DEMIR at ImageCLEFMed 2013: The Effects of Modality Classification to Information Retrieval

Okan Ozturkmenoglu, Nefise Meltem Ceylan, and Adil Alpkocak

Dokuz Eylül University, Department of Computer Engineering, DEMIR Dokuz Eylül Multimedia Information Retrieval Research Group, Tinaztepe Izmir, 35160, Turkey
okan.ozturkmenoglu@deu.edu.tr, meltem.ceylan@gmail.com, alpkocak@cs.deu.edu.tr

Abstract. This paper present the details of participation of DEMIR (Dokuz Eylül University Multimedia Information Retrieval) research team to the ImageCLEF 2013 Medical Retrieval task. This year, we participated to two subtasks: modality classification and ad-hoc image-based retrieval. For them, our central method is integrated combination multimodal retrieval applied to retrieved documents sets of each visual and text features, after than our information retrieval based classification algorithm is performed. In modality classification subtask, we proposed an approach for modality classification based on information retrieval techniques. The main elements of this method are information retrieval techniques. Additionally, in ad-hoc image-based retrieval subtask, we assumed as a baseline that our methods which were obtained our best performances in ImageCLEF 2012. We added on our proposed classification method to these baseline runs and we evaluated impact of classification on the modalities of documents.

Keywords: Modality classification, multimodal information retrieval, content-based image retrieval, medical image retrieval, information retrieval

1   Introduction

In this paper we present the experiments performed by Dokuz Eylül University Multimedia Information Retrieval (DEMIR) Group, Turkey, in the context of our participation to the ImageCLEF 2013 AMIA: Medical task. In this year, we participated to two subtasks: modality classification and ad-hoc image-based retrieval. The main focus of this work is evaluation of result improvement using the classification methods on the modalities of documents in data collection. We performed these methods as a baseline to our best results in ImageCLEF 2012 Medical Image Classification and Retrieval (Vahid et al., 2012) and so we are expected to increase performance than last year. In this year, as a baseline, we assumed that we applied inter modality integrated combination of text and low-level image features; and we utilized this method to combine result of different low level features of images as well as last year. On the other hand, we tried to filter out of irrelevant documents using classification algorithm.

After explanation of modality classification and ad-hoc image-based retrieval tasks definition (Section 2), we describe the textual and visual features (Section 3) we used. Section 4 contains that our classification technique for filtering the data collection out and our studies on modality classification subtask. After we describe methods, submitted runs and results on ad-hoc image-based retrieval subtask (Section 5), and then Section 6 concludes the paper by pointing out the open issues and possible avenues of further research for content-based image retrieval.

2 Task Definition

2.1 Modality Classification

For this task, past studies showed that methods based on visual features gives better results than text-based techniques. These studies provide better results as filtering out search results using modality information, which can be extracted from the image itself using visual features, such as Goldminer or Yottalook. So, the search results can be improved significantly using the modality classification. We expected to get better results using this modality information. With this in mind, our current and previous studies (Alpkocak et al., 2011) and the studies of IBM group have shown that mixed methods perform slightly better results in an appreciable ratio. For example, in modality classification task, IBM group classified in %80.79 for textual modality but for mixed modality, the count of correctly classified documents increased to %81.68 at ImageCLEF 2013 AMIA: Medical. Within that perspective; firstly we apply our information retrieval based classifier to textual data. Then, we performed integrated combination multimodal retrieval technique and information retrieval based classifier technique as mixed method approach.

In this task, the data collection contains a modality hierarchy, which has three main modality categories of image and totally 31 categories, training set and test set.

2.2 Ad-hoc Image-based Retrieval

The data collection of ad-hoc image-based retrieval task in ImageCLEF 2013 AMIA: Medical Retrieval has textual and visual information. It has about 75K articles in collection with approximately 306K figures. Participants were given a set of 35 queries with 2-3 sample images for each query. The queries are classified into mixed, visual and semantic, based on the methods that are expected to yield the best results.

We performed our experiments using ImageCLEF 2012 Medical Image Classification and Retrieval track's data. We check the variation of retrieval methods on textual and visual information to gain the best result.

