the number of terms whose NTFs are larger than a threshold v. Then, when the estimated MEF seems linear in v, applying the least squares method to estimate the slope and intercept provides the estimation results for (ϕ, σ) according to Eq. (64). Note that Eq. (64) only holds for $\phi < 1$ and therefore, $0 < \phi < 1$ is implicitly assumed for this parameter estimation method. I set the TED by ${}^{d_i}\!/_{e_i'} - \mu$ and therefore, the assumption of $0 < \phi < 1$ indicates that the CDF of ${}^{d_i}\!/_{e_i'}$, that is, F, does not have a finite upper limit and there is a huge risk that the large realizations happen. It also simultaneously assumes that F belongs to the domain of attraction for the Frechet distribution $\left(G_\phi, \phi > 0\right)$. These are the assumptions for my proposed model.

Although I excluded the case of $\phi = 0$, if the assumption for the EVI is changed to $0 \leq \phi < 1$, the F belongs to either Gumbel or Frechet distribution. Then the GPD also becomes the exponential distribution $\lim_{\phi \to 0} \mathrm{GPD}(\sigma, \phi) = 1 - e^{-y/\sigma}$. In that case, the IM becomes

$$\begin{aligned} \mathrm{score}(q', d) &= \sum_{i, q_i' = 1, d_i > 0} \frac{y}{\sigma} \\ &\propto_q \sum_{i, q_i' = 1, d_i > 0} \frac{d_i}{e_i'} - \mu \end{aligned} \tag{66}$$

When μ is set to 1, the model becomes equivalent to DFI in Eqs. (59) and (60). When $\phi < 0$, the TED has the upper limit specified by $-\sigma/\phi$ and this setting was excluded in the proposed model since such case is not applicable to my case. Generally speaking, $\phi > 0$, that is, the case that F does not have a finite upper limit and that there is a huge risk that the large realizations of x happen seems suitable to most of the IR cases. And, due to the limitation of using the MEF, the region of EVI can be restricted to $0 < \phi < 1$. Although this is one of my future words, investigating this assumption of the EVI for the other kinds of data is intriguing.

μ, that is, the threshold for the ${d_i}/{e_i'}$, is the control (tuning) parameter of the proposed model. When the estimated MEF for a certain μ is clearly deviated from a linear function, the aforementioned parameter estimation method is not theoretically applicable. I can vary μ until the estimated MEF becomes somehow linear, however, I cannot provide the sophisticated selection method of μ. Therefore, the μ is the parameter for the proposed model. To summarize, the proposed IR model is

$$\mathrm{score}(q', d) = \sum_{i, q_i' = 1, d_i > 0} \log\left(1 + \frac{\hat{\phi}\left({d_i}/{e_i'} - \mu\right)_+}{\hat{\sigma}}\right) \tag{68}$$

where, $(\hat{\phi},\hat{\sigma})$ are the estimated values based on the least squares results.

C. Relation to DFI

The DFI is expressed as follows:

$$\text{score}(q',d)=\sum_{i,q_i'=1,d_i>0}\log\left(1+\frac{d_i}{e_i'}\right) \tag{69}$$

Although the DFI is parameter-free, the function form is arbitrary. Indeed, the following form is also possible:

$$\text{score}(q',d)=\sum_{i,q_i'=1,d_i>0}\log\left(1+\frac{(d_i-e_i')_+}{e_i'}\right) \tag{70}$$

In this case, the terms whose TFs are larger than the expected values are only taken into account.

Comparing with the DFI with the model in Eq. (66), it is readily shown that $(\phi,\sigma,\mu)=(1,1,0)$ makes Eq. (66) identical to Eq. (69). Moreover, $(\phi,\sigma,\mu)=(1,1,1)$ in Eq. (66) leads to Eq. (70). On the other hands, Eq. (68) is based on $(\phi,\sigma,\mu)=(\hat{\phi},\hat{\sigma},\mu)$. Therefore, the proposed model can be regarded as the extension of the DFI in which the parameters are estimated according to the data to be searched. Since $\mu=0\ or\ 1$ for DFI, I can also think of setting $\mu=0\ or\ 1$ for Eq. (68). Then, the proposed model is data-driven and also becomes parameter-free.

The setting of $\phi=1$ for the GPD corresponds to using the LLD since the LLD is a special case of the GPD. Therefore, the DFI can be regarded as the IM using the LLD with $\sigma=1$.

When the σ is left as a parameter and defined as $\frac{\lambda'}{1-\lambda'}$, the resulting model becomes equivalent to the LMJM.

D. Difference from the Model in [13]

The IR model in [13] also uses the EVS. It is based on the following P model:

$$\mathrm{score}(q', d) = \sum_{i, q'_i = 1, d_i > 0} P(X \leq d'_i | d)\, \mathrm{IDF}_i \tag{71}$$

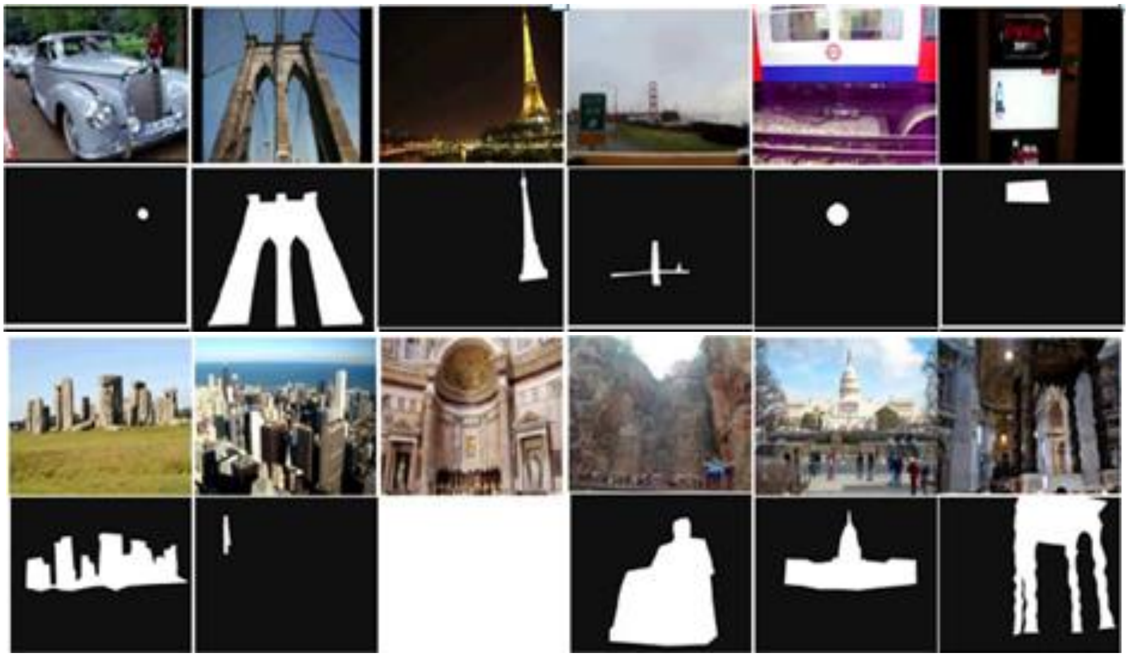
Then, $P(X \leq d'_i | d)$ is replaced with the CDF that a maximum NTF follows under the EVS basic assumption. IDF_i is an arbitrary inverse document frequency for the ith term. Note that the adopted distribution is not the distribution that the d'_i actually follows. Therefore, the implicit assumption is that the weight becomes large for the case that the d'_i approaches the right tail of the CDF that the maximum value follows.

