IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-6, Issue-9, Sept. 2018 E-ISSN: 2347-2693

A Brief Survey Ondynamic Topic Model for Unsupervised Object Discovery and Localization

Mereena Johny[1*], L. Haldurai[2]

[1]Dept. of Computer Science, Kongunadu Arts and Science College, Coimbatore, India
[2]Dept. of Computer Science (PG), Kongunadu Arts and Science College, Coimbatore, India.

[*]Corresponding Author: mereenajohny123@gmail.com

Available online at: www.ijcseonline.org

Accepted: 12/Sept/2018, Published: 30/Sept/2018

Abstract— With the explosion of the number of images in personal and on-line collections, efficient techniques for navigating, indexing, labelling and searching images become more and more important. In several studies, the representation of images by topic models in its various aspects and extend the current models. This paper aims to present a brief survey on knowledge based topic model for Unsupervised Object Discovery and Localization techniques in which the goal is to maximize the amount of work needed to re-optimize the solution when the object changes. Number of relative studies namely Latent Dirichlet allocation (LDA) with Multi-Domain Knowledge (MDK), Collaborative randomized search algorithm, Conditional random field and LDA with mixture of Dirichlet trees algorithms are discussed and evaluate the accuracy performance on the several datasets. Comparing to these algorithms the LDA with mixture of tree technique methods having better performance than other methods.

Keywords— Object discovery, object localization, topic model, and latentDirichlet allocation.

I. INTRODUCTION

Image data mining is a development potential technology for data mining which involves in multiple disciplines; it is also a challenging field which extends traditional data mining from structured data to unstructured data such as image data.

Knowledge-Based Object localization and detection is highly challenging because of intra-class variations, background clutter, and occlusions present in real-world images. While significant progress has been made in this area over the last decade, as shown by recent benchmark results [1, 2], most state-of-the-art methods still rely on strong supervision in the form of manually-annotated bounding boxes on target instances. Since those detailed annotations are expensive to acquire and also prone to unwanted biases and errors, recent work has explored the problem of weakly-supervised object discovery where instances of an object class are found in a collection of images without any box-level annotations.

Typically, weakly-supervised localization [4, 7] requires positive and negative image-level labels for a target object class. On the other hand, co-segmentation [5] and co-localization [3, 6] assume less supervision and only require the image collection to contain a single dominant object class, allowing noisy images to some degree. This paper addresses unsupervised object localization

Currently, indexing and search of images is mainly based on surrounding text, manually entered tags and/or individual and group usage patterns. However, manually entered tags have the disadvantage of being very subjective and noisy as they usually reflect the author's personal view with respect to the image content. A good example, for instance, is the tag Christmas in Flickr. Only a fraction of the images depicts the religious event as one might expect. Instead, the tag often denotes the time and date of creation. Thus thousands of vacation and party photos pop up with no real common theme. Moreover there are cases where no associated text is available for the images, as for instance many users do not label their pictures in their personal photo collection. In this survey conclude that image retrieval and indexing solely based on tags/text is difficult.

In existing studies area of content-based image retrieval deals mainly with techniques that enable searching and finding one or more images out of a possibly very large database. It can identify the following sub-areas of images retrieval with respect to their search goal [8]:

© 2018, IJCSE All Rights Reserved

Associative search: The user has no specific result image in mind when searching, only a vague idea of his/her search goal. During searching and browsing (result) images he/she interactively defines what constitutes an appropriate result image. Some examples for interactive image retrieval systems are [9].

Category search: The user searches images of a specific category. These could be scene images such as a beach during sunset, or specific object classes, for instance cats or flowers, as well as landmark images (e.g. Eiffel tower, Golden Gate Bridge).

Targeted search: The user searches for one special image. He/she has a very precise idea of how the result image has to look, e.g., he/she has already seen it before.

Most previous works in this area have only been designed and applied to relatively small and unrealistic image databases ranging from a few thousand to ten-thousands of images [12, 13]. For example the widely used COREL database consists of very clean and homogeneous images with almost perfect annotations.

