DOI: https://doi.org/10.3217/978-3-85125-609-3-30

Immersive analytics for the suggestion of opinions in the analysis of a medical report from a collaborative social network

Riadh Bouslimi, Mouhamed Gaith Ayadi, Khaoula Hamaied and Mariem Medini

Computer Science Department, BESTMOD Lab, ISG-University of Tunis, Bardo, Tunisia

{bouslimi.riadh, mouhamed.gaith.ayadi}@gmail.com, kawlahamaied@yahoo.fr and medinimeriem@hotmail.fr

Abstract. Doctors are mostly in need of several opinions in order to make a good decision. Indeed, medical social networks have become at place of exchange of experiences between doctors. In addition, supervised learning of medical record will help physicians quickly find the right concise explanation of their medical images. To do this, we present in this article a model of opinion suggestion based on a collaborative social radiological network. Indeed, the opinions shared on a medical image in a medico-social network are represented in the form of a textual description which in most cases requires a cleaning using a medical dictionary. In addition, we describe the textual description of medical image with TF-IDF weight vector using a " bag of words" approach. We use latent semantic analysis to establish relationships between textual terms and visual terms in shared opinions about the medical image. Multimodal modeling looks for medical information through multimodal queries. Our model is evaluated against the ImageCLEFmed baseline, which is the basic truth for our experiments. We have conducted many experiments with different descriptors and many combinations of modalities. The analysis of the results shows that the model based on two methods makes it possible to increase the performance of a search system based on a single modality, visual or textual.

Keywords: Multimodal modeling of medical information, Radiological social network, Multimodal information retrieval, Immersive analytics, Bag-of-words, Big data.

1 Introduction

Social networks on the theme of health seem to flourish on the Web in recent years. These social networks bring together the health professionals or only doctors and open to industry (pharmaceutical, ..) or to patients. However, these networks their mission is to revolutionize medicine in the real world by accomplishing a connection between doctors, exchange ideas safely without wasting time. Patients can have real-time notifications of doctors and thus have different opinions.

© Springer-Verlag Berlin Heidelberg 2011

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

Content-based medical image retrieval is relative to the context of his capture and interpretation by a radiologist. Current techniques do not allow us to extract images, visual characteristics of low level (color, texture or interest point). The question frequently asked: how best to use the visual characteristics of low level to link automatically to concepts? This problem is known by the name of "semantic gap" [1]. When both textual and visual modalities are reunited, it will be essential to exploit the two assemblies. Any time, we must take into account that it is easier to automatically associate meaning to a text than to an image.

Many image retrieval systems by the text are realized such as (Google Images, Yahoo!, Flickr…) which are based on information from image annotations. For example, Google indexes web images according the surrounding text (file name, description, link, ...) and Flickr indexes the database images according the keywords that the users assign themselves to images. However, the indexing of text associated with the image is the simplest solution to implement what gives the idea to associate with new images annotations existing in a database [2].

To automatically annotate new images, there are two approaches: the first which is based on a supervised learning and training images are manually classified. Another approach is to automatically discover hidden links between visual and textual elements using unsupervised learning methods [3]. This technique introduces a set of latent variables meant to represent the co-distribution of visual and textual elements.

To annotate a new image, we must first extract the visual description and a function of probabilistic similarity, will return the state that maximizes the probability density of the text annotation and the visual element. Finally, the annotations are sorted by probability values.

Medical information in collaborative medical social networks do not cease to grow, from which comes the necessity to develop medical image retrieval and annotation systems appropriate self while using both modalities. A natural approach is to use the representation based on "bag of words" for modeling the image. This approach has already proven effective, especially for image medical annotation applications [4][5]. Indeed, standard collections as TREC, or ImagEVAL ImageCLEFmed for the evaluation of these systems. We propose a model that combines both visual and textual information in collaborative social networks of health. The relevance of our model is evaluated on a search for medical information. This makes comparing the results obtained with our model to those obtained with a single modality, textual or visual.

The information seekers whether it be a patient, a physician, health personnel etc. can search useful information in medical social network and that is the reason which we propose in this article our model of representation of information in a social network Medical. Next, we present a model based of multimodal research by applying a fusion by latent analysis. Finally, we show the results of the application on the ImageCLEFMed dataset.

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

2 Literature Review

Social networks have not only revolutionized the way people communicate and interact with each other [6][7]. But also play an important role in health [8]. In the next two sections we will present immersive analysis and work similar to our work.

