chist-era

M Ü C K E

Report on Algorithm and Prototype Evaluation

tuw**,cea*, uaic[†], bu[‡]
*CEA, LIST, LVIC
** Vienna University of Technology, ISIS, IMP
† "Al. I. Cuza" University
‡ Bilkent University
Contacts: lupu@ifs.tuwien.ac.at, adrian.popescu@cea.fr

MUCKE Project, Deliverable 5.3
10/12/2015

Abstract

Throughout the project, algorithms and their implementations have been constantly evaluated as part of the research process. The results of these evaluations have been published by the consortium partners. This deliverable provides an overview of these publications. Rather than repeating the results, we point the reader to the corresponding documents.

1 SUMMARY OF EVALUATIONS

1.1 VISUAL DATA

The use of convolutional networks for recognising concepts in noisy Web images has been extensively tested in MUCKE [20, 19]. The empirical results generated by the team at CEA show that convolutional neural networks (convnets) can obtain good generalization despite imprecise image labeling, outliers, and data biases. They also show that: image reranking algorithms are useful to improve generalization; the proposed 13-layer convnet reveals that the depth of convnet is even more helpful in improving generalization; and finally that inexpensive convnets trained on noisy Web images could be a good initialization for fine-tuning in fine-grained category classification tasks.

The team at Bilkent [18] proposed ConceptFusion as a framework for combining concept groups from many different levels and perspectives for the purpose of scene categorization. The proposed framework provides flexibility for supporting any type of concept groups, such as those that have semantic meanings like objects and attributes, or low-level features that have no meanings semantically but can provide important information about the structure of an image. There is no limit in the definition of concepts, and it is easy to be expanded through inclusion of any other intermediate representation describing the whole or part of the image in content or semantics. Additionally, Golge and Sahin [6] have also benchmarked their proposed system for face recognition, as a particular instance of concept recognition, for although the proposed method is tested for identification of faces, it is a general method that could be used for other domains as we aim to attack as our future work. The report an accuracy of 90.75%, above the existing state-of-the-art (87.1%).

1.2 TEXT AND VISUAL DATA

Since 2014, MUCKE partners have consistently benchmarked their algorithms in the MediaEval Retrieving Diverse Social Images task [5, 12, 17] and consistently demonstrated top performance, as reported in the track's overview reports [9, 8].

Additionally, the use of statistical semantics has been thoroughly evaluated by the team at TU Wien [13, 14]. Rekabsaz and colleagues explore the use of semantic similarity in text-based image retrieval in the social media domain by applying Word2Vec and Random Indexing together with two similarity methods. They ran experiments on the SemEval2014 Task 10 and the MediaEval Retrieving Diverse Social Images Task 2013/2014. Beside achieving state-of-the-art results on both datasets, they show that the similarity method has more effect on the results rather than the number of dimensions or word representation training method. In addition, by using a two- phase approach, they reduced in half the processing time of the best run while keeping precision within the boundary of statistically insignificant difference.

In addition to aspects of precision and diversity, the team at TU Wien also considered retrievability in its experiments. Sabetghadam et al. [15, 16] showed that this can be substantially increased when using explicit semantics modelled as a graph containing both image and text facets.

Explicit semantics were also tested by the team at UAIC. In 2014 their paper [7] addressed the diversification aspects that can be useful for an image retrieval system. For that we perform text processing on user query with Yago and WordNet resources and image processing in order to create clusters with similar images. The obtained results on the available collection are comparable and better in some perspectives.

The algorithms created in Iasi were also benchmarked against the Scalable Concept Image Annotation Challenge from ImageCLEF 2015 [3]. Here, for Subtask 1, the main components of the system are related to text processing (the translation of non-English words, stop-words elimination, and concept identification) and to visual processing (face recognition and body parts identification). From the presented results we can conclude that the most important component is component related to body parts identification which increased significantly our results. For Subtask 2, the main components are related to applying templates on selected concepts, based on a resource with triplets (concept1, verb, concept2). From what we see, the most important part is related to the selection of most important concepts, and from this reason the results for clean track are much better than results for noisy track.

The team in Iasi also introduced and benchmarked a method that is able to detect redundant objectives at the runtime within a multi-objective evolutionary algorithm (MOEA) [2]. By eliminating redundant objectives and working in a reduced objective space the evaluation cost is lowered and better convergence is achieved. Objective selection is performed automatically, using an efficient unsupervised clustering algorithm that is very popular in the data mining community being able to detect natural groups in data. No parameter is involved and the number of clusters does not need to be specified in advance.

Experiments showed that, after eliminating the redundant objectives, when the cardinality of the selected set of conflicting objectives is still high, reducing further the objective space by retaining fewer of the most conflicting objectives is important. Thus, the method proposed goes beyond elimination of redundant objectives and proposes a general scheme capable to auto-adapt and select at various moments during the search a small number of specific objectives such that a many-objective problem could be efficiently solved.

References

[1] *13th International Workshop on Content-Based Multimedia Indexing, CBMI 2015, Prague, Czech Republic, June 10-12, 2015*. IEEE, 2015.