There has been a significant amount of research in image retrieval systems over the last years.
Detailed overviews can be found in [8,10,and 11]. Most image search or indexing systems consist of two steps:

1. Image content analysis or recognition of all images in the database;
2. Subsequent search for similar images based on the extracted content features.

This survey paper explores the Knowledge-Based Topic Model for Unsupervised Object Discovery and Localization to process sequences of object classification to discover dominant object classes frequently present in a given image collection, and localize corresponding object instances in each image.

II. RELATED WORK

This survey works in Knowledge-Based Topic Model for Unsupervised Object Discovery and Localization which follows either image processing and image classification techniques used with machine learning algorithms on the extracted features to classify the images as object discovery and localization. Some of the existing object localization detection method is described in this section.

A. Joulin, F. Bach, and J. Ponce **(2010)**[5]discussed to combine the existing tools for bottomup image segmentation such as normalized cuts, with kernel methods commonly used in object recognition. These two sets of techniques are used within a discriminative clustering framework: the goal is to assign foreground/background labels jointly to all images, so that a supervised classifier trained with these labels leads to maximal separation of the two classes.

D. Mimno, et.al., **(2011)**[14] addressed the dimensionality reduction methods for text, such as latent Dirichlet allocation, often produce low-dimensional subspaces (topics) that are obviously flawed to human domain experts. The contributions of this techniques are threefold: (1) An analysis of the ways in which topics can be flawed; (2) an automated evaluation metric for identifying such topics that does not rely on human annotators or reference collections outside the training data; (3) a novel statistical topic model based on this metric that significantly improves topic quality in a large-scale document collection from the National Institutes of Health (NIH).

T. Deselaers, B. Alexe, and V. Ferrari**(2012)**[15] presented novel approach that can cope with extensive clutter as well as large scale and appearance variations between object instances. To make this possible to exploited generic knowledge learned beforehand from images of other classes for which location annotation is available. Generic knowledge facilitates learning any new class from weakly supervised images, because it reduces the uncertainty in the location of its object instances. Meanwhile, they proposed a conditional random field that starts from generic knowledge and then progressively adapts to the new class.

Z. Chen, et.al., **(2013)**[16] addressed to go one step further to study how the prior knowledge from other domains can be exploited to help topic modeling in the new domain. This problem setting is important from both the application and the learning perspectives because knowledge is inherently accumulative. The human beings gain knowledge gradually and use the old knowledge to help solve new problems.

M. Rubinstein, J. Kopf, C. Liu, and A. Joul**(2013)** [17] presented a new unsupervised algorithm to discover and segment out common objects from large and diverse image collections. In contrast to previous co-segmentation method algorithm performs well even in the presence of significant amounts of noise images (images not containing a common object), as typical for datasets collected from Internet search. The key insight to the algorithm is that common object patterns should be salient within each image, while being sparse with respect to smooth transformations across images. They proposed to use dense correspondences between images to capture the sparsity and visual variability of the common object over the entire database, which enables us to ignore noise objects that may be salient within their own images but do not commonly occur in others. To performed extensive numerical evaluation on established co-segmentation datasets, as well as several new datasets generated using Internet search.

© 2018, IJCSE All Rights Reserved

A. Faktor and M. Irani**(2014)** [18] discussed to define a "good image cluster" as one in which images can be easily composed (like a puzzle) using pieces from each other, while are difficult to compose from images outside the cluster. The larger and more statistically significant the pieces are, the stronger the affinity between the images. This gives rise to unsupervised discovery of very challenging image categories. To further show how multiple images can be composed from each other simultaneously and efficiently using a collaborative randomized search algorithm. This collaborative process exploits the "wisdom of crowds of images", to obtain a sparse yet meaningful set of image affinities, and in time which is almost linear in the size of the image collection.