2.1 Immersive Analysis

In the last decade, the "Immersive Analytics" [9] is an emerging research field. It is an axis that explores how new interaction, display and analysis technologies can be used to support analytical reasoning and emphasize decision-making. This new area of research aims to provide panoply of multi-sensory interfaces that allow collaboration between imagination, scene and intuition. In other words, allowing a user to deeply extend the data and to go very far in his analysis. In fact the immersive analysis [10] is about presenting us an immersive environment that relies on new display technologies such as large touch surfaces, sensors and many other features that make it easy to display data in complete clarity using 2D or 3D techniques which in turn facilitates their deep analysis that will later serve decision makers. Visual analysis [11] is the result of a combination of analytical reasoning of projected data on interactive user interfaces. It's actually the key to exploring and understanding a mass of data. According to several researches [12], there is a big difference between simply presented data, and data represented in a more visual way. The representation of data is the key to good reasoning. We are in a world of big data, certainly there is always hidden information, it is here that appear the importance of data analysis. And this is actually the goal of a visual analysis that allows you to discover the hidden data and display it effectively and in a more pleasant way. What to say if we add intuition and human perception! So, immersive analysis is the combination of visual analysis and geographical representation. That's why in fact we thought to take advantage of this science to allow the user to live the situation by making comparisons, experimenting once he is connected to his immersive environment in order to support the analysis of the situation and therefore guide him in his decision-making process. In the next section, a presentation of our approach that results from a combination studies already presented in the state of the art section.

2.2 Related Work

The co-occurrence model proposed by Mori et al. represents the first approach for associations between text and image [13]. First, images of the training set are divided into regions that inherit all the keywords of original images from which they depend. Visual descriptors are then extracted from each region. All descriptors are clustered into a number of groups, each of which is represented by its center of gravity. Last, the probability of each keyword for each of the region groups can be measured.

[14] proposed a translation model to represent the relationships between text and content. According to their view, visual features and text are two languages that can be translated from one to the other. Thanks to a translation table having estimations of

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

probability of the translation between image's regions and keywords, an image is annotated by choosing the most probable keyword for each of regions.

[15] have extended the translation model of [14] to a hierarchical model. It combines the "aspect" model [16] which builds a joint distribution of documents and features, with a soft clustering model [15] which maps documents into clusters. Images and text are generated by nodes arranged in a tree structure. The nodes generate image regions using a Gaussian distribution, and keywords using a multinomial distribution.

[17] suggested improvements to the results of [14] by introducing a language generation model, called Cross-Media Relevance Model (CMRM). First, they use the same process as Duygulu et al. for calculating the representation of images (represented by blobs). Then [14] made the assumption that there is a one-to-one correspondence between regions and words, while Jeon et al. assume that a set of blobs is related only to a set of words. Thus, instead of seeking a probabilistic translation table, CMRM simply calculates the probability of observing a set of blobs and keywords in a given image.

[18] proved that the process of features quantifying using a Continuous-space Relevance Model (CRM) can avoid losing information related to the production of the dictionary in the CMRM model. [17] Using continuous features of probability density to estimate the probability of observing a particular region in an image, they showed that the model performance on the same dataset is much more efficient than the models proposed by [14] [17].

Some studies have attempted to use the LSA technique for combining visual and textual features, including [16, 19] who applied the Probabilistic Latent Semantic Analysis for automatic image annotation. With this approach, text and visual features are considered as "terms". It assumes that each term may come from a number of latent subjects, and each image can contain multiple subjects.

In the transformation model, [20] the text query is automatically converted into visual representations for image retrieval. First, the relationship between text and images are taken from a set of images annotated with text descriptions. A transmedia dictionary which is similar to a bilingual dictionary is set up in the training set.

[21] propose to do the opposite, which is to translate an image query into a text query. Based on both textual and visual queries, the authors transform visual queries into textual queries, and acquire new textual queries. After that, they apply text retrieval techniques to deal with initial textual queries and new textual queries constructed from the visual query for image retrieval. Finally, they merge the results.

Recently, nearest neighbor methods which treat image annotation as image retrieval problem, have received more attention. [22] introduce a baseline technique that transfers keywords to images using its nearest neighbors. A combination called Joint Equal Contribution (JEC) of basic distances to find nearest neighbors is used on low-level image features; the keywords are then assigned using a greedy label transfer mechanism. A more complex nearest-neighbor-type model called TagProp is proposed by [23]. The model combines a weighted nearest-neighbor approach with metric learning capabilities in a discriminative framework which allows the integration of metric learning by directly maximizing the log-likelihood of the tag predictions in the training set.

