MATHEMATICAL AND COMPUTATIONAL LINGUISTICS PROJECT N.5
A KANERVA NETWORK MODEL OF SYNTACTIC PARAMETERS

CS101/MA191 CLASS WINTER 2015, MATILDE MARCOLLI

1. Kanerva Networks

Sparse Distributed Memory (or Kanerva Networks) is a technique first introduced as a model of memory in neuroscience, [2] (see also §13 of [1] for a quick summary). In a large Boolean space $\mathbb{F}_2^N$ one chooses a uniform random sample of *hard locations*. The number 2^k of hard locations is small compared to the size 2^N of the ambient space. At each of these hard locations a datum (a Boolean string of length N) is stored. Compute the median Hamming distance between hard locations. An *access sphere* of a point in the Boolean space $\mathbb{F}_2^N$ is a Hamming sphere of radius slightly larger than the median distance of hard locations (see some estimates in §6 of [2]). The a given location ξ in $\mathbb{F}_2^N$, the datum $D(\xi)$ assigned to that location is *distributively stored* by writing it at all the hard locations within the access sphere of ξ. When the operation is performed for a set of locations ξ, each hard location stores a datum whose i-th entry is determined by the majority rule of the corresponding entries $D(\xi)_i$ for all the stored data $D(\xi)$. Similarly one reads at the location ξ a new datum $D'(\xi)$ whose i-th entry $D'(\xi)_i$ is given by the majority rule on the i-th entries of all the data stored at all the hard locations accessible from ξ. If the datum $D(\xi)$ stored at ξ is ξ itself as a Boolean string of length N, this method works at reconstructing the correct datum $D'(\xi) = D(\xi) = \xi$ because the intersections between access spheres are small and so copies of the datum where some entries are affected are in minority with respect to those copies that are faithful. When a datum is corrupted by noise, iterated readings converge to the original restored datum, provided the distance from the original is not too large. For details, see [2] and the summary in §13 of [1]. Thus, this can be viewed as a method to detect which part of a datum consisting of a binary string is sufficient to reconstruct the rest of it.

2. Plan of the project

One of the main problems in the Principles and Parameters model of linguistics is identifying dependencies between the syntactic parameters. We adapt the main idea of Kanerva networks as a possible method to study dependencies and structures in the space of syntactic parameters of languages. The plan is to construct Kanerva networks associated to the data of linguistic parameters, using as data the boolean strings of syntactic parameters from the database [3]. The idea then is to experiment on corrupting the data of small subsets of parameters and checking whether the Kanerva network can reconstruct the correct ones. That would provide another possible way of arguing that certain parameters are dependent variables determined by the remaining ones.

References

[1] S. Franklin, *Artificial Minds*, MIT Press, 2001.
[2] P. Kanerva, *Sparse Distributed Memory*, MIT Press, 1988.
[3] SSWL Database of Syntactic Parameters:
`http://sswl.railsplayground.net/`