To the contrary, the proposed model estimates the GPD that the d'_i with μ follows under the EVS basic assumption and does not rely on the assumption such as the aforementioned one. This is the primary difference from the method in [13], although the selection of the NTFs is also different. As already mentioned, according to the EVS and the EVS basic assumption, the d'_i asymptotically follow the GPD as μ increases. However, it is difficult to decide the value of μ for which the GPD assumption holds and the proposed GPD parameter estimation method based on the least squares is also for the determination of μ. That is, by varying μ, I manually plot and check the MEF estimated and utilize the least squares results to determine the proposed model in Eq. (68). When the plotted MLF is obviously deviated from a linear function, the proposed methodology is no longer applicable. For that case, the EVS basic assumption does not hold for the data under consideration and the knowledge of the EVS is not applicable.

5. EXPERIMENTS

A. Instance Search Dataset

I show the effectiveness of the proposed model for the instance search experiment. It is an image-query video retrieval task and the specific object is shown in the image-query. The system is required to search and rank videos in which the objects are shown in the decreasing order of relevance degrees. The following images are object examples.



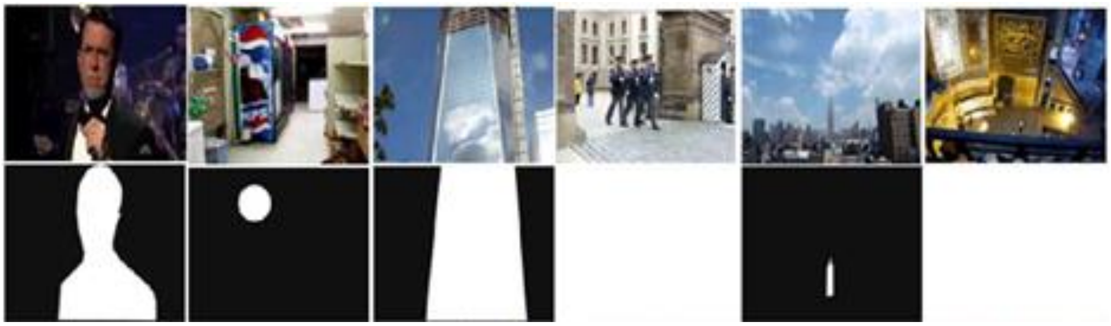


Figure 12: Examples of image-queries. Image queries are sets of original and region-of-interest (ROI) images. The white regions in the ROI images specify the objects in the original images. (Copyright©2017 ACM, [R3] Fig. 1)

The dataset is provided in the TRECVID2012 instance search task [1]. It is composed of 76751 short videos (the average duration is about 10 sec.) and 21 objects such as person/object/place. Each object is provided by five original and ROI images on average. The relevance judgement data is binary and the MAP is adopted as the search accuracy measure.

The frames are extracted from each video by 1 frame/sec and key-points are detected by the Harris-Laplace detector [21]. Then, the key-points are described by 128-dimensional SIFT feature vectors [20]. Same key-point detection and description methods are performed for all of the image-queries and the duplicated key-points are removed. This removal procedure is based on the cosine similarity value (CSV) between two SIFT vectors and the pair of key-points whose CSV is larger than a certain threshold is identified as matched. I varied the duplication threshold such as 0.999, 0.95 and 0.9. Then, for each key-point extracted from frames, the nearest key-point extracted from all of the image-queries is matched based on the CSV with a threshold of 0.9. Then, q and v already explained in Chapters 2 and 3 are obtained. The q' and d in the proposed model are replaced with q and v to retrieve videos in the database.

B. Results

Figures 13, 14, and 15 depict the estimated MEFs (μ=0) for the duplication thresholds 0.999, 0.95, and 0.9, respectively. The regions specified by the red circles seem linear and the proposed parameter estimation is performed only for these regions. Note that for $v \geq 200$ in Fig. 13, since $\#\{j|d'_j > v\}$ becomes small, the MEF becomes less trustworthy.

For all cases of the duplication thresholds 0.999, 0.95, and 0.9, I confirm that when $\#\{j|d'_j > v\}$ is sufficiently large, the corresponding MEF can be approximately regarded as linear. This fact supports the use of the GPD assumption and the proposed parameter estimation method is performed only for the linear region. I then execute Eq. (68) to rank videos. I also vary μ and the same parameter estimation procedure is performed for the other settings of μ.

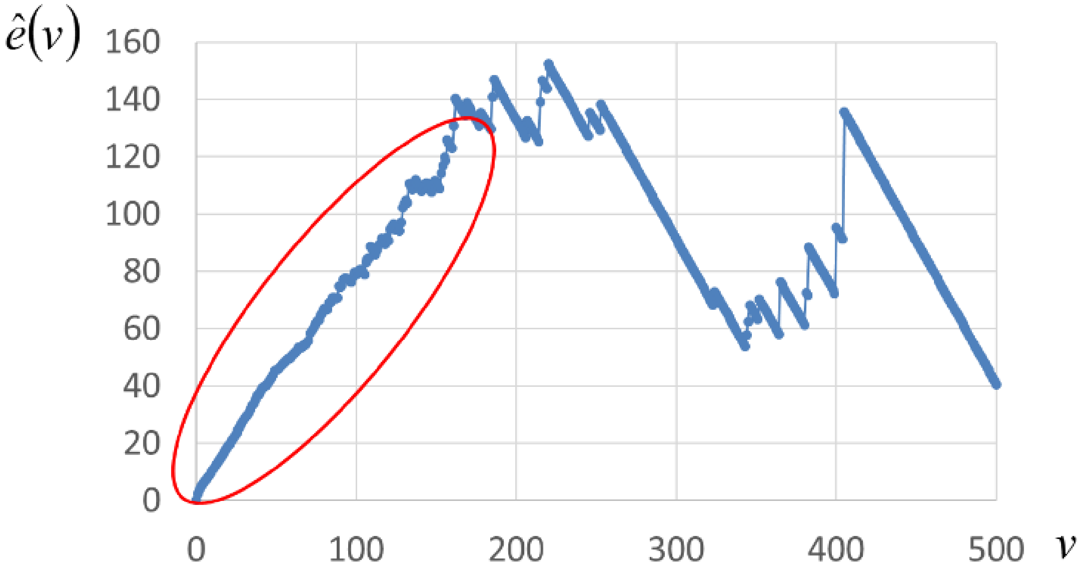


Figure 13: Estimated MEF for the duplication threshold 0.999. The region in the red circle seems linear. For $v \geq 200$, since $\#\{j|d'_j > v\}$ becomes small, the MEF becomes unstable and less trustworthy. (Copyright©2017 ACM, [R3] Fig. 2)