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

2.3 Our proposed model

As part of this work, we aim at giving answers to issues raised in the previous section. First, in order to avoid the dependence on the quality of the segmentation phase, we are working in a context without segmentation. In order to support image retrieval by textual queries independently of any manual annotation, we propose to add the bidirectional transformation between text and image. Finally, we place ourselves in a system with incremental knowledge learning, which requires no special knowledge at the beginning of the life of the system. This constraint seems essential, but also realistic, because most applications do not have specialized knowledge in their early life.

In our model, text/image associations are learnt by an incremental learning method via relevance feedback without any knowledge at first. Unlike other models where prior knowledge is available, in our system, knowledge comes from user interactions. Therefore, our system knowledge is progressively improved over time through interactions, without requiring any off-line learning stage.

3 Representation of information

Content-based image retrieval is difficult to automate, since it depends on the representation of the information. Several methods have been proposed in recent years to build a content-based image retrieval systems. However, the lack of explicit information on user request, and the real difficulties for a computer to effectively interpret the content of an image, make this problem of indexing and image retrieval is especially hard. In fact, the content-based image retrieval is performed on the signified image, ie on its interpretation by a human reader, as well as to the context of his capture. This context should be considered as a meta information. This is, to retrieve image with the name of the place visited, the name of the animal photographed, the name of the person close-up etc... However, current technical computer vision only allow us to extract images, visual characteristics of low level (color, texture or point of interest). The distance involved to automatically interpret the content of an image is still very important. The recurrent question is increased in these terms: how best to use the visual characteristics of low level to connect automatically to concepts ? This problem is known as the semantic gap. When the textual and visual modalities are combined into one document, it seems better to simultaneously operate these two types of content. You should know, firstly it is easier to automatically associate meaning to a text than an image, and secondly, it is easier to compare similar images, it is hoped that complementarity between the image and the text is conducive to better indexing, relative to the two media separately. However, several studies have been published using bag of word model to image retrieve.

3.1 Representation of information in a medical social network

In a medical social network, several physicians' can share their opinion on a medical image. However, the shared opinions require automatic processing to extract relevant keywords. We present our model of representation of medical social network, which is

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

to describe the text and images with textual terms and visual terms. The two modalities are first processed separately using the approach BoW to the visual and textual description. Indeed, they are represented as TF-IDF weight vector characterizing the frequency of each visual or textual words. The vector describing the textual content is cleaned using the UMLS thesaurus. To use the same mode of representation for the two modalities can be combined with a fusion method by Latent Semantic Analysis (LSA), after making multimodal queries to retrieve information.

Representation of textual modality

To represent a text report in the form of a weight vector, it is first necessary to define an index of textual or vocabulary terms. For that, we will apply initially a stemming algorithm with Snowball and we delete the black words from all reports. The indexing will be performed by the Lemur software[1]. However, the terms selected are then filtered using the thesaurus UMLS. Each report is then represented following the model of Salton [24], is as a weight vector $r_i^T = (w_{i,1}, \dots, w_{i,j}, \dots w_{i,|T|})$, where $w_{i,j}$ represents the weight of the term t_j in a report r_i. This weight is calculated as the product of two factors $tf_{i,j}$ and idf_j. The factor tf_{ij} is the frequency of occurrence of the term t_j in the report r_i and the factor idf_j measures the inverse of the frequency of the word in the corpus. Thus, the weight $w_{i,j}$ is even higher than the term t_j, and frequent in the report r_i and rare in the corpus. For the calculation of tf et idf, we are using the formulations defined by Robertson [25].

where $|R|$ is the size of the corpus and $\left|\{r_i | t_j \in r_i\}\right|$ is the number of documents in corpus, which the term t_j appears at least once. A textual query q_k can be considered a very short text report, it can also be represented by a weight vector. This vector is noted q_k^T, will be calculated with formulas of Roberson but with b = 0. To calculate the relevance score of a report r_i opposite an query q_k, we apply the formula given by Zhai in [26] and defined as:

$$score_T(q_k, r_i) = \sum_{j=1}^{|T|} r_{i,j}^T q_{k,j}^T \quad (1)$$

Representation of medical image.