3 Feature Definition

3.1 Textual Features

We used Terrier IR Platform API (Ounis et al., 2006), which is an open source search engine written in Java and is developed at the School of Computing Science, University of Glasgow, to generate VSM. Terrier provides efficient and effective search methods supported by many different parameters.

In modality classification task, we used caption tags of each test and train image as textual features and modeled in vector space model (VSM). We can summarize applied parameters of Terrier under three topics. Firstly; we apply token normalization; remove all punctuation characters and convert all characters to lower-case. After then, stop words are removed from the collection, and we applied porter stemmer algorithm on normalized tokens. But we did not apply stemming for all runs. The parameters of different runs are given in detail in section 4.2. Beside these parameters, we apply different weighting schemes. According to defined parameters, two different text collections are obtained. In the first set, we used porter stemmer and InL2 weighting approach. For the second set, we did not use stemmer and did apply $TF \times IDF$ weighting scheme.

In ad-hoc image-based retrieval task, we first split the XML file for textual metadata and represented each image in the collection as a structured document of xml file. We also expanded the XML file using related article full text, abstract and title as new tags. We used Terrier for our text-based information retrieval subsystem and we performed our experiments on textual features using $TF \times IDF$. The order in which transformations were applied is as follows:

1. Noise character removal: characters with no meaning, like punctuation marks or blanks, are all eliminated;
2. Stop-word removal: discarding of semantically empty words, very high-frequency words;
3. Token normalization: converting all words to lower case;
4. Stemming: we used the Porter stemmer (Porter, 1997) as a process for removing the common morphological endings from words in English.

3.2 Visual Features

We extracted features for all images in test collection and query examples using Rummager tool (Chatzichristofis et al., 2009), which is developed in the Automatic Control Systems & Robotics Laboratory at the Democritus University of Thrace-Greece. We used FCTH, CLD and CEDD features. Our previous studies (Vahid et al., 2012) have shown that these three low-level features allowed us access to our best scores so we extracted features for all images. Here we explained these features we used below:

- Color and edge directivity descriptor (CEDD): The CEDD includes texture information produced by the six-bin histogram of the fuzzy system that uses the five digital filters proposed by the MPEG-7 EHD. Additionally, for color information the CEDD uses the 24-bin color histogram produced by the 24-bin fuzzy-linking system. Overall, the final histogram has 144 regions. This feature combines EHD with color histogram information and named "Color and Edge Directivity Descriptor". Important attribute of the CEDD is the low computational power needed for its extraction, in comparison to the needs of the most MPEG-7 descriptors (Chatzichristofis and Boutalis, 2008a).
- Fuzzy color and texture histogram (FCTH): The FCTH descriptor includes the texture information produced in the eight-bin histogram of the fuzzy system that uses the high frequency bands of the Haar wavelet transform. For color information, the descriptor uses the 24-bin color histogram produced by the 24-bin fuzzy-linking system. Overall, the final histogram includes192 regions. This feature fuzzy version of CEDD feature which contains fuzzy set of color and texture histogram and named "Fuzzy Color and Texture Histogram". This feature contains results from the combination of 3 fuzzy systems including histogram, color and texture information (Chatzichristofis and Boutalis, 2008b).
- Color layout descriptor (CLD): This descriptor effectively represents the spatial distribution of color of visual signals in a very compact form. This compactness allows visual signal matching functionality with high retrieval efficiency at very small computational costs. It provides image-to-image matching as well as ultra-high-speed sequence-to-sequence matching, which requires so many repetitions of similarity calculations (Lux and Chatzichristofis, 2008).

4   Modality Classification

4.1   Methods

Text. We applied a new approach for modality classification based on information retrieval System (IRS) where IRS system is used as text classifier.

Mixed. Mixed method for modality classification based on combining CEDD, FCTH and CLD features or visual terms with textual information. Algorithm of this method is similar to the algorithm given for text based classification. Additionally, it combines the retrieved documents of different visual features and text features.

