

Dynamic Routing Between Capsules

Authors: Sara Sabour, Nicholas Frosst, Geoffrey E Hinton

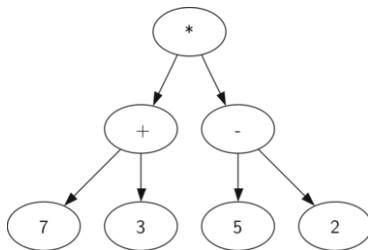
November 2017

Introduction

- A capsule is a group of neurons whose activity vector represents an entity - either an object or an object part
- The length of the vector represents the probability that a certain entity exists
- Achieves state-of-the-art performance on MNIST and is better at recognizing highly overlapping digits than a convolutional network

Reference With Human Vision

- The human vision processes only a tiny fraction of the optic array at the highest resolution
- Assumption that the visual system creates a parse tree on each fixation



(a) Parse tree example:
 $(7 + 3) * (5 - 2)$

Capsules as Parse Trees

- A parse tree can be represented by a multilayer neural network
- Each layer will be divided into many capsules
- Using an iterative routing process, each capsule will choose a capsule in the layer above to be its parent in the tree - this process solves the problem of assigning parts to wholes

Describing the Capsule

- The vector of activities of a capsule represent the various properties of a particular entity that is present in the image (pose, deformation, velocity, texture etc.)
- The existence of an object is defined by using the overall length of the activity vector of a capsule - a non-linearity is applied so that the length of the vector output cannot exceed 1
- A lower level capsule is assigned to a higher level capsule by a "routing-by-agreement" algorithm