The representation of the visual modality is carried out in two steps: the creation of a visual vocabulary and the representation of the medical image using thereof. The vocabulary V of the visual modality is obtained using the approach BoW [27]. The process consists of three steps: the choice of regions or interest points, the description by calculating a descriptor of points or regions and grouping of descriptors into classes constituting the visual words. We use two different approaches for the first two steps. The first approach uses a regular cutting of the image into n^2 thumbnails. Then a color descriptor with 6 dimension, denoted Meanstd, is obtained for each thumbnail, by calculating the mean and standard deviation of normalized components $\frac{R}{R+G+B}$, $\frac{G}{R+G+B}$ et

[1] http://www.lemurproject.org/

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

$\frac{R+G+B}{3\times255}$ where R, G and B are the colors components. The second approach uses the characterization of images with regions of interest detected by the MSER [28] and presented by their bounding ellipses (according to the method proposed by [29]). These regions are then described by the descriptor SIFT [30]. For the third step, the grouping of classes is performed by applying the k-means algorithm on the set of descriptors to obtain k clusters descriptors. Each cluster center then represents a visual word. The representation of an image using the vocabulary defined previously for calculating a weight vector r_i^V exactly as for the text modality. To obtain a visual words from the medical image, we first calculate the descriptors on the points or regions of the image, and then is associated, at each descriptor, the word vocabulary, the nearest in the sense of the Euclidean distance.

Fusion Model.

In this section, we show the fusion of both textual and visual descriptors using latent semantic technology. However Latent Semantic Indexing (LSI) was first introduced in the field of research information and has proven its effectiveness in recent years [29]. This technique involves reducing the indexation matrix in a new space sensible to express more "semantic" dimensions. This reduction is intended to make it appear the "hidden semantics" in the co-occurrence links. This is called latent semantic. This latent semantic allows for example to reduce the effects of synonymy and polysemy. It is also used to index without translation, no dictionary, parallel corpus, that is to say composed of documents in different languages, but supposed to be translations of each other. Technically, LSA method is a matrix transformation operation M of co-occurrence between terms and documents. This is in fact a singular value decomposition[2] it of the matrix : $M_{i,j}$ describes the occurrences of term i in document j.

Suggestion opinions through research similar medical images.

Each modality of reports (text and image) is processed independently. We obtain a textual matrix report-term $M_{r,t}$ and the visual matrix report-term matrice $M_{r,v}$. The fusion of these two methods is first obtained simply by concatenating the columns of the two matrices $M_{r,t}$ and $M_{r,v}$ in a matrix $M_{r,vt}$ because it is of different coordinates on the same set of documents. This merged matrix is then projected in a latent space to obtain the latent matrix $M_{r,k}$, with k the new reduced size. Therefore, each document is represented by a line latent matrix $M_{r,k}$. For a query containing text and images, we apply the same process with the reports. Then, this vector is projected into the reduced space for a pseudo-vector $q_k = q * \sum_k V_k^t$. Finally, the calculation of the value of relevance of a document to the query (Relevance Status Value or RSV) is calculated according the similarity of the query vector q_k with the lines of the latent matrix by using the function *cosinus*.

[2] Singular Value Decomposition (SVD)

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

We present below our algorithm that shows the search steps in a medical social network :

Algorithm SuggestionOpinions
Input :
R_q: Query of a medical report (Text or Medical Image or Text with Medical Image)
Output :
TImages : The list of similar images
Var
$M_{rq,vt}, M_{r,vt}$: Result from the concatenation of two matrices
$M_{rq,k}, M_{r,k}$: Latent Matrix
q_k : Query projected in a small space
0. **Begin**
1. $M_{rq,vt}$ ← Extract_Matrix_Terms(R_q) U Extract_Matrix_Visual_Terms(R_q)
2. $M_{rq,k}$ ← Extract_Lalent_Matrix ($M_{rq,vt}$)
3. $q_k \leftarrow q * \sum_k V_k^t$
4. **ForEach** R **In** the social network DataBase **Do**
5. $M_{r,vt}$ ← Extract_Matrix_Terms(R) U Extract_Matrix_Report_Visual_Terms(R)
6. $M_{r,k}$ ← Extract_Lalent_Matrix ($M_{r,vt}$)
7. TImages ← Extraction_of_the_nearest_Medical_Images (Cosinus (q_k, $M_{rq,k}$))
8. **End ForEach**
9. **End**.