4.2   Runs

In order to assess the above mentioned methods, we set up a set of experiments on the data collection of modality classification task in ImageCLEF 2013 AMIA: Medical Retrieval. We submitted 6 runs to ImageCLEF 2013 AMIA: Medical Retrieval for modality classification task two categories as following:

Textual.

- DEMIR_MC_1: In this run we applied porter stemmer algorithm during text processing procedure and used InL2 (Divergence from Randomness Framework) matching model to calculate weighting factor of each term.
- DEMIR_MC_2: The main difference of this run and the previous run are no-stemming and usage of TF×IDF weighting scheme.

Mixed.

- DEMIR_MC_3: This run uses Visual feature Set $_2$ and the same textual features with run DEMIR_MC_2. This weighting factor try to balance text and visual features, where the weighting factor for both visual and texture feature is equal to 1
- DEMIR_MC_4: The only difference of this run and the run DEMIR_MC_3 is the weight of text similarity score is 1.7.
- DEMIR_MC_5: Text feature set and ICWF factor of this run is the same with run DEMIR_MC_3. Includes CLD, FCTH and CEDD as visual features and weight of visual and textual fearures are equal.
- DEMIR_MC_6: This run is similar to run DEMIR_MC_5, weight for textual similarity score is to 1.7.

4.3 Results

For modality classification subtask, we submit runs for text and mixed methods. Among our runs we get the best performance in mixed category, inputs are visual CEDD, CLD, FCTH and textual features. Integrated combination multimodal retrieval applied to retrieved image sets of each visual and text features, after than our information retrieval based classification algorithm is performed. If we evaluate run results for each method separately, for text based-method, applying information retrieval based classification algorithm to text feature set correctly classify %62.70 of test images by run DEMIR_MC_1 and DEMIR_MC_2. This performance is the third best performance among other groups' runs. Text based run results show that applying a stemming algorithm during text processing or different term weighting algorithms do not affect the performance directly. Following, best run performance for the mixed method is obtained by run DEMIR_MC_5. Textual features of this run is the same with the run DEMIR_MC_2 and CLD, FCTH and CEDD visual features' values used as input to our algorithm. The most important point of this run is the value of ICWF value is 1.0. Mixed method runs indicate the effect of two major parameters. Firstly; weighting visual and text features equally by ICWF gives better results. Following, our synthetic visual terms generated by clustering do not have a positive impact on test results.

Table 1. Runs of DEMIR group for modality classification task in ImageCLEFMed 2013 AMIA: Medical Retrieval

RunID	Type	Correctly classified in %	Text Parameters		Mixed Parameters	
			Stemming Method	Matching Model	Visual Feature Set	ICWF
DEMIR_MC_5	Mixed	64.60	No Stemming	TF_IDF	CEDD, CLD, FCTH	1.0
DEMIR_MC_3	Mixed	64.48	No Stemming	TF_IDF	VT_{CEDD}, VT_{CLD}, VT_{FCTH},	1.7
DEMIR_MC_6	Mixed	64.09	No Stemming	TF_IDF	CEDD, CLD, FCTH	1.0
DEMIR_MC_4	Mixed	63.67	No Stemming	TF_IDF	VT_{CEDD}, VT_{CLD}, VT_{FCTH},	1.7
DEMIR_MC_1	Textual	62.70	Porter Stemmer	InL2	-	
DEMIR_MC_2	Textual	62.70	No Stemming	TF_IDF	-	

5 Ad-hoc Image-based Retrieval

5.1 Methods

In this year, we tested effectiveness of our information retrieval System (IRS) based test classification approach on information retrieval performance. So, in addition to the results of last year's baseline runs, we preferred to filter out such documents using classification methods, which is, explained details in section 4.1 and narrowing data collection down.

For the results of textual modality, we applied the basic processes in information retrieval system that consists of preprocessing, indexing and retrieval stages using Terrier. Additionally, in some runs, we assumed the results of modality classification as textual modality's result.

For the results of visual modality, we extracted CEDD, FCTH and CLD features using Rummager tool (Chatzichristofis et al., 2009). We created a VSM for each feature, for each document in collection. After we calculated the similarities using using

Euclidean distance function. Then, we normalized among them and combined by averaging as a visual modality result.

