INAOE's participation at ImageCLEF 2016: Text Illustration task

Luis Pellegrin*, A. Pastor López-Monroy,
Hugo Jair Escalante and Manuel Montes-Y-Gómez

Computer Science Department
Instituto Nacional de Astrofísica, Óptica y Electrónica
Tonantzintla, Puebla, 72840, México
*pellegrin@inaoep.mx

Clef2016 5-8 September, Évora, Portugal

Content

Text Illustration (TI)

- The goal of the TI task consists in finding the best image that describes a given document query.

Given a document

I want to travel. A lot. My whole life if I have to.
I want to watch beautiful sunsets and see beautiful flowers and feel the sun on my skin and the salt water in my eyes. I want to meet people.
Lots of them. Kind people. Interesting people.
I want to see the world and get lost in it. I want to look at the stars until I forget everything. I want to know what it's like elsewhere. I want to experience things for myself and decide where I want to live one day. I don't want to feel like I'm trapped in a lonely, gray town.

- m.s.

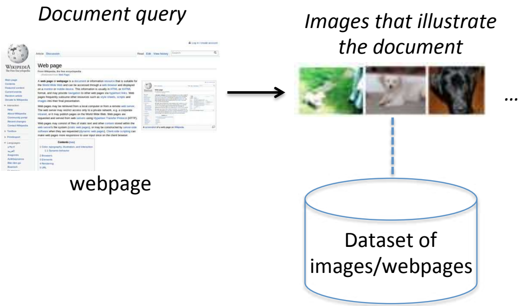
textual

Illustrating

visual

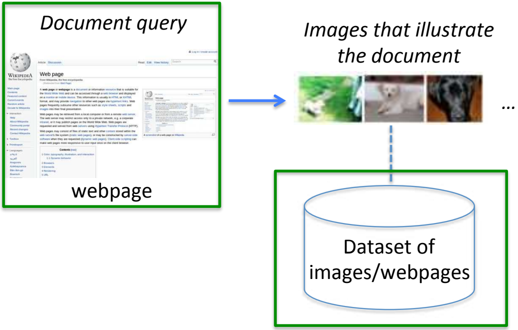
TI at ImageCLEF

- A set of images embedded in webpages are available to illustrate document queries.



TI as IR-based approach

1. **Representing documents.**
2. **Retrieval stage.**



The hypothesis is that relevant images can be derived from related webpages.

Representing documents

We use two popular representations:

1. Bag-of-words (BoW) representation defines each document as histograms of word occurrences.
2. Word2vec representation incorporates distributional semantics to text documents with learned word vectors.

Representing documents

We use two popular representations:

1. Bag-of-words (BoW) representation defines each document as histograms of word occurrences.
2. Word2vec representation incorporates distributional semantics to text documents with learned word vectors.

Retrieval stage

A document query q_j under a specific representation is used to retrieve a set of relevant items $\{(d_h, l_h) : (d_h, l_h) \in \mathcal{D}\}$.

$$similarity(q_j, d_i) = cosine(\mathbf{q}_j, \mathbf{d}_i) = \frac{\mathbf{q}_j * \mathbf{d}_i}{||\mathbf{q}_j||||\mathbf{d}_i||} \quad (1)$$

where $\mathbf{q}_j, \mathbf{d}_i$ are the representations of the document query q_j, and the i^{th} document d_i from the collection $\mathcal{D}$, respectively.

Some details about document representation

1. BoW representation:
   - It was built filtering terms with high frequency.
   - Using TF-IDF (term frequency inverse document frequency) weighting scheme.
   - Dimension of the extracted vocabulary is about 15,000.
2. Word2vec representation:
   - Each document is represented by the average word vectors (Le & Mikolov, 2014).
   - Using learned word vectors from Wikipedia (Mikolov et al., 2013).
   - Dimension of the vectors: 300 values.

Some details about document representation

1. BoW representation:
   - It was built filtering terms with high frequency.
   - Using TF-IDF (term frequency inverse document frequency) weighting scheme.
   - Dimension of the extracted vocabulary is about 15,000.
2. Word2vec representation:
   - Each document is represented by the average word vectors (Le & Mikolov, 2014).
   - Using learned word vectors from Wikipedia (Mikolov et al., 2013).
   - Dimension of the vectors: 300 values.

Quantitative results

Cuadro: Recall@K for full 180K test set.

Team	RUN	Recall (%) R@1	R@5	R@10	R@25	R@50	R@75	R@100
Baseline	chance	0.00	0.00	0.01	0.01	0.03	0.04	0.05
CEA	cbs.flickrgroup.FS.valid	0.02	0.10	0.22	0.48	0.84	1.16	1.44
	cbs.wordnet.FS.valid	0.03	0.12	0.23	0.53	0.97	1.38	1.74
	cbs.mergeA.valid	0.14	0.56	0.97	1.90	2.98	3.82	4.47
	cbs.mergeB.valid	0.18	0.63	1.05	1.97	3.00	3.87	4.51
	cbs.mergeC.valid	0.18	0.62	1.04	1.95	2.99	3.85	4.50
	wam5.kcca1.idsQueries.all.valid	0.11	0.36	0.62	1.11	1.68	2.11	2.47
	warm7.idsQueries.10BWS.all.valid	0.18	0.63	1.07	1.93	2.93	3.69	4.33
INAOE	**run1.bow+tfidf.thr5p**	**28.75**	**63.50**	**75.48**	**84.39**	**86.79**	**87.36**	**87.59**
	run2.d2v.thr5p	2.57	5.65	7.71	11.76	16.69	20.34	23.40
	run3.d2v+tfidf.thr5p	3.68	7.73	10.46	15.62	21.36	25.48	28.78