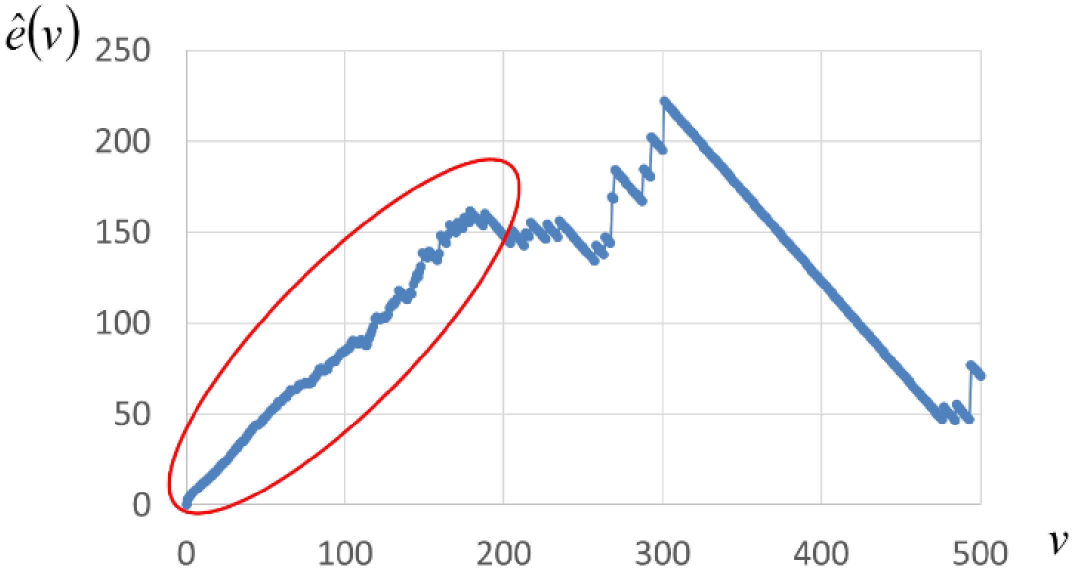


Figure 14: Estimated MEF for the duplication threshold 0.95. For $v \gg 200$, the MEF fluctuates since $\#\{j|d_j' > v\}$ is small. (Copyright©2017 ACM, [R3] Fig. 3)

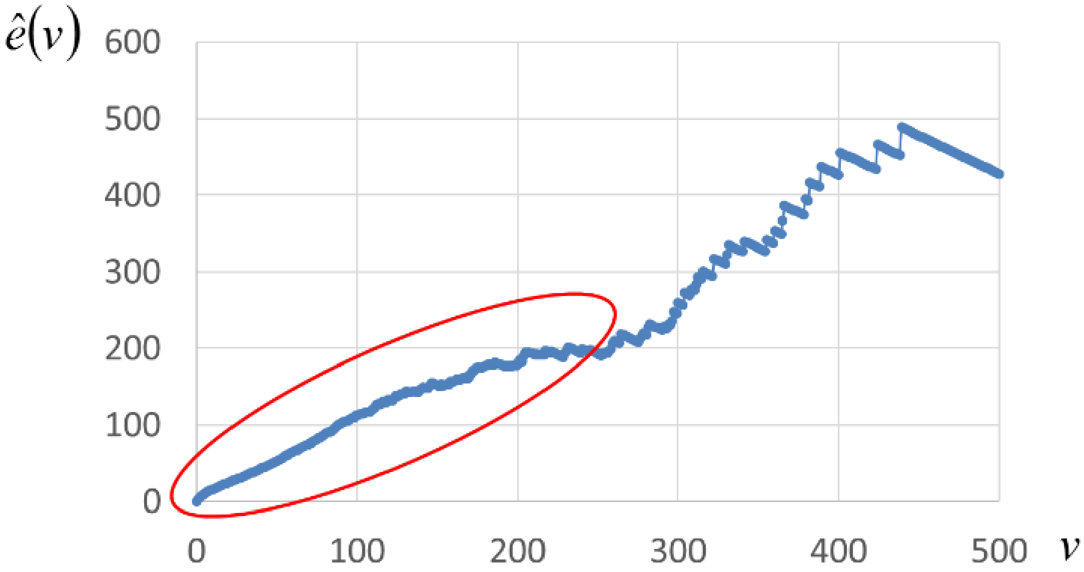


Figure 15: Estimated MEF for the duplication threshold 0.9. Compared with Figs. 13 and 14, the size of unstable region is decreased because $\#\{j|d_j' > v\}$ becomes sufficiently large. (Copyright©2017 ACM, [R3] Fig. 4)

Table 6 shows the MAP results for DFI and proposed models with the duplication threshold of 0.999. I only list the results for this setting since they are the best MAP results among ours. As shown in this table, the search accuracy of the proposed models is significantly better than those for the DFI models. As mentioned in Sections 4-C and 4-D, the proposed model is data-driven and I confirmed that using the estimated parameters from the data leads to the improvement in the search accuracy.

Table 6: MAP Results (Copyright©2017 ACM, [R3] Table 1)

IR model	MAP
DFI (Eq. (69))	0.23
DFI (Eq. (70))	0.24
Proposed model with μ=0	0.29
Proposed model with μ=1	0.29
Proposed model with μ=5	0.30
Proposed model with μ=10	0.31
Proposed model with μ=20	0.31
Proposed model with μ=50	0.31
Proposed model with μ=100	0.31

The MAP values scored by the proposed models are comparable with the highest MAP in the TRECVID2012 instance search task [1]. It is clearly shown that as μ increases, the MAP value is improved further. I discuss this issue in the next subsection.

C. Discussion

From Table 6, I confirm that the search accuracy is improved for large μ. As mentioned in Section 4-D, as μ increases, the TED, that is, $d_i' = {d_i}/{e_i'} - \mu$, asymptotically follows the

GPD when the EVS basic assumption holds for the NTF, that is, $d_i \big/ e_i'$. Roughly speaking, when the NTF follows a heavy-tailed distribution, such a proposition is true.

When large d_i often occurs in a video, the NTF is expected to follow a heavy-tailed distribution. Considering a video which is the time series of frame images, similar or even same key-points often repeatedly occur in the video, which contributes to large d_i. Such key-points are described as "bursty" in the image/video retrieval community and it is known that the adequate treatment is essential for the successful retrieval. Therefore, for the video retrieval task such as the instance search, the NTF can be expected to follow a heavy-tailed distribution. The larger μ supports the GPD assumption for the TED further and I expect that this tendency is shown in Table 6.