For mixed modality, we kept the results for textual and visual modalities and performed integrated weighted CombSUM combination such that coefficient of text modality was 1.7 folds of visual modality.

5.2 Runs

We submitted 10 runs to ImageCLEF 2013 AMIA: Medical Retrieval for ad-hoc image-based retrieval task three categories. Below, we provide a short description of each run, shortly.

Textual.

- DEMIR1: This run is our baseline retrieval result for textual modality. In this run, caption tag is used in indexing documents. Retrieval weighting model is TF×IDF. UTF tokeniser and stop-word list were used and porter stemmer was applied. We used *en-description* field in topics file for each topic when retrieved. We obtained the best result in ImageCLEF2012 Medical Image Classification and Retrieval, using this method so we have got it as a baseline for this year.
- DEMIR6: We assumed that the result of modality classification is a result of textual modality, instead of Terrier result. We evaluated according to result of order 1 (k=1; first k modalities of sorted candidate modalities given in text based modality classification method) using modality classifications.
- DEMIR8: After classification of documents and topics using our modality classification based methods, we filtered out the documents and retrieved from the classified documents which were result of order 0 using textual features distance.
- DEMIR9: This run is similar to run DEMIR8, but in this run, we retrieved from the classified documents that were ordered in the first three documents (k={1,2,3}; first k modalities of sorted candidate modalities given in text based modality classification method) using textual features distance.

Visual.

- DEMIR2: This run is our baseline retrieval for visual retrieval type. We used CEDD, FCTH and MPEG7-CLD features, and Euclidean distance to calculate the similarities between topics and documents. We normalized distance scores as topic level. We calculated median of sum used features scores and combined them. We used the maximum values for getting topic results from their figures results that has a combination of used visual features.
- DEMIR4: After classification of documents and topics, we filtered out the documents and retrieved from the classified documents which were result of order 0 using visual features distance.

- DEMIR5: This run is similar to run DEMIR4, but in this run, we retrieved from the classified documents that were ordered in the first three documents using visual features distance.

Mixed.

- DEMIR3: We performed integrated weighted combination such that coefficient of text modality was 1.7 folds of visual modality. This run is baseline for mixed retrieval type.
- DEMIR7: We performed integrated weighted combination such that coefficient of text modality in DEMIR6 was 1.7 folds of visual modality in DEMIR2. We obtained text modality results from modality classification results be applied to documents.
- DEMIR10: We performed integrated weighted combination such that coefficient of text modality in DEMIR1 was 1.7 folds of visual modality in DEMIR2 and we filtered out results using modality classification results according to order 0.

5.3 Results

For ad-hoc image-based retrieval subtask, we submitted 10 runs to ImageCLEF Medical Retrieval task, in three different categories: textual-only, visual-only and mixed retrieval types. Among our runs, we get the best performance in mixed category, in run DEMIR3; inputs are visual CEDD, CLD, FCTH and textual features as in the modality classification subtask. If we evaluate run results (DEMIR9, DEMIR1, DEMIR6, DEMIR8) for textual-only retrieval type, when we filtered out documents using information retrieval based classification algorithm and retrieved from the classified documents, we achieved the best result as in run DEMIR9 than other runs. As in the textual-only, the information retrieval based classification algorithm improves the performance for visual-only retrieval type. We applied classification algorithm in runs DEMIR4 and DEMIR5, so these runs performances are better than DEMIR2 for visual-only.

Table 2. Runs of DEMIR group for ad-hoc image-based retrieval task in ImageCLEFMed 2013 AMIA: Medical Retrieval

RunID	Type	MAP	GM-MAP	bpref	P10	P30
DEMIR3	Mixed	0.2168	0.0345	0.2255	0.3143	0.1914
DEMIR9	Textual	0.2003	0.0352	0.2158	0.2943	0.1952
DEMIR1	Textual	0.1951	0.0289	0.2036	0.2714	0.1895
DEMIR6	Textual	0.1951	0.0289	0.2036	0.2714	0.1895

