PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/199472

Please be advised that this information was generated on 2019-09-19 and may be subject to change.
Exploiting Local Semantic Concepts for Flooding-related Social Image Classification
Zhengyu Zhao, Martha Larson, Nelleke Oostdijk
Radboud University, Netherlands
z.zhao@cs.ru.nl, m.larson@cs.ru.nl, n.oostdijk@let.ru.nl

ABSTRACT
In this paper, we present an approach to identification of the images that depict passable and non-passable roads, from a collection of flood-related tweet images. Our key insight is that the local information from domain-specific concepts ('boat', 'person' and 'car') can be exploited to help determine whether an image depicts a location that is passable. We use concept detection as the basis for features that encode local information. We use conventional features, i.e., presence of concepts and visual features extracted from the concept region, but also a novel light-weight feature, i.e., the aspect ratio of the bounding box. Experimental results show that integrating local semantic information yields slightly better performance than only using image-level CNN representation. Text features are not competitive.

1 INTRODUCTION
Despite achieving impressive performance in various visual recognition tasks, convolutional neural network (CNN) representations do not fully capture local-level discriminative information when only trained at a single scale, i.e., input size of 224x224 for most conventional CNNs. In order to complement global CNN features, recent work on fine-grained object classification [5, 6, 8, 12] and scene recognition [2, 9–11] has also tried to exploit discriminative information from local semantic regions. Building on these insights, here, we demonstrate that the task of differentiating two road conditions (passable vs. non-passable) [1] will also benefit from local semantic information. Our starting point is the observation that images with similar global appearance have differentiable local patterns, as shown in Figure 1. Intuitively, we consider that three specific concepts ('boat', 'person' and 'car') will show different properties in the context of road passability. Moreover, based on our exploratory experiments, we observed that the images containing the three concept classes account for a large proportion (46%) of the passability-relevant images. As shown in Figure 2, the images with these three concepts span over the entire passability-relevant dev-set without any specific bias related to time order, which is reflected by the numerical order of the tweet ID. These two observations indicate that using local information from these concepts is not accidental but can be generally applicable.

2 APPROACH
We start with a light-weight approach by only using text information. By manual inspection of the patterns in the dev-set, we created a set of rules that apply to a vocabulary that has been annotated with basic part-of-speech and semantic-word class information. On the basis of these rules, we create a set of ngrams, which represents strings of lexical items that we would expect to occur in tweets related to road passability. Whenever any created ngram is encountered in the text, the associated class label is assigned (either passable or non-passable). As we target mostly texts indicating that roads are not passable, there are only few ngrams that yield the label passable. In the case of no matching, the image will be regarded not relevant to road passability.

For the visual-based approach, the basic pipeline is hierarchical classification with two SVM classifiers. The first classifier is applied to differentiate the images that are relevant to road passability from the others. Here we only use image-level features extracted from a ResNet50-based CNN model, which is pre-trained on the large-scale
scene-centric database Places2 [13]. Exploratory experiments on the dev-set showed that this option performed better than using the object-centric ImageNet [3] as pre-training data. Then, the second classifier will further predict the images that have been classified as relevant into passable or non-passable classes. Here, we use both the Places2 and ImageNet as the pre-training data, resulting in better performance than using only one of them. This result suggests that discriminative information from scene-level and object-level will complement each other for differentiating passable vs. non-passable images.

Alternatively, we add a pre-filtering step before the second classifier that allows test images containing the three concepts to be treated differently. We adopt the state-of-the-art YOLOv3 [7], which is pre-trained on the union of VOC2007 and VOC2012 trainval set [4], for automatic concept detection. In order to capture differences accurately, we exclude the image candidates that have incomplete bounding boxes in the image area, or a confidence score below 0.9. When multiple instances are detected in one image, we use the average values of their features as the final feature.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 passable images.

The subtle differences in local information can also be encoded by a single value derived from concept bounding boxes. Specifically, we look at the height-width aspect ratio of the bounding box, or visual features derived from the bounding box. From the analysis of the text-based approach, we concluded that the text information might be useful if we would use deep visual features extracted from the bounding box region of ‘car’ to train a SVM classifier for pre-filtering. Note that no local features for ‘boat’ and ‘car’ are used for this run.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 non-passable images.

The subtle differences in local information can also be encoded by a single value derived from concept bounding boxes. Specifically, we look at the height-width aspect ratio of the bounding box, or visual features derived from the bounding box. From the analysis of the text-based approach, we concluded that the text information might be useful if we would use deep visual features extracted from the bounding box region of ‘car’ to train a SVM classifier for pre-filtering. Note that no local features for ‘boat’ and ‘car’ are used for this run.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 non-passable images.

The subtle differences in local information can also be encoded by a single value derived from concept bounding boxes. Specifically, we look at the height-width aspect ratio of the bounding box, or visual features derived from the bounding box. From the analysis of the text-based approach, we concluded that the text information might be useful if we would use deep visual features extracted from the bounding box region of ‘car’ to train a SVM classifier for pre-filtering. Note that no local features for ‘boat’ and ‘car’ are used for this run.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 non-passable images.

The subtle differences in local information can also be encoded by a single value derived from concept bounding boxes. Specifically, we look at the height-width aspect ratio of the bounding box, or visual features derived from the bounding box. From the analysis of the text-based approach, we concluded that the text information might be useful if we would use deep visual features extracted from the bounding box region of ‘car’ to train a SVM classifier for pre-filtering. Note that no local features for ‘boat’ and ‘car’ are used for this run.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 non-passable images.

The subtle differences in local information can also be encoded by a single value derived from concept bounding boxes. Specifically, we look at the height-width aspect ratio of the bounding box, or visual features derived from the bounding box. From the analysis of the text-based approach, we concluded that the text information might be useful if we would use deep visual features extracted from the bounding box region of ‘car’ to train a SVM classifier for pre-filtering. Note that no local features for ‘boat’ and ‘car’ are used for this run.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 non-passable images.

The subtle differences in local information can also be encoded by a single value derived from concept bounding boxes. Specifically, we look at the height-width aspect ratio of the bounding box, or visual features derived from the bounding box. From the analysis of the text-based approach, we concluded that the text information might be useful if we would use deep visual features extracted from the bounding box region of ‘car’ to train a SVM classifier for pre-filtering. Note that no local features for ‘boat’ and ‘car’ are used for this run.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 non-passable images.

The subtle differences in local information can also be encoded by a single value derived from concept bounding boxes. Specifically, we look at the height-width aspect ratio of the bounding box, or visual features derived from the bounding box. From the analysis of the text-based approach, we concluded that the text information might be useful if we would use deep visual features extracted from the bounding box region of ‘car’ to train a SVM classifier for pre-filtering. Note that no local features for ‘boat’ and ‘car’ are used for this run.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 non-passable images.

The subtle differences in local information can also be encoded by a single value derived from concept bounding boxes. Specifically, we look at the height-width aspect ratio of the bounding box, or visual features derived from the bounding box. From the analysis of the text-based approach, we concluded that the text information might be useful if we would use deep visual features extracted from the bounding box region of ‘car’ to train a SVM classifier for pre-filtering. Note that no local features for ‘boat’ and ‘car’ are used for this run.

Since ‘boat’ is not a conventional means to pass a road, the presence of any boat in the image indicates the road is very likely to be non-passable. So we use the +/- presence of ‘boat’ in an image as a feature. The experiments on the dev-set show that boats can be detected in 46 of 1179 non-passable images, and only in 5 of 951 non-passable images.
REFERENCES


