sorodoc_2017_multimodal_topic_labelling

Year

2017

Author(s)

Sorodoc, Ionut and Lau, Jey Han and Aletras, Nikolaos and Baldwin, Timothy

Title

Multimodal Topic Labelling

Venue

EACL

Topic labeling

Fully automated

Focus

Primary

Type of contribution

Novel

Underlying technique

Deep Neural Network

Topic labeling parameters

Nr. of folds (k-fold cross-validation): 10
Epochs: 10

Label generation

Baseline:

That is, we generate and rank textual and image labels based on Bhatia et al., 2016 and `aletras_2017_labeling_topics_with_images_using_a_neural_network` respectively, and then generate a combined ranking based on the predicted ratings. The baseline model views the two modalities (image and textual labelling) as two distinct tasks and does not leverage potential complementarity between them.

joint-NN

We propose a simple feed-forward neural that jointly re-ranks the two topic label modalities.

In joint-NN, we first generate the candidate image labels and textual labels using the methodologies of `aletras_2017_labeling_topics_with_images_using_a_neural_network` and Bhatia et al., 2016, respectively.

However, unlike baseline where the labels are ranked separately, joint-NN feeds both label modalities into a single network to predict their ratings.

Each input modality is fed into two dense layers that are unconnected.

The hidden representation at the 4th layer of the networks is then passed to a joint/shared hidden layer before the final output layer.

All connections between layers are dense connections and the final output layer has an sigmoid activation, while all other hidden layers have ReLU activations.

The first four layers are kept separate to allow the network to transform the embeddings from the two different modalities to a common hidden representation.

The shared layers leverage potential complementarity between the two label modalities to predict the final label rating.

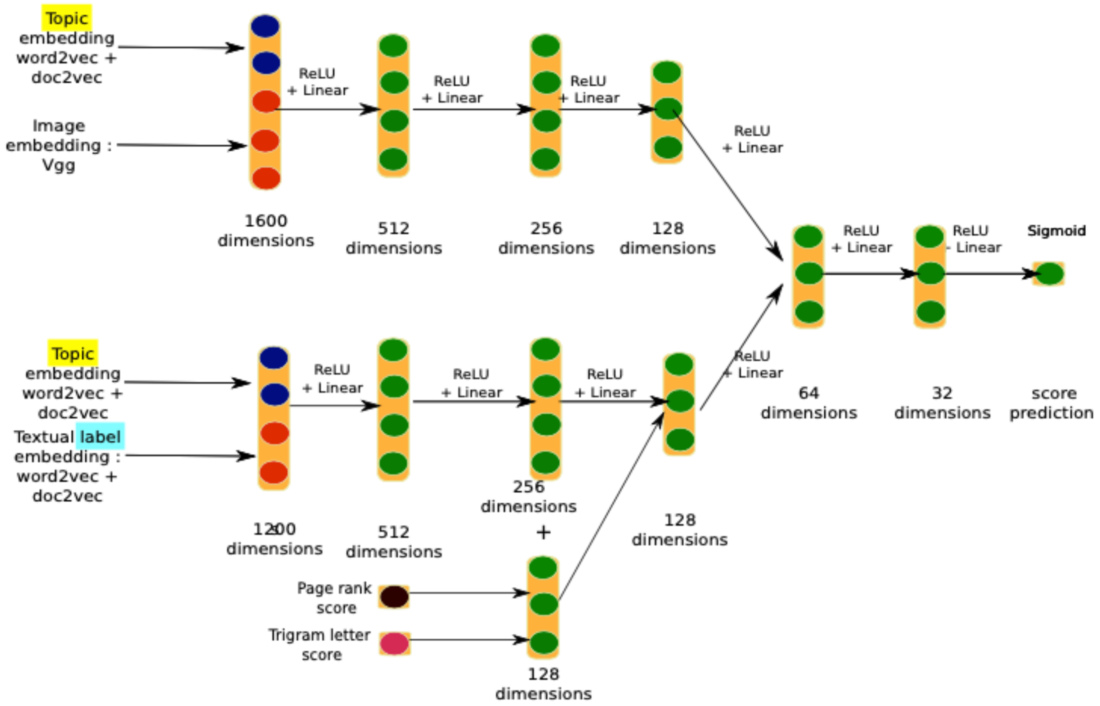


Figure 1: Multimodal model for topic labelling (joint-NN)

Generation of textual labels

Following the label generation methodology of Bhatia et al., 2016, as part of which, the labels and topic terms each have representations based on doc2vec and word2vec embeddings, respectively.