A. Joulin, K. Tang, and L. Fei-Fei**(2014)**[19] tackled the problem of performing efficient co-localization in images and videos. Co-localization is the problem of simultaneously localizing (with bounding boxes) objects of the same class across a set of distinct images or videos. Building upon recent state-of-the-art methods, to show how they are able to naturally incorporate temporal terms and constraints for video co-localization into a quadratic programming framework. Furthermore, by leveraging the Frank-Wolfe algorithm (or conditional gradient), To showed how their optimization formulations for both images and videos can be reduced to solving a succession of simple integer programs, leading to increased efficiency in both memory and speed.

Z. Niu, G. Hua, X. Gao, and Q. Tian**(2014)**[20] addressed the problem of recognizing images with weakly annotated text tags. Most previous work either cannot be applied to the scenarios where the tags are loosely related to the images; or simply take a pre-fusion at the feature level or a post-fusion at the decision level to combine the visual and textual content. Instead, to first encode the text tags as the relations among the images, and then propose a semi-supervised relational topic model (ss-RTM) to explicitly model the image content and their relations. In such way, it can efficiently leverage the loosely related tags, and build an intermediate level representation for a collection of weakly annotated images.

C. Wang, K. Huang, W. Ren, J. Zhang, and S. Maybank**(2015)** [21] proposed the latent category learning (LCL) in large-scale cluttered conditions. LCL is an unsupervised learning method which requires only image-level class labels. Firstly, they use the latent semantic analysis with semantic object representation to learn the latent categories, which represent objects, object parts or backgrounds. Secondly, to determine which category contains the target object, they proposed a category selection strategy by evaluating each category's discrimination. Finally, they proposed the online LCL for use in large-scale conditions. Evaluation on the challenging PASCAL VOC 2007 and the large-scale ILSVRC 2013 detection datasets shows that the method can improve the annotation precision by 10% over previous methods.

M. Cho, S. Kwak, C. Schmid, and J. Ponce**(2015)** [22] addressedanunsupervised discovery and localization of dominant objects from a noisy image collection with multiple object classes. The setting of this problem is fully unsupervised, without even image-level annotations or any assumption of a single dominant class. This is far more general than typical co-localization, co-segmentation, or weakly-supervised localization tasks. To tackled the discovery and localization problem using a part-based region matching approach: To use off-the-shelf region proposals to form a set of candidate bounding boxes for objects and object parts. These regions are efficiently matched across images using a probabilistic Hough transform that evaluates the confidence for each candidate correspondence considering both appearance and spatial consistency. Dominant objects are discovered and localized by comparing the scores of candidate regions and selecting those that stand out over other regions containing them.

ZhenxingNiu, et.al.**(2018)** [23] addressed to tackle common object discovery in a fully unsupervised way. Generally, object co-localization aims at simultaneously localizing objects of the same class across a group of images. Traditional object localization/detection usually trains specific object detectors which require bounding box annotations of object instances, or at least image-level labels to indicate the presence/absence of objects in an image. Given a collection of images without any annotations, they proposed fully unsupervised method is to simultaneously discover images that contain common objects and also localize common objects in corresponding images. Without requiring knowing the total number of common objects, to formulate this unsupervised object discovery as a sub-graph mining problem from a weighted graph of object proposals, where nodes correspond to object proposals and edges represent the similarities between neighboring proposals.

ZhenxingNiu, et.al.**(2018)** [24] addressed an unsupervised discovery and localization of dominant objects from a noisy image collection with multiple object classes. The setting of this problem is fully unsupervised, without even image-level annotations or any assumption of a single dominant class. This is far more general than typical co-localization, co-segmentation, or weakly-supervised localization tasks. To tackled the discovery and localization problem using a part-based region matching approach: To use off-the-shelf region proposals to form a set of candidate bounding boxes for objects and object parts. These regions are efficiently matched across images using a probabilistic Hough transform that evaluates the confidence for each candidate correspondence considering both appearance and spatial consistency. Dominant objects are discovered and localized by comparing the scores of candidate regions and selecting those that stand out over other regions containing them.

