
Additional Material for the paper: Documents as multiple overlapping windows into grids of counts

1 Recipe Grid

In `AdditionalGrid.pdf` we reported the grid learned from the corpus of recipes that we have used as main example throughout all the paper (e.g., Fig. 1 and Fig. 4).

2 Additional Results

In `AdditionalResults.pdf` we reported the classification accuracy of CCG and LDA on 3 new Datasets. We used the same procedure described in Sec. 3-“Document Classification”.

The first two plates (a-b), are relative to the `Different` and `Similar` subsets analyzed in the main text. Similarly to Fig. 3a of the main paper, on the x-axis we have the different model size, in term of capacity κ , whereas in the y-axis we reported the accuracy. The same κ can be obtained with different choices of **E** and **W** therefore we represented the grid size **E** using gray levels, the lighter the marker the bigger the grid. The capacity κ is roughly equivalent to the number of LDA topics as it represents the number of independent windows that can be fit in the grid and we compared the with LDA using this parallelism.

In the bottom plate, we have also considered the UIUC Sports dataset [1]. This computer vision dataset contains 8 classes of images each one relative to a different sport event. As words z we extracted SIFT features and we quantized them in $Z = 200$ visual words. This dataset is particularly challenging as composing elements and objects must be identified in order to correctly classify the sport event. The rationale here is that different classes share some elements, like “water” for sailing and rowing classes, but they also will have peculiar elements that distinguish them.

3 Videos

We uploaded some video that presents a simple interface on youtube (anonymousNips channel). We could not compress the videos to fit the 10Mb limit due to the presence of text: high compression made the fonts unreadable.

In the videos we considered a new text corpus composed by all the Science Paper (1601 Reports or Research articles) published in 2011 and 2012. We extracted the words of each document from title, abstract, author names and references (we ignored the full text); this process resulted in a vocabulary size of 11038 unique words and a total of $\sim 2K$ tokens.

1. The interface is zoomable and pannable <http://www.youtube.com/watch?v=YUPj4iIy6zU>
2. Words are clickable http://www.youtube.com/watch?v=oIvwC_6JwQ0
3. Search for the word “Memory” <http://www.youtube.com/watch?v=ij2-XYUmhSI>

4. Refinement of a search by using additional search terms http://www.youtube.com/watch?v=SmwDHrC-_tE
5. Search for the word “Weather” <http://www.youtube.com/watch?v=60Yon4P-KYk>
6. Search for the word “Forest” <http://www.youtube.com/watch?v=oqM6uEDkqq8>
7. Words shown after a search changes based on the co-occurrence with the search term <http://www.youtube.com/watch?v=FF2a9WX1CrQ>

References

- [1] jia Li, L.: What, where and who? classifying event by scene and object recognition. In: In IEEE International Conference on Computer Vision. (2007)