We concatenate all four embeddings and use them as the input for the network.

Bhatia et al., 2016 found that letter trigram features and PageRank features were strong features when re-ranking the labels.

We borrow this idea, and incorporate these two features into the network by mapping the 2-dimensional input (representing the letter trigram and PageRank features) into a 128-dimension vector and concatenating it with the 256-dimension hidden representation at the third layer (thus yielding a 384-dimension vector).

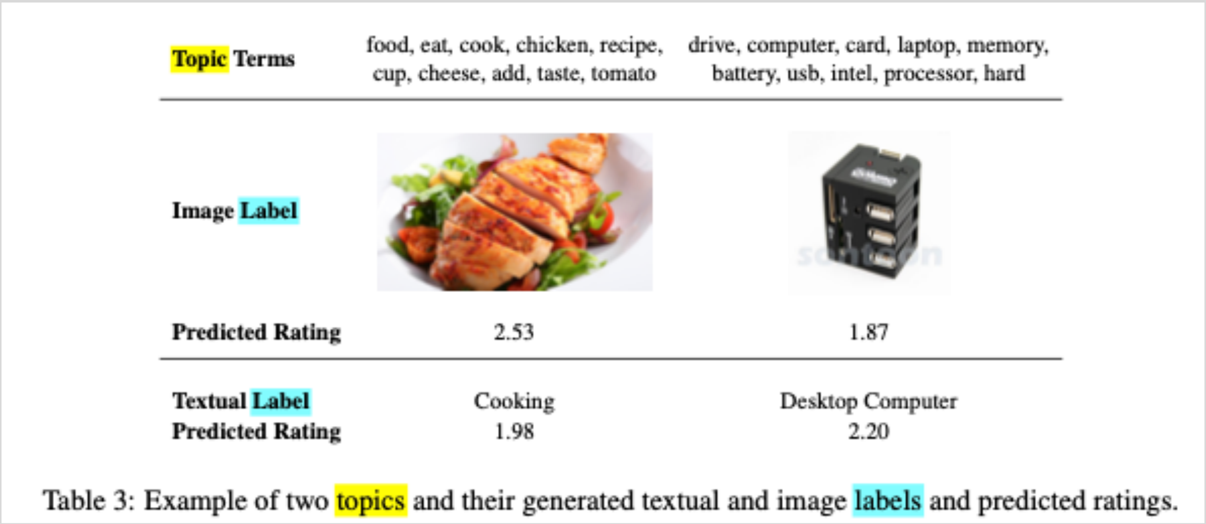
Generation of visual labels

The topic terms use the same doc2vec and word2vec embeddings.

For the image labels, we use the representation of the last layer of the VGG Neural Network (Simonyan and Zisserman, 2014).

As before, the vectors for the topic terms and image labels are concatenated and fed as input to the network.

disjoint-NN
As a control to test whether the sharing of weights helps with the prediction of label ratings, we experiment with another network that has the same architecture as joint-NN, except that the final few layers are not shared and the two networks are trained independently.



Topic Terms	food, eat, cook, chicken, recipe, cup, cheese, add, taste, tomato	drive, computer, card, laptop, memory, battery, usb, intel, processor, hard
Image Label		
Predicted Rating	2.53	1.87
Textual Label	Cooking	Desktop Computer
Predicted Rating	1.98	2.20

Table 3: Example of two topics and their generated textual and image labels and predicted ratings.

Motivation

Reducing the cognitive load of end-users when interpreting topics.