The results in Table 6 also indicate that the TED with large μ are sufficient for the successful video retrieval and that taking the whole data into account results in the deteriorated search accuracy. This is also the main finding of this chapter. It is interesting to see whether this proposition also holds for the document retrieval task. Since heavy-tailed distributions are often assumed for the NTF in the document retrieval community [8], there is also a possibility that such a proposition holds for the text retrieval task. Then, the inverted index may be dramatically shortened since only the effective TED ($d_i' > 0$) are sufficient for the successful retrieval. Further investigating this issue is my immediate future work.

6. Conclusions

I proposed a IR model based on the GPD in the IM framework. I also proposed the parameter estimation method based on the least squares method. The MEF is estimated from data and the manually selected linear region is processed by the least squares, which provides the estimates of the shape parameters for the GPD model. As shown in this chapter, the proposed IR model corresponds to the extension of the DFI. Since the model is data-driven, that is, the parameters are estimated according to the data to be searched, its retrieval accuracy is significantly improved. I confirmed it using the instance search. I also discussed the validation of the EVS basic assumption for the proposed IR model.

The immediate future work is the application of the proposed model to the document retrieval tasks. As mentioned in Section 5-C, when the same tendency in Table 6 is shown, it may be possible to dramatically decrease the retrieval cost since the effective TED are sufficient for the successful search. With the implicit assumption of the EVI for the proposed model in mind, investigating the validation of this assumption for the document retrieval dataset such as using the TREC dataset is highly desired. Similar experiments for the other kinds of dataset such as music and audio retrieval tasks are also interesting.

Chapter 5. Conclusions

1. Summary

The development of IR models was addressed in this dissertation and I proposed three different weighting methods. The first two methods were for achieving the state-of-the-art retrieval accuracy and the last one was also for pursuing the model simplicity. The first method is called the EIDF and it was designed to enhance the discriminative power of BM25 IDF in the BM25. The significantly enhanced retrieval accuracy was confirmed for the instance search experiments. The second contribution is for the theoretical analysis of the EIDF and the Bayesian weighting method to enhance the BM25 IDF was proposed. Using the prior information modeled by Beta distributions for retrieval features led to the EIDF-like weight called BEIDF whose effectiveness was also confirmed for the instance search experiments. I also showed that within the proposed framework, the BM25 IDF-like weight called the BIDF can be derived by setting the uniform prior (non-informative prior) distributions.

Finally, the latest IR model called the IM was addressed and I proposed a methodology to determine the information model based on the knowledge of the EVS and GPD. The parameter estimation method for the GPD was also proposed by analyzing the plotted MEF. I explained that LMJM and DFI are the special cases of the proposed model and showed that the proposed model becomes parameter-free (data-driven), that is, the model parameters are estimated and specified by data to be searched. The improved retrieval accuracy was also experimentally confirmed by comparing with those for the DFI. Theoretically, since the LLD often assumed as the basic distribution for IR models becomes the special case of the GPD, the existing models relying on the LLD assumption were interpreted from the GPD viewpoint.

2. Future Research

A. Relationship between GPD and Zipfian Distribution

The GPD is the extension of the LLD which has been often adopted as a fundamental distribution for developing IR models. Therefore, the GPD is expected to become another basic distribution for constructing a new family of IR models. In this dissertation, the video retrieval dataset was analyzed and the GPD was applied to model the weighting scheme in the IR models. Investigating the applicability of the GPD to the other kinds of dataset such as documents, audio and music is highly desired. However, the moment calculation for the GPD is not still obvious. A recent paper has reported the characteristic and moment generating functions to obtain these values [83] and I am certain that further research on this issue will deepen our knowledge of the GPD and will prompt the exploration of new IR models.

Furthermore, it should be noted that the GPD is mathematically related to Zipfian distribution [84] [85]. Therefore, observing the current and proposed IR models discussed in this dissertation from the Zipfian viewpoint will further enhance our knowledge. I am especially interested in pursuing a formal IR model purely based on the Zipf's law, which is a fundamental principle in word collection and it well explains the word frequency distribution in terms of the word rank among an entire word collection. Since the Zipf's law, that is, the fundamental principle may be adopted as an axiom to design a term weighting formula, the further research on this issue is interesting. The Zipfian distribution is also related to the Riemann zeta function and therefore, utilizing the mathematical properties of the zeta function to model the term weighting formulae is also intriguing.

B. Other Directions

I list the other interesting research directions as follows. Goyal et al. showed that the BM25, LM, and DFR are derived using their IR model based on the lexical association between words in a document [86]. Their model is related to the LM and the importance weights of words are re-distributed based on the lexical association with the context words. It is interesting to investigate whether the other representative IR models mentioned in this dissertation can be also derived from their model.

Wilkie et al. reported on the experimental comparison between the representative IR models, especially focusing on the retrieval bias problem for which particular documents with the same or similar characteristics are ranked higher. They also experimentally showed that the BM25 generally exhibited the least bias on the collection tested [87]. Further investigating this issue also for the models with EVS (PM+EVS in [13] and IM+EVS proposed in this dissertation) are desired.

Recently, the term weighting scheme proposed by Paik [88] showed outperforming retrieval accuracy over that in [79]. Therefore, following the methodology in [12], modeling this weighting scheme in the PM+EVS framework is interesting. The idea in [76] was also developed and the new exploration method for the space of IR term scoring functions was proposed [89]. Incorporating the IR models based on the EVS into this framework is also desired.

Word embedding techniques were also taken into account for the LM [90]. In this framework, words are expressed by low-dimensional vectors and the similarities between words are calculated by the distance between the vectors. Its natural extension is incorporating these techniques for the other IR models such as IM, PM, and DFI. Yang et al. released the evaluation platform for IR models to facilitate the performance comparison [91]. Using their platform, experimentally checking the effectiveness of the recent models discussed in this dissertation provides new insights and values.

Finally, the application of the deep learning techniques to the IR problems is now attracting attentions from IR researchers. Recent papers [92] [93] [94] are beginning to show positive results in the IR area and the comprehensive experimental comparisons with non-neural network models as explained in this dissertation are therefore necessary to further scrutinize the development of IR models.

Bibliography

[1] M. Murata, H. Nagano, M. Ryo, K. Kashino, S. Satoh, “BM25 with Exponential IDF for Instance Search”, IEEE Transactions on Multimedia, Vol.16, Issue 6, pp. 1690-1699, Oct. 2014.

[2] M. Murata, H. Nagano, K. Hiramatsu, K. Kashino, S. Satoh, “Bayesian Exponential Inverse Document Frequency and Region-of-Interest Effect for Enhancing Instance Search Accuracy”, IEICE Transactions on Information and Systems, Vol. E99-D, No. 9, pp. 2320-2331, Sep. 2016.

[3] M. Murata, K. Hiramatsu, S. Satoh, "Information Retrieval Model using Generalized Pareto Distribution and Its Application to Instance Search", In Proceeding of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), short paper, to appear, Aug. 2017.

[4] S. Buttcher, C. L. A. Clarke, G. V. Cormack, Information Retrieval, Implementing and Evaluating Search Engines, The MIT Press, London, 2010.

[5] S. E. Robertson, “The Probability Ranking Principle in IR”, Journal of Documentation, Vol. 33, Issue. 4, pp. 294–304, 1977.