DEMIR10	Mixed	0.1583	0.0292	0.1775	0.2771	0.1867
DEMIR8	Textual	0.1578	0.0267	0.1712	0.2714	0.1733
DEMIR7	Mixed	0.0225	0.0003	0.0355	0.0543	0.0543
DEMIR4	Visual	0.0185	0.0005	0.0361	0.0629	0.0581
DEMIR5	Visual	0.011	0.0004	0.0257	0.04	0.0448
DEMIR2	Visual	0.0044	0.0002	0.0152	0.0229	0.0229

6   Conclusion

In this year, we examined effects of modality classification to retrieval performance. Among our runs, we get the best performance rather than our other submitted runs in mixed category for both subtasks. Integrated combination multimodal retrieval applied to retrieved image sets of each visual and text features, after than our information retrieval based classification algorithm is performed. We also used our integrated combination method, which was used by our team DEMIR at last year, on different level of multimodality retrieval system and again, we agree that proper combination model can improve the performance of multimodal retrieval systems. We used CEDD, FCTH and CLD features as low-level features in visual modality.

For modality classification subtask, we apply integrated combination multimodal retrieval and our information retrieval based classification algorithm. Text based run results show that applying a stemming algorithm during text processing or different term weighting algorithms do not affect the performance directly. Mixed method runs indicate the effect of two major parameters. Firstly; weighting visual and text features equally by ICWF gives better results. Another one, our synthetic visual terms generated by clustering do not have a positive impact on test results.

For ad-hoc image-based retrieval subtask, our proposed approach is based on the information retrieval based classification algorithm. We aim to evaluate the effects of classification method on information retrieval system. So, in addition to the results of last year's baseline runs, we preferred to filter out such documents using modality classification method and retrieved from the classified documents. As in the textual-only, the information retrieval based classification algorithm improves the performance for visual-only retrieval type.

References

Alpkocak A, Ozturkmenoglu O, Berber T, Vahid AH, and Hamed RG (2011) DEMIR at ImageCLEFMed 2011: Evaluation of Fusion Techniques for Multimodal Content-based

Medical Image Retrieval. In *Conference and Labs of the Evaluation Forum (CLEF)*. Petras V, Forner P, andClough PD (ed.), (ed.), Vol. pp. Amsterdam, The Netherlands.

Chatzichristofis SA and Boutalis YS (2008a) CEDD: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In *Proceedings of the 6th international conference on Computer vision systems*. (ed.), Vol. pp. 312-22, Springer-Verlag, Santorini, Greece.

Chatzichristofis SA and Boutalis YS (2008b) FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval. In *Proceedings of the 2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*. (ed.), Vol. pp. 191-6, IEEE Computer Society,

Chatzichristofis SA, Boutalis YS, and Lux M (2009) Img(Rummager): An Interactive Content Based Image Retrieval System. In *Proceedings of the 2009 Second International Workshop on Similarity Search and Applications*. (ed.), Vol. pp. 151-3, IEEE Computer Society, Prague, Czech Republic.

Lux M and Chatzichristofis SA (2008) Lire: lucene image retrieval: an extensible java CBIR library. In *Proceedings of the 16th ACM international conference on Multimedia*. (ed.), Vol. pp. 1085-8, ACM, Vancouver, British Columbia, Canada.

Ounis I, Amati G, Plachouras V, He B, Macdonald C, and Lioma C (2006) Terrier: A High Performance and Scalable Information Retrieval Platform. In *ACM SIGIR'06 Workshop on Open Source Information Retrieval (OSIR)*. (ed.), Vol. pp. Seattle, Washington, USA.

Porter MF (1997) An algorithm for suffix stripping. In *Readings in information retrieval*. Karen Sparck J and Peter W (ed.), Vol. pp. 313-6. Morgan Kaufmann Publishers Inc.,

Vahid AH, Alpkocak A, Hamed RG, Ceylan NM, and Ozturkmenoglu O (2012) DEMIR at ImageCLEFMed 2012. In *CLEF 2012 Evaluation Labs and Workshop*. Forner P, Karlgren J, andWomser-Hacker C (ed.), (ed.), Vol. pp. Rome, Italy.