In the following, we present in figure 1 our search engine based on immersive analysis to suggest opinions about similar clinical cases.

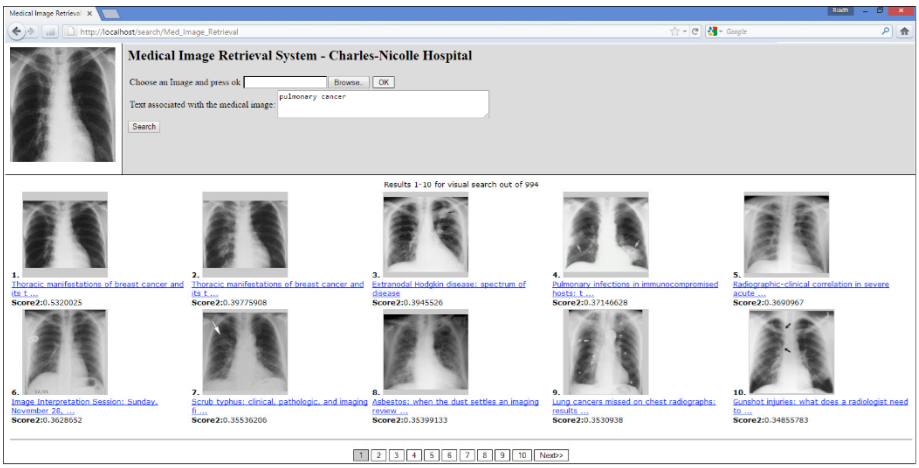

Fig. 1. Our medical image retrieval system on a text + image query.

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

4 Experimental evaluation

4.1 Test data and evaluation criteria

The pertinence of our model is evaluated on the collection provided for the competition ImageCLEFMed [30]. This collection is composed over 45,000 biomedical research articles of the PubMed Central (R). Each document is composed of an image and a text part. The images are very heterogeneous in size and content. The text part is relatively short with an average of 33 words per document. The goal of the information search task is to return to the 75 queries supplied by ImageCLEFMed a list of pertinent documents. All queries have a textual part, but many do not have a query image. Order to have a visual part for each query, we use the first two pertinent medical images returned by our system when we use only the textual part. This corresponds to a relevance feedback fact by the system user. The criteria of average accuracy (*Mean Average Precision - MAP*), which is a classic criteria in information retrieval, is then used to evaluate the pertinence of the results.

4.2 Results and Discussion

To demonstrate the contribution of the use of our model compared to only textual or visual model, we realized experiments using a single modality, textual or visual, then experiments combining two modalities, modality text with visual descriptor, this visual features for both Meanstd and SIFT previously presented. The text vocabulary is consists of approximately 200000 words whereas the two visual vocabularies are constituted of 10000. Table summarizes the MAP values obtained for each experiment. On the one hand, it can be stated that the use of the single visual modality irrespective of the descriptor used leads to poorer results that the use of the only modality text. On the other hand, combine a visual descriptor with the text improves search performance with the only textual descriptor. These overall observations are confirmed by the precision/recall curves presented in Figure 2. A detailed analysis per query show that, for some, the first results returned by the visual modality is best for text modality. We can add, about the performance obtained with a visual modality, that the regular division of the associated image at Meanstd color descriptor is more robust than MSER + SIFT. We explain this behavior by clustering problems. With the color descriptor, we work with 6 characteristic parameters and 4 million thumbnails to consolidate vocabulary words. With SIFT descriptor, we have 128 features and settings 54 million thumbnails. In the second case, the thumbnails are divided very irregularly in the space of descriptors, because of the use of MSER, the large size and the large amount of data. This situation is very unfavorable for clustering algorithms such as K-means [31]. Also, it has been shown in [32] [33] that the descriptors of the most densely spaces of the parameter space are not necessarily the most informative.