[6] S. E. Robertson, S. Walker, “Some Simple Effective Approximations to the 2-poisson Model for Probabilistic Weighted Retrieval”, In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 232–241, July 1994.

[7] J. M. Ponte, W. B. Croft, “A Language Modeling Approach to Information Retrieval”, In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 275–281, Aug. 1998.

[8] G. Amati, C. J. V. Rijsbergen, “Probabilistic Models of Information Retrieval Based on Measuring the Divergence From Randomness”, ACM Transactions on Information Systems (TOIS), Vol. 20, Issue 4, pp. 357–389, Oct. 2002.

[9] H. Fang, T. Tao, C. X. Zhai, “A Formal Study of Information Retrieval Heuristics”, In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 49–56, July 2004.

[10] S. Clinchant, E. Gaussier, “Bridging Language Modeling and Divergence From Randomness Models: A Log-logistic Model for IR”, In Proceedings of the 2nd International Conference on the Theory of Information Retrieval (ICTIR), pp. 54-65, 2009.

[11] S. Clinchant, E. Gaussier, “Information-based Models for Ad Hoc IR”, In Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 234-241, July 2010.

[12] Y. Lv, C. X. Zhai, “A Log-Logistic Model-Based Interpretation of TF Normalization of BM25”, In Proceeding of the 34th European Conference on Information Retrieval (ECIR). pp. 244-255, 2012.

[13] J. H. Paik, “A Probabilistic Model for Information Retrieval Based on Maximum Value Distribution”, In Proceeding of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 585-594, Aug. 2015.

[14] I. Kocabas, B.T. Dincer, B. Karaoglan, “A Nonparametric Term Weighting Method for Information Retrieval Based on Measuring the Divergence From Independence”, Information Retrieval, Vol. 17, Issue 2, pp. 153-176, April, 2014.

[15] S. E. Robertson, K. S. Jones, “Relevance Weighting of Search Terms”, Journal of the American Society for Information Science, Vol. 27, Issue 3, pp. 129-146, May/June 1976.

[16] K. S. Jones, “A Statistical Interpretation of Term Specificity and Its Application in Retrieval”, Journal of Documentation, Vol. 28, Issue 1, pp. 11-21, 1972.

[17] S. E. Robertson, S. Walker, M. Hancock-Beaulieu, A. Gull, M. Lau, “Okapi at TREC4”, NIST SPECIAL PUBLICATION SP, pp. 73-96, 1996.

[18] S. E. Robertson, H. Zaragoza, “The Probabilistic Relevance Framework: BM25 and Beyond”, Foundations and Trends in Information Retrieval, Vol. 3, No. 4, pp. 333-389, Dec. 2009.

[19] S. E. Robertson, “On the History of Evaluation in IR”, Journal of Information Science, Vol. 34, Issue 4, pp. 439-456, June 2008.

[20] D. G. Lowe, “Distinctive Image Features from Scale-invariant Key Points”, International Journal of Computer Vision, Vol. 60, Issue 2, pp. 91-110, Nov. 2004.

[21] K. Mikolajczyk, C. Schmid, “Scale and Affine Invariant Interest Point Detectors”, International Journal of Computer Vision, Vol. 60, Issue 1, pp. 63-86, Oct. 2004.

[22] L. Page, S. Brin, R. Motwani, T. Winograd, “The Pagerank Citation Ranking: Bringing Order to the Web”, Stanford Univ., Stanford, CA, USA, Tech. Rep. Stanford Digital Libraries SIDL-WP-1999-0120, 1998.

[23] R. Yan, A. G. Hauptmann, "A Review of Text and Image Retrieval Approaches for Broadcast News Video", Information Retrieval, Vol. 10, Issue 4, pp. 445-484, Oct. 2007.

[24] J. Liu, W. Lai, X. S. Huang, Y. Huang, S. Li, "Video Search Re-ranking via Multi-graph Propagation", In Proceedings of the 15th ACM International Conference on Multimedia (MM), pp. 208-217, Sep. 2007.

[25] S. C. H. Hoi, M. R. Lyu, "A multimodal and multilevel ranking scheme for large-scale video retrieval", IEEE Transactions on Multimedia, Vol. 10, Issue 4, pp. 607-619, June 2008.

[26] Y. Liu, T. Mei, X. S. Hua, "CrowdReranking: Exploring Multiple Search Engines for Visual Search Reranking", In Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 500-507, July 2009.

[27] J. Jeon, V. Lavrenko, R. Manmatha, "Automatic Image Annotation and Retrieval using Cross-media Relevance Models", In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 119-126, Aug. 2003.

[28] V. Lavrenko, R. Manmatha, J. Jeon, "A Model for Learning the Semantics of Pictures", NIPS. Vol. 1, No. 2, 2003.

[29] D. M. Squire, W. Muller, H. Muller, T. Pun, "Content-based Query of Image Databases: Inspirations from Text Retrieval", Pattern Recognition Letters, Vol. 21, Issues 13-14, pp. 1193-1198, Dec. 2000.

[30] A. P. Vries, H. M. Blanken, "The Relationship Between IR and Multimedia databases", In Proceedings of the 20th Annual BCS-IRSG Colloq. IR, 1998.

[31] A. P. Vries, "Content and Multimedia Database Management Systems", Ph.D. thesis, Centre Telemat. Inf. Technol., Univ. Twente, Enschede, The Netherlands, 1999.

[32] A. P. Vries, T. Westerveld, "A Comparison of Continuous vs. Discrete Image Models for Probabilistic Image and Video Retrieval", In Proceedings of the 2004 International Conference on Image Processing (ICIP), pp. 2387-2390, Oct. 2004.

[33] J. Siciv, A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos", In Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV), pp. 1470-1477, Oct. 2003.

[34] D. Nister, H. Stewenius, "Scalable Recognition with a Vocabulary Tree", In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2161-2168, June 2006.

[35] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, “Object Retrieval with Large Vocabularies and Fast Spatial Matching”, In Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, June 2007.

[36] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, “Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases”, In Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, June 2008.

[37] W. Zhou, Y. Lu, H. Li, Y. Song, Q. Tian, “Spatial Coding for Large Scale Partial-duplicate Web Image Search”, In Proceedings of the 18th ACM International Conference on Multimedia (MM), pp. 511-520, Oct. 2010.

[38] C. Z. Zhu, S. Satoh, “Large Vocabulary Quantization for Searching Instances from Videos”, In Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (ICMR), Article No. 52, June 2012.

[39] D. D. Le, C. Z. Zhu, S. Poullot, V. Q. Lam, V. H. Nguyen, N. C. Duong, T. D. Ngo, D. A. Duong, S. Satoh, “National Institute of Informatics, Japan at TRECVID2012”, TRECVID Workshop, pp. 276-282, Nov. 2012.

[40] Z. Zhao, Y. Zhao, Y. Hua, W. Wang, D. Wan, G. Jia, Z. Li, F. Su, A. Cai, “BUPT-MCPRL at TRECVID2012”, TRECVID Workshop, pp. 66-73, Nov. 2012.

