Title: A spatio-temporal map of reading processes in the brain

Author keywords:
Magnetoencephalography
Conditional Independence Testing
Naturalistic experiment
Reading
Language Processing

Abstract:

Reading is a fast and complex process that requires us to perceive incoming words and gradually integrate them into a representation of sentence meaning. As words are read, it takes about 100ms for the visual input to reach the visual cortex. 50ms later, the visual input is processed as letter strings in a specialized region of the left visual cortex. We know that between 200-600ms, that the word's semantic properties are processed. Less is understood about the cortical dynamics of word integration, as multiple theories exist that contradict each other.

To shed light on the cortical dynamics of reading, we use in this work a natural reading paradigm in which subjects read an unmodified chapter of a book while their brain activity is recorded using Magnetoencephalography (offering recordings with great temporal resolution). We also use various natural language processing tools to model the content of the text (including Recurrent Neural Network Language Models, Vector Space Models representing semantic properties, automated parsers used to represent syntactic structures etc.). Finally, we build a predictive model that expresses the brain activity related to reading any segment of text as a function of the properties of that text. We manipulate this model in a specific manner that allows us to perform an indirect conditional independence test, telling us which part of the signal is related to a specific word property when taking into account other properties.

The final result is a novel, detailed spatiotemporal map of where and when in the brain different properties of words (semantics, part of speech, grammatical function in a sentence...) are represented as the brain reads. We show the progressive perception of each word in a posterior to anterior fashion. For each region along this pathway we show a differentiation of the word properties that best explain the activity. Our results are consistent with some broad lines of the classical reading models, however, most interestingly, they show that the perception of the syntactic and semantic features of a word happen earlier than previously reported. We hypothesise that this is because classical reading experiments operate by breaking a semantic or syntactic expectation in order to detect semantic or syntactic processing, respectively. Brain processes involved in anomaly detection are bound to be slower than the initial perception of a word. We therefore advocate that using naturalistic paradigms such as ours, that do not rely on breaking expectations but model perception directly, offer results that are more representative of how the brain reads normally.

Title: What's the point: Semantic segmentation with point supervision

Author keywords: semantic segmentation
weak supervision
deep learning

Abstract:

At the forefront of visual recognition is the question of how to most effectively teach computers about new visual concepts. Algorithms trained from large-scale carefully annotated data enjoy better performance than their weakly supervised counterparts; however, obtaining strongly supervised data is very expensive.It is particularly difficult to collect training data for semantic segmentation, i.e. the task of assigning a class label to every pixel in the image. Detailed per-pixel annotations enable training accurate models but are very expensive to obtain; image-level class labels are an order of magnitude cheaper but result in less accurate models. In this work, we take a natural step towards stronger supervision for semantic segmentation at negligible additional cost, compared to image-level labels. The most natural way for humans to refer to an object is by pointing: “That cat over there (point)” or “What is that over there? (point)”

We demonstrate that simply pointing once to target objects in training images can be a surprisingly effective means of supervision. We extend a state-of-the-art convolutional neural network (CNN) framework for semantic segmentation [1] to incorporate point supervision in its training loss function. With just one annotated point per object class, we are able to considerably improve semantic segmentation accuracy.

One lingering concern with supervision at the point level is that it is difficult to infer the full extent of the object. To overcome this issue, we additionally modify the training loss function to incorporate a generic objectness prior [2] which enables learning the full object extent. Objectness helps separate objects (e.g., car, sheep, bird) from background (e.g., grass, sky, water), by providing a probability that a pixel belongs to an object. To the best of our knowledge we are the first to employ this directly in the loss to guide the training of a CNN.

Contributions: Our primary contribution is introducing a novel supervision regime for semantic segmentation based on humans pointing to objects. This supervision is (1) cheap to obtain, and (2) significantly improves segmentation accuracy. Our secondary contribution is improving the accuracy of weakly supervised segmentation by using an objectness prior directly for training a segmentation CNN. The combined effect of our contributions is a substantial increase of 12.9% mean intersection over union on the PASCAL VOC 2012 dataset compared to training with image-level labels.

Full text of this work is available [3] and additional details are at http://ai.stanford.edu/~olga/whats_the_point.html

References:

[1] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015.
[2] B. Alexe, T. Deselares, and V. Ferrari. Measuring the objectness of image windows. In PAMI, 2012.
[3] O. Russakovsky*, A.L. Bearman*, V. Ferrari, L. Fei-Fei. What’s the point: semantic segmentation with point supervision. ArXiv 1506.02106, 2015.

Title: Deep Graph Kernels

Author keywords: graph kernels
social networks
graphs
graph similarity
neural networks

Abstract: In domains such as social networks, bioinformatics, chemoinformatics and robotics, we are often interested in computing similarities between structured objects. Graphs, including sequences and trees as special cases, offer a natural way to represent structured data. To illustrate one example where graph similarity can be useful, consider the problem of identifying a sub-community (also referred as subreddits) on Reddit. To tackle this problem, one can represent an online discussion thread as a graph where nodes represent users, and edges represent whether two users interact, for instance, by responding to each other's comments.

Then, the task is to predict which sub-community a discussion thread belongs to based on its communication graph. Similarly, in bioinformatics, one might be interested in the problem of identifying whether a given protein is an enzyme or not. In this case, the secondary structure of a protein is represented as a graph where nodes correspond to atoms and edges represent the chemical bonds between atoms. If the graph structure of the protein is similar to known enzymes, one can conclude that the given graph is also an enzyme. Therefore, computing semantically meaningful similarities between graphs is an important problem in various domains.

One of the increasingly popular approaches to measure the similarity between structured objects is to use kernel methods. Roughly speaking, kernel methods measure the similarity between two objects with a kernel function which corresponds to an inner product in reproducing kernel Hilbert space (RKHS). The challenge for kernel methods is then to find a suitable kernel function that captures the semantics of the structure while being computationally tractable. R-convolution is a general framework for handling discrete objects where the key idea is to recursively decompose structured objects into "atomic" sub-structures and define valid local kernels between them. Several graph kernels proposed to compute the similarity between graphs by decomposing them into various sub-structures.

In this work, we present Deep Graph Kernels, a unified framework to learn latent representations of sub-structures for graphs, inspired by latest advancements in language modeling and deep learning. Our framework leverages the dependency information between sub-structures by learning their latent representations. We demonstrate instances of our framework on three popular graph kernels, namely Graphlet kernels, Weisfeiler-Lehman subtree kernels, and Shortest-Path graph kernels. Our experiments on several benchmark datasets show that Deep Graph Kernels achieve significant improvements in classification accuracy over state-of-the-art graph kernels.

Note: This work is presented at KDD 2015 as a full research paper.