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

Table 1. Result of average precision obtained for different modalities

Modality type	MAP
SIFT	0.1287
Meanstd	0.0962
Text	0.2346
Fusion : Text + SIFT	0.3667
Fusion : Text + Meanstd	0.2734

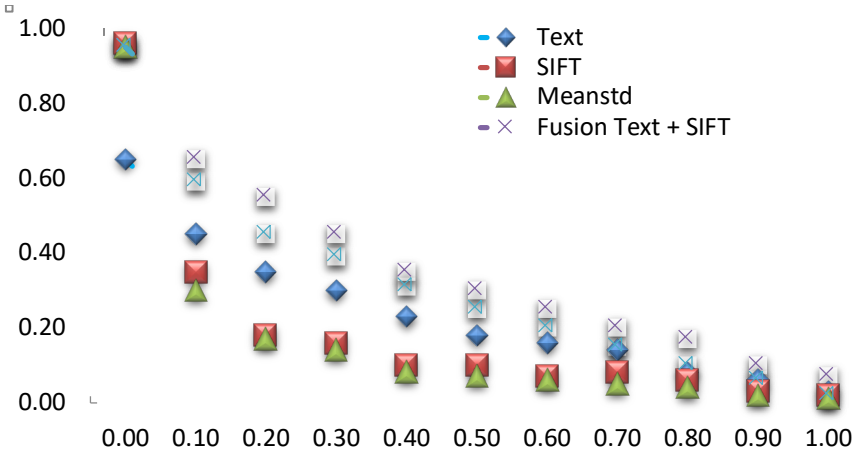

Fig. 2. Precision / Recall curve of modalities (text only, visual only and fusion text/visual)

5 Conclusion

We have presented in this paper a representing model of multimedia data extracted from radiological social networks where they are used for annotating medical images. This model is based on a fusion with LSA of the textual and visual information using the "bag-of-words" approach.

The performance of the indexing and search system has been evaluated based on real dataset of ImageCLEFMed and the obtained results were very promising to use a media model specialized radiology, like the one proposed to retrieve information from a collection of radiological media. Indeed, the fusion of both textual and visual methods allows each time to increase the performance of the system. In this context, Larlus [31] proposes a clustering method that allows uniform quantifications of spaces contrary to K-means which focuses on dense spaces. This method could be used to improve our system when creating the visual vocabulary.

References

1. Priyatharshini R, Chitrakala S (2012) Association Based Image Retrieval: A Survey. Second International Joint Conference, AIM/CCPE 2012, Bangalore, India, pp 17-26.

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

2. Duan L, Yuan B, Wu C, Li J, Guo Q (2014) Text-Image Separation and Indexing in Historic Patent Document Image Based on Extreme Learning Machine. Proceedings of ELM-2014 Volume 2, Volume 4 of the series Proceedings in Adaptation, Learning and Optimization pp 299-307.
3. Clinchant S, Csurka G, Ah-Pine J (2011) Semantic combination of textual and visual information in multimedia retrieval. in Proc. 1st ACM Int. Conf. Multimedia Retrieval, New York, NY, USA.
4. Bouslimi R, Akaichi J (2015) Automatic medical image annotation on social network of physician collaboration. Journal of Network Modeling Analysis in Health Informatics and Bioinformatics, 4(10): 219-228.
5. Bouslimi R, Akaichi J, Ayadi M G, Hedhli H (2016) A medical collaboration network for medical image analysis. Journal of Network Modeling Analysis in Health Informatics and Bioinformatics,5(10): 145-165.
6. Bilenko M, Mooney R J (2003). Adaptive duplicate detection using learnable string similarity measures. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 39–48. ACM.
7. Chen Z, Kalashnikov D V, and Mehrotra S (2009) Exploiting context analysis for combining multiple entity resolution systems. In Proceedings of the2009 ACM SIGMOD International Conference on Management of data, pages 207–218. ACM.
8. Gao H, Barbier G, and Goolsby R (2011) Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intelligent Systems, 26(3):10–14.
9. Donalek, S. Djorgovski, A. Cioc, A. Wang, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, A. Drake, S. Davidoff, J. Norris, and G. Longo (2014) Immersive and collaborative data visualization using virtual reality platforms, in 2014 IEEE International Conference on Big Data (Big Data), pp. 609–614.
10. Mahfoud E, Yuemeng L, Wegba K, Hongley H, Aidong L (2018) Immersive Visualization for Abnormal Detection in Heterogeneous Data for On-site Decision Making: In Hawaii International Conference on System Sciences, pp 1300-1309, Hawaii.
11. Paravati G, Lamberti F, Sanna A, Ramirez E H, Demartini C (2012). An immersive visualization framework for monitoring, simulating and controlling smart street lighting networks. In Proceedings of the 5th International ICST Conference on Simulation Tools and Techniques, pp 17-26.
12. Bach B, Dachselt R, Carpendale S, Dwyer T, Collins Christopher, Lee B (2016) . Immersive Analytics : Exploring Future Interaction and Visualization Technologies for Data Analytics. ISS: 529-533.
13. Mori Y, Takahashi H, Oka R (1999), Image-to-word transformation based on dividing and vector quantizing images with words, in Proc. First Int. Workshop Multimedia Intelligent Storage and Retrieval Management (Orlando, Florida, USA, 1999), pp. 405–409.
14. Duygulu P, Barnard K, Freitas J. F. G, Forsyth D. A (2002) Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary, in Proc. ECCV'02 (Springer-Verlag, London, UK, 2002), pp. 97–112.
15. Barnard K, Duygulu P, Forsyth D, deFreitas N, Blei D M, Jordan M I. (2003). Matching Words and Pictures. J. Mach. Learn. Res., 3:1107–1135.
16. Hofmann T (1998), Learning and representing topic. A hierarchical mixture model for word occurrences in document databases, in Proc. CONALD'98, Pittsburgh.
17. Jeon J, Lavrenko V, Manmatha R (2003), Automatic image annotation and retrieval using cross-media relevance models, in Proc. SIGIR'03 (ACM, New York, NY, USA, 2003), pp. 119–126.