[41] Y. Peng, X. Zhai, J. Zhang, C. Yao, T. Xiao, N. Li, X. Luo, “PKUICST at TRECVID2012: Instance Search Task”, TRECVID Workshop, pp. 305-311, Nov. 2012.

[42] K. E. A. van de Sande, T. Gevers, C. G. M. Snoek, “Color descriptors for object category recognition”, In Proceedings of Conference on Colour in Graphics, Imaging, and Vision, Vol. 2008, No. 1, pp. 378-381, Jan. 2008.

[43] W. Zhang, C. C. Tan, S. A. Zhu, T. Yao, L. Pang, C. W. Ngo, “VIREO@TRECVID2012: Searching with Topology, Recounting with Small Concepts, Learning with Free Examples”, TRECVID Workshop, pp. 420-432, Nov. 2012.

[44] R. Aly, A. Doherty, D. Hiemstra, F. D. Jong, A. F. Smeaton, “The Uncertain Representation Ranking Framework for Concept-based Video Retrieval”, Information Retrieval, Vol. 16, Issue 5, pp. 557-583, Oct. 2013.

[45] S. E. Robertson, H. Zaragoza, M. Taylor, “Simple BM25 Extension to Multiple Weighted Fields”, In Proceedings of the 13th ACM International Conference on Information and Knowledge Management (CIKM), pp. 42-49, Nov. 2004.

[46] K. M. Svore and C. J. C. Burges, "A Machine Learning Approach for Improved BM25 Retrieval", In Proceedings of the 18th ACM International Conference on Information and Knowledge Management (CIKM), pp. 1811-1814, Nov. 2009.

[47] A. Singhal, "Modern information retrieval: A brief overview", IEEE Data Eng. Bull., Vol. 24, Issue 4, pp. 35-43, 2001.

[48] Y. Lv, C. X. Zhai, "When Documents Are Very Long, BM25 Fails!", In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 1103-1104, July 2011.

[49] R. Blanco, P. Boldi, "Extending BM25 with Multiple Query Operators", In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 921-930, Aug. 2012.

[50] J. Matas, O. Chum, M. Urban, T. Pajdla, "Robust Wide-baseline Stereo from Maximally Stable Extremal Regions", Image and Vision Computing, Vol. 22, Issue 10, pp. 761-767, Sep. 2004.

[51] F. Korn, S. Muthukrishnan, "Influence Sets based on Reverse Nearest Neighbor Queries", In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 201-212, May 2000.

[52] M. Iwamura, N. Matozaki, K. Kise, "Fast Instance Search based on Approximate Bichromatic Reverse Nearest Neighbor Search", In Proceedings of the 22nd ACM International Conference on Multimedia (MM), pp. 1121-1124, Nov. 2014.

[53] C. Z. Zhu, H. Jegou, S. Satoh, "Query-adaptive Asymmetrical Dissimilarities for Visual Object Retrieval", In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1705-1712, Dec. 2013.

[54] W. Zhang, C. W. Ngo, "Topological Spatial Verification for Instance Search", IEEE Transactions on Multimedia, Vol.17, Issue 8, pp. 1236-1247, Aug. 2015.

[55] P. Over, J. Fiscus, G. Sanders, D. Joy, M. Michel, G. Award, A. Smeaton, W. Kraaij, G. Quenot, "TRECVID2014 – An Overview of the Goals, Tasks, Data, Evaluation Mechanisms and Metrics", TRECVID Workshop, pp. 52, Nov. 2015.

[56] C. Wilkie, L. Azzopardi, "Best and Fairest: An Empirical Analysis of Retrieval System Bias", Advances in Information Retrieval, Vol. 8416, pp. 13-25, Apr. 2014.

[57] C. Z. Zhu, X. Zhou, S. Satoh, "Bag-of-words Against Nearest-neighbor Search for Visual Object Retrieval", In Proceedings of the 2nd IAPR Asian Conference on Pattern Recognition, pp. 626-630, Nov. 2013.

[58] W. Zhang, C. W. Ngo, “Searching Visual Instances with Topology Checking and Context Modeling”, In Proceedings of the 3rd ACM International Conference on Multimedia Retrieval (ICMR), pp. 57-64, Apr. 2013.

[59] X. Zhou, C. Z. Zhu, Q. Zhu, S. Satoh, Y. T. Guo, “A Practical Spatial Re-ranking Method for Instance Search from Videos”, In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), pp. 3008-3012, Oct. 2014.

[60] R. Tao, E. Gavves, C. G. M. Snoek, A. W. M. Smeulders, “Locality in Generic Instance Search from One Example”, In Proceedings of the 2014 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2099-2106, June 2014.

[61] F. Perronnin, C. Dance, “Fisher Kernels on Visual Vocabularies for Image Categorization”, In Proceedings of the 2007 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, June 2007.

[62] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perezm, C. Shmid, “Aggregating Local Image Descriptors into Compact Codes”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 34, Issue 9, pp. 1704-1716, July 2012.

[63] J. Meng, J. Yuan, Y. P. Tan, G. Wang, “Fast Object Instance Search in Videos from One Example”, In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), pp. 4381-4385, Sep. 2015.

[64] Y. Yang, S. Satoh, “Efficient Instance Search from Large Video Database via Sparse Filters in Subspaces”, In Proceedings of the 2013 IEEE International Conference on Image Processing (ICIP), pp. 3972-3976, Sep. 2013.

[65] C. Z. Zhu, Y. H. Huang, S. Satoh, “Multi-image Aggregation for Better Visual Object Retrieval”, In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4304-4308, May 2014.

[66] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, “Speeded-up Robust Features (SURF)”, Computer Vision and Image Understanding, Vol. 110, Issue 3, pp. 346-359, June 2008.

[67] E. Apostolidis, V. Mezaris, I. Kompatsiaris, “Fast Object Re-detection and Localization in Video for Spatio-temporal Fragment Creation”, In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops, pp. 1-6, July 2013.

[68] A. Araujo, M. Makar, V. Chandrasekhar, D. Chen, S. Tsai, H. Chen, R. Angst, B. Girod, “Efficient Video Search using Image Queries”, In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), pp. 3082-3086, Oct. 2014.

[69] C. Z. Zhu, S. Satoh, “Evaluation of Visual Object Retrieval Datasets”, In Proceedings of the 2013 IEEE International Conference on Image Processing (ICIP), pp. 3954-3958, Sep. 2013.

[70] S. Kaavya, G. G. LakshmiPriya, “Multimedia Indexing and Retrieval: Recent Research Work and Their Challenges”, In Proceedings of the 3rd International Conference on Signal Processing, Communication and Networking, pp. 15, Mar. 2015.

[71] M. Iwamura, T. Sato, K. Kise, “What is the Most Efficient Way to Select Nearest Neighbor Candidates for Fast Approximate Nearest Neighbor Search?”, In Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), pp. 3535-3542, Dec. 2013.

[72] R. Takahashi, T. Shimura, “Extreme Values Statistics”, Kindaikagakusha (Only available in Japanese), Aug. 2016.