[2] Mihaela Elena Breaban and Adrian Iftene. Dynamic objective sampling in many-objective optimization. In *19th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, KES 2015, Singapore, 7-9 September 2015*, pages 178–187, 2015.

[3] Alexandru Calfa, Dragos Silion, Andreea Cristina Bursuc, Cornel Paul Acatrinei, Razvan Iulian Lupu, Alexandru Eduard Cozma, Cristian Padurariu, and Adrian Iftene. Using textual and visual processing in scalable concept image annotation challenge. In Cappellato et al. [4].

[4] Linda Cappellato, Nicola Ferro, Gareth J. F. Jones, and Eric SanJuan, editors. *Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum, Toulouse, France, September 8-11, 2015*, volume 1391 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2015.

[5] Alexandru-Lucian Gînsca, Adrian Popescu, and Navid Rekabsaz. CEA list's participation at the mediaeval 2014 retrieving diverse social images task. In *Working Notes Proceedings of the MediaEval 2014 Workshop, Barcelona, Catalunya, Spain, October 16-17, 2014.*, 2014.

[6] Eren Golge and Pinar Duygulu. FAME: face association through model evolution. In *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Boston, MA, USA, June 7-12, 2015*, pages 43–49, 2015.

[7] Adrian Iftene and Lenuta Alboaie. Diversification in an image retrieval system based on text and image processing. *The Computer Science Journal of Moldova*, 22(3):339–348, 2014.

[8] Bogdan Ionescu, Alexandru-Lucian Gînsca, Bogdan Boteanu, Adrian Popescu, Mihai Lupu, and Henning Müller. Retrieving diverse social images at mediaeval 2015: Challenge, dataset and evaluation. In Larson et al. [11].

[9] Bogdan Ionescu, Adrian Popescu, Mihai Lupu, Alexandru-Lucian Gînsca, and Henning Müller. Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. In Larson et al. [10].

[10] Martha A. Larson, Bogdan Ionescu, Xavier Anguera, Maria Eskevich, Pavel Korshunov, Markus Schedl, Mohammad Soleymani, Georgios Petkos, Richard F. E. Sutcliffe, Jaeyoung Choi, and Gareth J. F. Jones, editors. *Working Notes Proceedings of the MediaEval 2014 Workshop, Barcelona, Catalunya, Spain, October 16-17, 2014*, volume 1263 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.

[11] Martha A. Larson, Bogdan Ionescu, Mats Sjöberg, Xavier Anguera, Johann Poignant, Michael Riegler, Maria Eskevich, Claudia Hauff, Richard F. E. Sutcliffe, Gareth J. F. Jones, Yi-Hsuan Yang, Mohammad Soleymani, and Symeon Papadopoulos, editors. *Working Notes Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, September 14-15, 2015*, volume 1436 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2015.

[12] João R. M. Palotti, Navid Rekabsaz, Mihai Lupu, and Allan Hanbury. TUW @ retrieving diverse social images task 2014. In Larson et al. [10].

[13] Navid Rekabsaz, Ralf Bierig, Bogdan Ionescu, Allan Hanbury, and Mihai Lupu. On the use of statistical semantics for metadata-based social image retrieval. In *13th International Workshop*

on Content-Based Multimedia Indexing, CBMI 2015, Prague, Czech Republic, June 10-12, 2015 [1], pages 1–4.

[14] Navid Rekabsaz, Ralf Bierig, Mihai Lupu, and Allan Hanbury. Toward optimized multimodal concept indexing. In Cappellato et al. [4], pages 141–152.

[15] Serwah Sabetghadam, Mihai Lupu, Ralf Bierig, and Andreas Rauber. Reachability analysis of graph modelled collections. In *Advances in Information Retrieval - 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, March 29 - April 2, 2015. Proceedings*, pages 370–381, 2015.

[16] Serwah Sabetghadam, João R. M. Palotti, Navid Rekabsaz, Mihai Lupu, and Allan Hanbury. TUW @ mediaeval 2015 retrieving diverse social images task. In Larson et al. [11].

[17] Mustafa Ilker Sarac and Pinar Duygulu. Bilkent-retina at retrieving diverse social images task of mediaeval 2014. In *Working Notes Proceedings of the MediaEval 2014 Workshop, Barcelona, Catalunya, Spain, October 16-17, 2014.*, 2014.

[18] Mustafa Ilker Sarac, Ahmet Iscen, Eren Golge, and Pinar Duygulu. Conceptfusion: A flexible scene classification framework. In *Advances in Information Retrieval - 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, March 29 - April 2, 2015. Proceedings*, pages 55–60, 2015.

[19] Phong D. Vo, Alexandru-Lucian Gînsca, Hervé Le Borgne, and Adrian Popescu. Effective training of convolutional networks using noisy web images. In *13th International Workshop on Content-Based Multimedia Indexing, CBMI 2015, Prague, Czech Republic, June 10-12, 2015* [1], pages 1–6.

[20] Phong D. Vo, Alexandru-Lucian Gînsca, Hervé Le Borgne, and Adrian Popescu. On deep representation learning from noisy web images. *CoRR*, abs/1512.04785, 2015.