Structural Learning for Web Design

Maxine Lim, Cesar Torres, Arvind Satyanarayan

Design is difficult to quantify.

How can we better encode human-driven design in data representations?

THE WEB DESIGN DATA

1713 dimensions for each visual block
DOM-based (e.g. tree level, numChildren)
CSS (e.g. font, color, size)
Computer Vision (e.g. GIST)

CROWDSOURCED LABELS

style:

- 365 classes (e.g. clean or modern)
- 3292 pages

structural semantic:

- 416 classes (e.g. header, nav-bar)
- 13319 page-node pairs

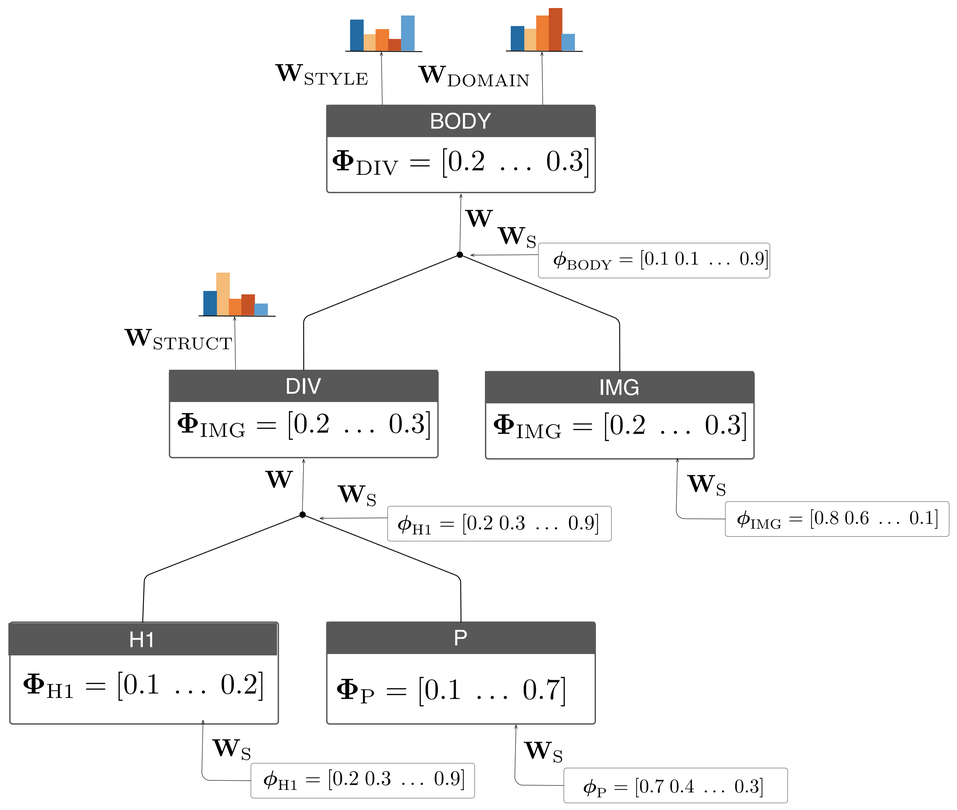
OUR APPROACH

Socher et. al (2011) used Recursive Neural Networks (RNNs) for structured prediction in domains such as language and image processing. **Web pages already possess explicit hierarchical structure** in their underlying Document Object Models (DOMs).

Our method differs from Socher et. al. by directly leveraging this structure as part of our 1713-dimension feature vector -- raw features at the each level of the RNN augment those from their children during learning.

To train the RNN, we learn softmax classifiers for Wsemantic, W, and Wlabel by running batch gradient descent with a feed-forward step, which absorbs children feature vectors. It is followed by back-propagation through the tree structure, which calculates the contribution of each neuron to the error.

Recursive Neural Network (RNN)



DISCUSSION

Adjusting the parameters of the learning algorithm can have a significant impact on the results. We tweaked the number of hidden dimensions (the dimensionality of the RNN), in increments of 50 starting at 50, and the regularization constant, by orders of magnitude starting at 0.0001. 100 hidden dimensions and a regularization of 0.01 gave us the greatest test accuracy for structural labels.

However, test accuracy for style labels remains low (~40%). We suspect that this is due to two reasons: while a node can only have one structural label, a page could have many style labels. As a result, while softmax is a good technique for predicting structural labels, we believe it "smears" out the probabilities for style labels. Thus, we are now investigating the effects of using multiple, independent logistic regression classifiers (one per style label). Additionally, as compared to the structural ground truth labels, the style labels are far more sparse. We are conducting another round of crowdsourced label collection to increase this density.

RESULTS

Our accuracy metric judged whether a ground truth structural label was in the top 5 predicted labels.

Training Accuracy

Softmax Only (numHid=100, reg=0.01)	RNN (numHid=100, reg=0.01)
86%	90%

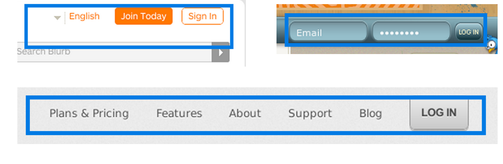
Test Accuracy

Softmax Only (numHid=100, reg=0.01)	RNN (numHid=100, reg=0.01)
50%	58%

Correctly predicted 'advertisement'

Incorrectly predicted 'advertisement'

Correctly predicted 'login'



Incorrectly predicted 'login'