[73] L. Su, C. C. M. Yeh, J. Y. Liu, J. C. Wang, Y. H. Yang, “A Systematic Evaluation of the Bag-of-frames Representation for Music Information Retrieval”, IEEE Transactions on Multimedia, Vol. 16, Issue 5, pp. 1188-1200, Aug. 2014.

[74] C. Zhai, J. Lafferty, “A Study of Smoothing Methods for Language Models Applied to Ad Hoc Information Retrieval”, In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 334-342, Sep. 2001.

[75] C. Zhai, J. Lafferty, “A Study of Smoothing Methods for Language Models Applied to Information Retrieval”, ACM Transactions on Information Systems, Vol. 22, Issue 2, pp. 179-214, Apr. 2004

[76] H. Fang, C. X. Zhai, “An Exploration of Axiomatic Approaches to Information Retrieval”, In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 480-487, Aug. 2005.

[77] S. Clinchant, E. Gaussier, “The BNB Distribution for Text Modeling”, In Proceedings of the 30th European Conference on Information Retrieval (ECIR), pp. 150-161, Mar. 2008.

[78] G. Salton, C. Buckley, “Term-weighting Approaches in Automatic Text Retrieval”, Information Processing and Management, Vol. 24, Issue 5, pp. 513-523, 1988.

[79] J. H. Paik, “A Novel TF-IDF Weighting Scheme For Effective Ranking”, In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 343-352, Aug. 2013.

[80] B. T. Dincer, I. Kocabas, B. Karaoglan, “IRRA at TREC2009: Index Term Weighting from Independence Model”, In Proceedings of the 18th Text Retrieval Conference (TREC), Nov. 2009.

[81] B. T. Dincer, I. Kocabas, B. Karaoglan, “IRRA at TREC2010; Index Term Weighting from Independence model”, In Proceedings of the 19th Text Retrieval Conference (TREC), Nov. 2010.

[82] B. T. Dincer, “IRRA at TREC2012; Divergence From Independence (DFI)”, In Proceedings of the 21st Text Retrieval Conference (TREC), Nov. 2012.

[83] G. Muraleedharan, C. G. Soares, “Characteristic and Moment Generating Functions of Generalized Pareto (GP3) and Weibull Distributions”, Journal of Scientific Research & Reports, Vol. 3, Issue 14, pp. 1861-1874, 2014.

[84] L. A. Adamic, “Zipf, Power-laws, and Pareto -A Ranking Tutorial”, Only Available on http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html, 2000.

[85] L. A. Adamic, B. A. Huberman, “Zipf’s Law and the Internet”, Glottometrics, Vol. 3, Issue 1, pp. 143-150, 2002.

[86] P. Goyal, L. Behera, T. M. McGinnity, “A Novel Neighborhood Based Document Smoothing Model for Information Retrieval”, Information Retrieval, Vo. 16, Issue 3, pp. 391-425, June 2013.

[87] C. Wilkie, L. Azzopardi, “Relating Retrievability, Performance and Length”, In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 937-940, July 2013.

[88] J. H. Paik, “Parameterized Decay Model for Information Retrieval”, ACM Transactions on Intelligent Systems and Technology (TIST), Vol. 7, Issue 3, Article No. 27, April 2016.

[89] P. Goswami, E. Gaussier, M. R. Amini, “Exploring the Space of Information Retrieval Term Scoring Functions”, Information Processing and Management, Vol. 53, Issue 2, pp. 454-472, March 2017.

[90] H. Zamani, W. B. Croft, “Embedding-based Query Language Models”, In Proceedings of the 2016 ACM on International Conference on the Theory of Information Retrieval (ICTIR), pp. 147-156, Sep. 2016.

[91] P. Yang, H. Fang, “A Reproducibility Study of Information Retrieval Models”, In Proceedings of the 2016 ACM on International Conference on the Theory of Information Retrieval (ICTIR), pp. 77-86, Sep. 2016.

[92] J. Guo, Y. Fan, Q. Ai, W. B. Croft, “A Deep Relevance Matching Model for Ad-hoc Retrieval”, In Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM), pp. 55-64, Oct. 2016.

[93] B. Mitra, N. Craswell, “Neural Models for Information Retrieval”, arXiv:1705.01509, 2017.

[94] N. Craswell, W. B. Croft, J. Guo, B. Mitra, M. Rijke, “Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval”, In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 1245-1246, July. 2016.

RELATED PUBLICATIONS

Journal Papers

[R1] M. Murata, H. Nagano, R. Mukai, K. Kashino, S. Sato, “BM25 with Exponential IDF for Instance Search”, IEEE Transaction on Multimedia, Vol. 16, Issue 6, pp. 1690-1699, Oct. 2014.
Chapter 1 in this dissertation is based on this paper [R1], Copyright©2014 IEEE. All the figures and tables in Chapter 1 are reused from this paper under the permission of the IEEE.

[R2] M. Murata, H. Nagano, K. Hiramatsu, K. Kashino, S. Satoh, "Bayesian Exponential Inverse Document Frequency and Region-of-Interest Effect for Enhancing Instance Search Accuracy", Transactions of the Institute of Electronics, Information and Communication Engineers (IEICE), Vol. E99-D, No. 9, pp. 2320-2331, Sep. 2016.
Chapter 2 in this dissertation is based on this paper [R2], Copyright©2016 IEICE. All the figures and tables in Chapter 2 are reused from this paper under the permission of the IEICE.

Reviewed Conference Papers

[R3] M. Murata, K. Hiramatsu, S. Satoh, "Information Retrieval Model using Generalized Pareto Distribution and Its Application to Instance Search", In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), short paper, to appear, Aug. 2017.
Chapter 3 in this dissertation is based on this paper [R3], Copyright©2017 ACM. All the figures and tables in Chapter 3 are reused from this paper under the permission of the ACM.

Non-Reviewed Articles and Conference Papers

村田眞哉, 永野秀尚, 平松薫, 川西隆仁, 柏野邦夫, "注目領域情報を使用したリランキングに基づく特定物体探索", 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, 114(521), pp. 245-249, 2015 年 3 月.

村田眞哉, 永野秀尚, 平松薫, 川西隆仁, 柏野邦夫, "映像中の特定物体を探索するインスタンスサーチ技術", 画像ラボ（Image Laboratory）, 日刊工業出版, pp. 28-34, 2015年6月.

M. Murata, H. Nagano, K. Hiramatsu, T. Kawanishi, K. Kashino, S. Satoh, "NTT Communication Science Laboratories at TRECVID2014 Instance Search Task", TRECVID2014 Workshop, Nov. 2014.

村田眞哉, 永野秀尚, 向井良, 平松薫, 柏野邦夫, "映像中の特定物体を探索するインスタンスサーチ技術", NTT技術ジャーナル, 26(9), pp. 16-19, 2014年9月.

村田眞哉, 永野秀尚, 柏野邦夫, 佐藤真一, "Exponential BM25によるインスタンスサーチ", 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, 113(402), pp. 189-194, 2014年1月.