Qualitative results: short texts

doc1.txt

cardiothoracicsurgery.org/about Bottom,Top,Right1 '); } Resolution: standard / high Figure 1. Severe aortic insufficiency seen on transthoracic echocardiography during catheterization . Severe aortic insufficiency is noted through the area of the left coronary cusp. Aortic insufficiency is terminated by placement of catheter into left coronary cusp. Martin et al. Journal of Cardiothoracic Surgery 2010 5 :94 doi:10.1186/1749-8090-5-94 Download authors' original image

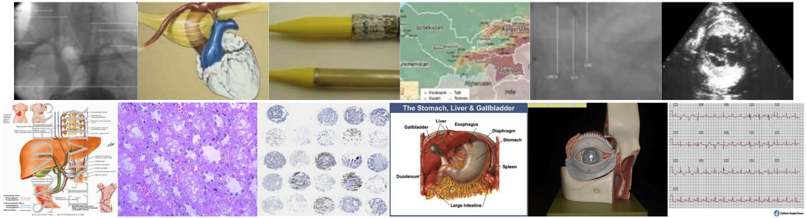
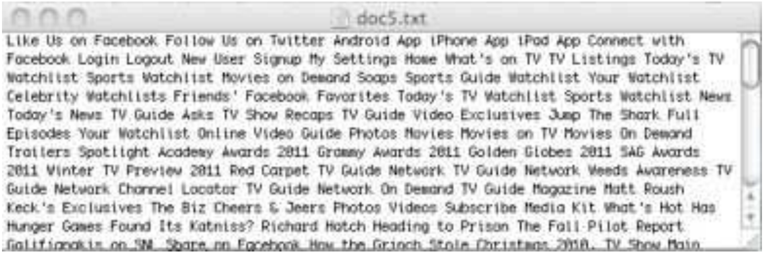


Figura: Given the text document (top), the top of retrieved images. First row, output images from BoW. Second row, output images from word2vec.

Qualitative results: long texts


docs5.txt

Like Us on Facebook Follow Us on Twitter Android App iPhone App iPod App Connect with Facebook Login Logout New User Signup My Settings Home What's on TV TV Listings Today's TV Watchlist Sports Watchlist Movies on Demand Soaps Sports Guide Watchlist Your Watchlist Celebrity Watchlists Friends' Facebook Favorites Today's TV Watchlist Sports Watchlist News Today's News TV Guide Asks TV Show Recaps TV Guide Video Exclusives Jump The Shark Full Episodes Your Watchlist Online Video Guide Photos Movies Movies on TV Movies On Demand Trailers Spotlight Academy Awards 2011 Grammy Awards 2011 Golden Globes 2011 SAG Awards 2011 Winter TV Preview 2011 Red Carpet TV Guide Network TV Guide Network Weeds Awareness TV Guide Network Channel Locator TV Guide Network On Demand TV Guide Magazine Matt Roush Keck's Exclusives The Biz Cheers & Jeers Photos Videos Subscribe Media Kit What's Hot Has Hunger Games Found Its Katniss? Richard Hatch Heading to Prison The Fall Pilot Report

Figura: Given the text document (top), the top of retrieved images. First row, output images from BoW. Second row, output images from word2vec.

Conclusions

- According to the performed evaluation we conclude that, relevant images can be derived from related webpages, so it is possible to use IR techniques.
- BoW representation obtained outstanding performance and was more robust to deal with long documents.
- Finally, the performed experiments under different representations give an initial point of comparison for future approaches.
- As future work, we plan to incorporate mix information (textual-visual).

Conclusions

- According to the performed evaluation we conclude that, relevant images can be derived from related webpages, so it is possible to use IR techniques.
- BoW representation obtained outstanding performance and was more robust to deal with long documents.
- Finally, the performed experiments under different representations give an initial point of comparison for future approaches.
- As future work, we plan to incorporate mix information (textual-visual).

Gilbert A., Piras L., Wang J., Yan F., Ramisa A., Dellandrea E., Gaizauskas R., Villegas M., Mikolajczyk K. (2016)
Overview of the ImageCLEF 2016 Scalable Concept Image Annotation Task.
In: *CLEF2016 Working Notes*, CEUR Workshop Proceedings, Évora, Portugal.

Mikolov T., Chen K., Corrado G., Dean J. (2013)
Efficient estimation of word representations in vector space.
In: *CoRR*, abs/1301.3781.

Jones K.S. (1972)
A statistical interpretation of term specificity and its application in retrieval.
In: *Journal of Documentation 28*, 11-21.

Le Q.V., and Mikolov T. (2014)
Distributed representations of sentences and documents.
In: *CoRR*, abs/1405.4053.

Hodosh M., Young P., Hockenmaier J. (2013)
Framing image description as a ranking task: Data, models and evaluation metrics.
In: *Journal Artif. Int. Res. 47*, 853-899.

Pellegrin, L., Vanegas, J.A., Arevalo, J., Beltrán, V., Escalante, H.J., Montes-Y- Gómez, M., González, F. (2015)
INAOE-UNAL at ImageCLEF 2015: Scalable Concept Image Annotation.
In: *CLEF2015 Working Notes*, CEUR Workshop Proceedings, Toulouse, France.

Thank you for your attention, questions?

pellegrin@inaoep.mx