© 2018, IJCSE All Rights Reserved

III. COMPARISON ANALYSIS

This survey paper aims to collect and consider papers that deal with Knowledge-Based Topic Model for Unsupervised Object Discovery and Localization techniques. The objective is not to undertake a conditions review, but quite to provide a broad state-of-the-art view on these related fields. Several existing approaches have been projected to assist object discovery process, distances method, which has mentioned in a body of literature that is spread over a wide variety of applications.

Table 1: SUMMARY TABLE FOR COMPARISON OF UNSUPERVISED OBJECT DISCOVERY AND LOCALIZATIONTECHNIQUES

Title	Algorithm	Key-Idea	Techniques	Limitations	Performance
Discriminative clustering for image co-segmentation (2010) [5]	Image segmentation and Object recognition.	Clustering framework for image co-segmentation.	Pre-clustering, Lowrank optimization and Manifold.	Weaklysupervised learning is not performed.	For 30 images, it takes between 4 and 9 hours.
Optimizing Semantic Coherence in Topic Models (2011) [14]	Expert-Driven Annotation Protocol.	To predict a class of low-quality topics images.	Topic Coherence and Annotation.	It does not improving the semantic quality of topics.	AUC (Area Under Curve) values predicting bad topics given coherence were 0.83 and 0.80, respectively.
Weakly Supervised Localization and Learning with Generic Knowledge (2012) [15]	Conditional random field.	Learning any new class from weakly supervised images.	Localization and Learning.	Donot learn separate models for different viewpoints of an object class from a single mixed training set.	To attain 87% accuracy.
Leveraging Multi-Domain Prior Knowledge in Topic Models (2013) [16]	LDA with Multi-Domain Knowledge (MDK)	To deal with multiple senses and add a new latent variable in LDA to model s-sets.	Collapsed Gibbs Sampling.	Topic models often do not correlate well with human judgments.	Average precision of each topic attains 85%.
"Clustering by Composition"—Unsupervised Discovery of Image Categories (2014) [18]	Collaborative randomized search algorithm.	To preserve unsupervised discovery of image categories.	Image clustering, image affinities, category discovery, unsupervised object recognition.	The region detection is in principle a hard problem even between a pair of images.	Accuracy of full search range is 89.8%.
Large-Scale Weakly Supervised Object Localization via Latent Category Learning (2015) [21]	The pipeline of latent category learning.	Semantic object representation to learn the latent categories, which represent objects, object parts or backgrounds.	Weakly supervised learning, object localization, latent semantic analysis,	To determine the number of latent categories for use in small-scale conditions.	Maximum average Selective Search is 80.30%.
Knowledge-Based Topic Model for Unsupervised Object Discovery and Localization (2018) [24]	LatentDirichlet allocation (LDA) with mixture of Dirichlet trees.	Allows to more efficiently exploiting discriminative prior knowledge from Web images.	Object discovery, object localization.	Groups of Must-Links with low confidence are assigned with smaller dimensional vector.	Object Localization on Pascal Dataset attains 87.17%

© 2018, IJCSE All Rights Reserved

IV. CONCLUSION

This paper presents brief survey about Object Discovery and Localization in Knowledge based topic models discussed with the different categories. This survey can be classified into Weak supervised and unsupervised object discovery and multi-domain knowledge representation of images by topic models in the context of retrieval on large, real-world databases and to conclude the discussion on object discovery algorithms with small and large categories. It also discussed the concept of finding relevant techniques to consider all possible datasets to localization measures which proves to be the most important criteria for image classification.

The further work enhanced and expanded for the Object Discovery and Localization technique in probabilistic Randomized Hough Transform (PRHT) with deep learning classification algorithm.