DOI: https://doi.org/10.3217/978-3-85125-609-3-30

18. Lavrenko V, Manmatha R, Jeon J (2004), A model for learning the semantics of pictures, NIPS, pp. 553–560.
19. Monay F, Gatica-Perez D (2003), On image auto-annotation with latent space models, in Proc. MULTIMEDIA'03 (ACM, New York, NY, USA), pp. 275–278.
20. Liu Y, Navathe SB, Pivoshenko A, Dasigi V, Dingledine R, Ciliax BJ (2006). Text Analysis of MEDLINE for Discovering Functional Relationships among Genes: Evaluation of Keyword Extraction Weighting Schemes. International Journal of Data Mining and Bioinformatics. 1:88-110.
21. Chang Y. C and Chen H. H (2008), Using an image-text parallel corpus and the web for query expansion in cross-language image retrieval, in Advances in Multilingual and Multimodal Information Retrieval (Springer-Verlag Berlin, Heidelberg), pp. 504–511.
22. Makadia A, Pavlovic V and Kumar S (2010), Baselines for image annotation, Int. J. Comput. Vis. 90, 88–105.
23. Guillaumin M, Mensink T, Verbeek J J, Schmid C (2009), TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation, in Proc. ICCV, pp. 309–316.
24. Salton G, Wong A, Yang C (1975) A vector space model for automatic indexing. Communication ACM, 18(11) :613–620.
25. Robertson S, Walker S, Hancock-Beaulieu M, Gull A, Lau. M Okapi (1994) at trec-3. In Text REtrieval Conference, pages 21–30.
26. Zhai C (2001). Notes on the lemur TFIDF model. Technical report, Carnegie Mellon University.
27. Csurka G, Dance C, Fan L, Willamowski J, Bray C (2004). Visual categorization with bags of keypoints. In ECCV'04 workshop on Statistical Learning in Computer Vision, pages 59–74.
28. Matas J, Chum O, Martin U, Pajdla T (2002) Robust wide baseline stereo from maximally stable extremal regions. In Proceedings of the British Machine Vision Conference, pages 384–393. BMVA.
29. Abd Rahman N , Mabni Z, Omar N, Fairuz H, Hanum M, Amirah N Nur, Mohamad Rahim T (2015) A Parallel Latent Semantic Indexing (LSI) Algorithm for Malay Hadith Translated Document Retrieval. First International Conference, SCDS 2015, Putrajaya, Malaysia, pp 154-163.
30. Garcia Seco de Herrera A, Muller H, Bromuri S (2015) : Overview of the ImageCLEF 2015 medical classification task. In: Working Notes of CLEF 2015 (Cross Language Evaluation Forum).
31. Larlus D, Dorkó G, Jurie F (2006). Création de vocabulaires visuels efficaces pour la catégorisation d'images. In Reconnaissance des Formes et Intelligence Artificielle.
32. Jurie F, Triggs W (2005) Creating efficient codebooks for visual recognition. In ICCV.
33. Vidal-Naquet M, Ullman S (2003). Object recognition with informative features and linear classification. In ICCV, pages 281–288.