Topic modeling

LDA

Topic modeling parameters

Nr of topics (K): 100

Nr. of topics

45, 38, 60, 85 topics for BLOGS, BOOKS, NEWS and PUBMED, respectively after filtering incoherent topics

Label

An image and textual label starting from a given topic, both associated with their respective predicted ratings

Label selection

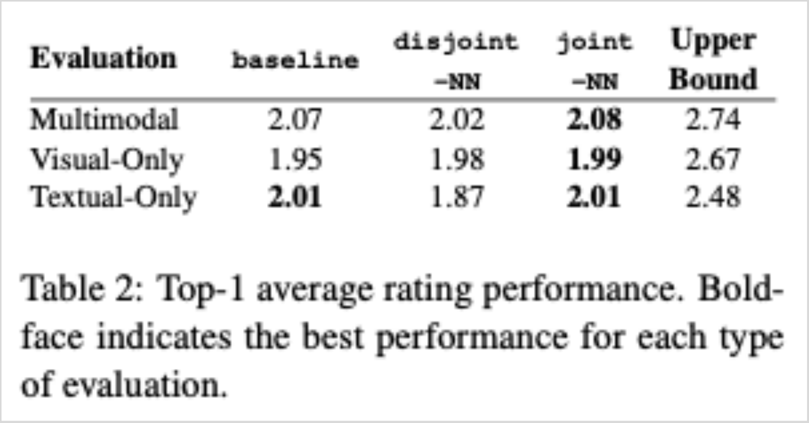
\

Label quality evaluation

“top-1 average rating” is used as the evaluation metric.
It computes the mean rating of the top-ranked label generated by the system, and provides an assessment of the absolute utility of the labels.
For example, if the top-ranked label predicted by the system has an aver- age rating of 3.0, that means the system are generating perfect topic labels.

In addition to the 3 systems, for each topic we determine the rating of the best label and compute its mean over all topics, as the upper bound for the task (labelled “upper bound”).

Comparison models



Evaluation	baseline	disjoint -NN	joint -NN	Upper Bound
Multimodal	2.07	2.02	**2.08**	2.74
Visual-Only	1.95	1.98	**1.99**	2.67
Textual-Only	**2.01**	1.87	**2.01**	2.48

Table 2: Top-1 average rating performance. Boldface indicates the best performance for each type of evaluation.

Assessors

\

Domain

Domain (paper): Topic labeling

Domain (dataset): BLOGS, BOOKS, NEWS and PUBMED

Problem statement

Proposing a multimodal approach to topic labelling using a simple feedforward neural network.

Given a topic and a candidate image or textual label, the model automatically generates a rating for the label, relative to the topic.

Developing a topic labelling dataset with manually-scored image and text labels for a diverse set of topics

Corpus

Origin: Bhatia et al., 2016

Content:

- Before pre-processing: 228 topics, each with 19 textual labels
- After pre-processing: 228 topics, each with 19 textual labels and 20 image labels
- Total: 4560 images and 4332 textual labels for 228 topics.

Details:

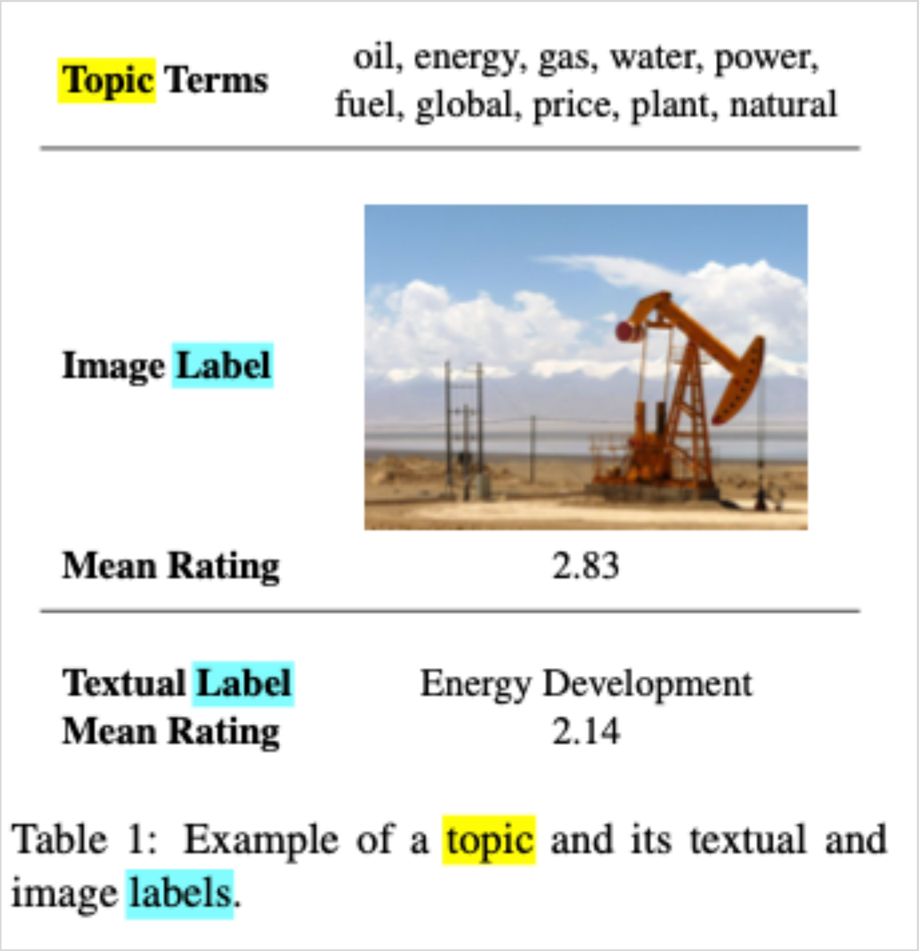
- 4 different domains: BLOGS, BOOKS, NEWS and PUBMED

Document

A topic associated with 19 textual labels which were rated by human judges on a scale of 0–3, where 0 represents a poor label and 3 indicates a perfect label.

Image labels are generated for each topic (see `pre-processing`)

An example of a topic and its image and textual labels, and their associated mean ratings:



Topic Terms	oil, energy, gas, water, power, fuel, global, price, plant, natural
Image Label	
Mean Rating	2.83
Textual Label	Energy Development
Mean Rating	2.14

Table 1: Example of a topic and its textual and image labels.

Pre-processing

20 Image labels are generated for each topic following the method of Aletras and Stevenson, 2013

- Following the annotation approach of Bhatia et al., 2016, ratings based on an ordinal scale of 0–3 are collected.
- Amazon Mechanical Turk is used to crowdsource the ratings (353 total participations)
    - Each image is labelled by 8 workers
    - For quality control, bad labels are embedded into the HIT for each topic by sampling a label candidate for a topic from a different domain
        - Workers who rate these control labels greater than 1 are recorded, and those who fail more than 50% of control labels are filtered out of the dataset.
        - A total of 42 participants were filtered out, on the basis of having an error rate of more than 50% (based on the control images).
- To aggregate the ratings for a label, its mean rating is computed

```
@inproceedings{sorodoc_2017_multimodal_topic_labelling,
    title = "Multimodal Topic Labelling",
    author = "Sorodoc, Ionut  and
      Lau, Jey Han  and
      Aletras, Nikolaos  and
      Baldwin, Timothy",
    booktitle = "Proceedings of the 15th Conference of the {E}uropean Chapter
of the Association for Computational Linguistics: Volume 2, Short Papers",
    month = apr,
    year = "2017",
    address = "Valencia, Spain",
    publisher = "Association for Computational Linguistics",
    url = "https://aclanthology.org/E17-2111",
    pages = "701--706",
    abstract = "Topics generated by topic models are typically presented as a
list of topic terms. Automatic topic labelling is the task of generating a
succinct label that summarises the theme or subject of a topic, with the
intention of reducing the cognitive load of end-users when interpreting these
topics. Traditionally, topic label systems focus on a single label modality,
e.g. textual labels. In this work we propose a multimodal approach to topic
labelling using a simple feedforward neural network. Given a topic and a
candidate image or textual label, our method automatically generates a rating
for the label, relative to the topic. Experiments show that this multimodal
approach outperforms single-modality topic labelling systems.",
}
```

#Thesis/Papers/Initial