REFERENCES

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In CVPR, 2009.
[2] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007)
[3] T. Deselaers, B. Alexe, and V. Ferrari. "Localizing objects while learning their appearance". In ECCV, 2010.
[4] R. G. Cinbis, J. Verbeek, and C. Schmid. "Multi-fold MIL training for weakly supervised object localization". In CVPR, 2014.
[5] A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In CVPR, 2010.
[6] A. Joulin, K. Tang, and L. Fei-Fei. "Efficient image and video co-localization with frank-wolfe algorithm". In ECCV, 2014.
[7] M. H. Nguyen, L. Torresani, F. de la Torre, and C. Rother. Weakly supervised discriminative localization and classification: a joint learning process. In ICCV, 2009.
[8] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell., 22(12):1349–1380, 2000.
[9] A. Dong and B. Bhanu. Active concept learning for image retrieval in dynamic databases. In ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision, page 90, 2003.
[10] T. Gevers and A. W. M. Smeulders. Content-based image retrieval: an overview. In G. Medioni and S. B. Kang, editors, Emerging Topics in Computer Vision, pages 333 –384. Prentice Hall, 2004.
[11] Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues. Journal of Visual Communication and Image Representation, 10(1):39–62, 1999.
[12] C. Schmid and R. Mohr. Local greyvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19:530–535, 1997.
[13] R. Lienhart and A. Hartmann. Classifying images on the web automatically. Journal of Electronic Imaging, 11(4):445–454, 2002.
[14] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," in Proc. EMNLP, 2011, pp. 262–272.
[15] T. Deselaers, B. Alexe, and V. Ferrari, "Weakly supervised localization and learning with generic knowledge," Int. J. Comput. Vis., vol. 100, no. 3, pp. 275–293, Dec. 2012.
[16] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Leveraging multi-domain prior knowledge in topic models," in Proc. IJCAI, 2013, pp. 2071–2077.
[17] M. Rubinstein, J. Kopf, C. Liu, and A. Joulin, "Unsupervised joint object discovery and segmentation in Internet images," in Proc. CVPR, Jun. 2013, pp. 1939–1946.
[18] A. Faktor and M. Irani, "Clustering by composition'—Unsupervised discovery of image categories," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 6, pp. 1092–1106, Jun. 2014.
[19] A. Joulin, K. Tang, and L. Fei-Fei, "Efficient image and video co-localization with Frank–Wolfe algorithm," in Proc. ECCV, 2014, pp. 253–268.
[20] Z. Niu, G. Hua, X. Gao, and Q. Tian, "Semi-supervised relational topic model for weakly annotated image recognition in social media," in Proc. CVPR, Jun. 2014, pp. 4233–4240.
[21] C. Wang, K. Huang, W. Ren, J. Zhang, and S. Maybank, "Largescale weakly supervised object localization via latent category learning," IEEE Trans. Image Process., vol. 24, no. 4, pp. 1371–1385, Apr. 2015.
[22] M. Cho, S. Kwak, C. Schmid, and J. Ponce, "Unsupervised object discovery and localization in the wild: Part-based matching with bottomup region proposals," in Proc. CVPR, Jun. 2015, pp. 1201–1210.
[23] Zhenzhen Wang ; Junsong Yuan, "Simultaneously Discovering and Localizing Common Objects in Wild Images", IEEE Transactions on Image Processing (Volume: 27, Issue: 9, Sept. 2018)
[24] Zhenxing Niu, Gang Hua, Le Wang, Member, and Xinbo Gao, "Knowledge-Based Topic Model for Unsupervised Object Discovery and Localization", IEEE TRANSACTIONS on image processing, vol. 27, no. 1, january 2018

Authors Profile

Mereena Johny M.sc, Pursuing Mphil Research Degree in Kongunadu Arts and Science College at Coimbatore. She did her UG and PG degree in Kongunadu Arts and Science College at Coimbatore.

Mr. L. Haldurai has done his Master of Computer Applications and Master of Philosophy at Bharathiar University. He currently works as Assistant Professor in Department of Computer Science (PG), Kongunadu Arts and Science College. His research work focuses on Data Mining and Data Communications. He has 10 years of Teaching experience and 3 years of Research experience.