M. Murata, H. Nagano, K. Kashino, S. Satoh, "NTT Communication Science Laboratories and National Institute of Informatics at TRECVID2013 Instance Search Task", TRECVID2013 Workshop, Nov. 2013.

村田眞哉, 永野秀尚, 向井良, 柏野邦夫, 佐藤真一, "画像をクエリとしたインスタンス映像検索", 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, 112(441), pp. 215-219, 2013年2月.

M. Murata, T. Izumitani, H. Nagano, R. Mukai, K. Kashino, S. Satoh, "NTT Communication Science Laboratories and National Institute of Informatics at TRECVID2012: Instance Search and Multimedia Event Detection Tasks", TRECVID2012 Workshop, Nov. 2012.

OTHER PUBLICATIONS

Journal Papers

M. Murata, K. Hiramatsu, "Non-Gaussian filters for Nonlinear Continuous-discrete models", SICE Journal of Control, Measurement, and System Integration (SICE JCMSI), Vol. 10, No. 2, pp. 53-61, Mar. 2017.

村田眞哉, 平松薫, "アンサンブルカルマンフィルタ，粒子フィルタ，ガウシアン粒子フィルタについて", システム制御情報学会論文誌 (ISCIE Transactions), Vol. 29, No. 10, Jan. 2016.

村田眞哉，高屋典子，市川裕介，内山匡，"EC サイトにおけるセールシミュレーション", 日本応用数理学会論文誌，Vol. 23, Issue 2, pp. 68-72, 2013 年 6 月.

村田眞哉, 戸田浩之, 松浦由美子,片岡良治, “クリックログ解析による情報要求ベースの検索結果ランキング”，日本データベース学会論文誌(DBSJ Journal), Vol. 7, No. 4, pp. 37-42, Mar. 2009.

M. Murata, T. Hiroyuki, Y. Matsuura, R. Kataoka, “Access Concentration Detection in Click Logs to Improve Mobile Web-IR”，Information Sciences, Vol. 179, Issue 12, pp. 1859-1869, May 2009.

村田眞哉, 戸田浩之, 松浦由美子, 片岡良治, “検索結果中のアクセス集中サイトを利用したクエリ拡張法の提案”，日本データベース学会論文誌(DBSJ Letters), Vol. 6, No. 4, pp. 45-48, Mar. 2008.

Reviewed Conference Papers

M. Murata, K. Hiramatsu, "Non-Gaussian Estimation of Nonlinear Continuous-discrete Models", In Proceedings of the 48th ISCIE International Symposium on Stochastic Systems Theory and Its Applications (SSS), to appear, 2017.

M. Murata, H. Nagano, K. Hiramatsu, K. Kashino, "Filter Design Based on Multiple Model Estimation", In Proceedings of the 2016 American Control Conference (ACC), pp. 7061-7066, July 2016.

M. Murata, H. Nagano, K. Kashino, "Gaussian Sum Resampling Filter", In Proceedings of the 54th IEEE Conference on Decision and Control (CDC), pp. 4338-4343, Dec. 2015.

M. Murata, H. Nagano, K. Kashino, "Monte Carlo Filter Particle Filter", In Proceedings of the 14th European Control Conference (ECC), pp. 2836-2841, July 2015.

M. Murata, H. Nagano, K. Kashino, “Iterative Unscented Statistically Linearized Filter for Nonlinear Gaussian Observation Models”, In Proceedings of the 53rd IEEE Conference on Decision and Control (CDC), pp. 4154-4159, Dec. 2014.

M. Murata, H. Nagano, K. Kashino, "Gaussian Unscented Filter", In Proceedings of the 46th ISCIE International Symposium on Stochastic Systems Theory and Its Applications (SSS'14), pp. 54-58, Oct. 2015.

M. Murata, H. Nagano, K. Kashino, “Suboptimal Kalman Filter for Dual Estimation under Dynamical Uncertainties”, In Proceedings of the 19th IFAC World Congress, Vol. 19, No. 1, pp. 5497-5502, Aug. 2014.

M. Murata, H. Nagano, K. Kashino, “Unscented Statistical Linearization and Robustified Kalman Filter for Nonlinear Systems with Parameter Uncertainties”, In Proceedings of the 2014 American Control Conference (ACC), pp. 5079-5084, June 2014.

M. Murata, H. Nagano, K. Kashino, “Robustifying Kalman Filter to Rapidly Adapt to Significant Changes in System Model Parameters of State-Space Models”, In Proceedings of the 52nd IEEE Conference on Decision and Control (CDC), pp. 7803-7808, Dec. 2013.

M. Murata, K. Kashino, “Normalized Unscented Kalman Filter and Normalized Unscented RTS Smoother for Nonlinear State-Space Model Identification”, In Proceedings of the 2013 American Control Conference (ACC), pp. 5462-5467, June 2013.

M. Murata，H. Toda，Y. Matsuura，R. Kataoka，T. Mochizuki, "Detecting Periodic Changes in Search Intentions in a Search Engine", In Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), pp. 1525-1528, Oct. 2010.

M. Murata， H. Toda， Y. Matsuura， R. Kataoka， "Query-page Intention Matching Using Clicked Titles and Snippets to Boost Search Rankings", In Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL), pp. 105-114, June 2009.

M. Murata, H. Toda, Y. Matsuura, R. Kataoka, "Improving Mobile Web-IR Using Access Concentration Sites in Search Results", In Proceedings of the 9th International Conference on Web Information Systems Engineering (WISE), pp. 221-234, Sep. 2008.

Non-Reviewed Journal Articles and Conference Papers

村田眞哉, 平松薫, "非線形連続-離散モデルの非ガウス推定", 第59回自動制御連合講演会, pp. 49-50, 2016年11月.

村田眞哉, 永野秀尚, 柏野邦夫, "ベンチマーク問題を使用した非線形フィルタの性能検証", システム制御情報学会研究発表講演会講演論文集 60, 4p, 2016年5月.

村田眞哉, 永野秀尚, 柏野邦夫, "モンテカルロ粒子フィルタとガウス和リサンプリングフィルタ", システム制御情報学会研究発表講演会講演論文集 59, 6p, 2015年5月.

村田眞哉, 永野秀尚, 柏野邦夫, "非線形ガウス型観測モデルに対する iterative unscented statistical linearization", システム制御情報学会研究発表講演会講演論文集 58, 4p, 2014年5月.

村田眞哉, 永野秀尚, 柏野邦夫, "構造変化を伴うダイナミクス下における準最適カルマンフィルタの設計", システム制御情報学会研究発表講演会講演論文集 57, 4p, 2013年5月.

村田眞哉, 戸田浩之, 松浦由美子, 片岡良治, "アクセス集中サイトを利用した検索精度の向上", NTT技術ジャーナル, Vol.21, No.1, pp. 58-61, 2009年1月.

村田眞哉, 戸田浩之, 松浦由美子, 片岡良治, "検索結果のアクセス分析に基づく情報要求ベースのランキング", 電子情報通信学会技術研究報告. DE, データ工学 108(329), 19, 2008年11月.