H Y P E R D I M E N S I O N A L

C O M P U T I N G

Computing in distributed representation
with high-dimensional random vectors

Pentti Kanerva

Redwood Center for Theoretical Neuroscience
UC Berkeley

pkanerva @ berkeley . edu

WHAT ARE WE TALKING OF?

- Internal REPRESENTATION of concepts, relations, sensations, perceptions, moments of experience

- COMPUTING with the representations:

  Cognitive Computing based on high/hyper-dimensionality

EXAMPLES of concepts and their relations

child, country, red, truth, rise, morning,
sun rises in the morning, the baby is happy

. SIMILARITY of concepts:

man $\cong$ woman
man $\neq$ lake

```
he drank          \
he drank water     > similar
she drank wine    /
  vs.
tree fell to the ground
```

EXAMPLES (cont'd)

- Easy BRIDGING of dissimilar concepts:

  man - fisherman - fish - lake
  man - plumber - water - lake

- Logical RELATIONS between concepts:

  5 : 10 :: 15 : 20
  5 : 10 :: 15 : 30

  man : woman :: son : daughter
  man : woman :: country : language

- FIGURATIVE use of concepts:

  "What's the dollar of Mexico?"

HEADS-UP

If you know of context vectors and Latent Semantic Analysis/Indexing (LSA/LSI): 1/3-way there

If you know Holographic Reduced Representation (HRR) and its kin: Not much new to report here

REPRESENTATION INSIDE A COMPUTER

Basic units: WORDS made of bits

Word SIZE

- 8, 16, 32, 64, 128-bit words
- Letters, numbers, pointers to memory
- LOCAL representation: meaning tied to location

```
       name                                  sex   age
      +---+---+---+---+---+-/ /-+---+----+----+
      | P | A | T | _ | _ | ... | _ |MALE| 66 |
      +---+---+---+---+---+-/ /-+---+----+----+
 byte   1   2   3   4   5        12   13   14
```

Data record with FIELDS

REPRESENTATION INSIDE A BRAIN

. Very large circuits:
  model with 10,000-bit words, N = 10,000

. More like in a holographic: individual neurons
  - partake in many representations
  - are not critical to any of them (neurons die)

. Hence DISTRIBUTED REPRESENTATION

```
    +-------------------------------------------+
    | (name = Pat) & (sex = male) & (age = 66) |
    +-------------------------------------------+
bit  1 2 3 ...                              10,000
```

HOLISTIC record: "Look Mom, no Fields!"

HIGH-D VECTORS / POINTS OF H-D SPACE

. H-D vectors as basic computational units

. CONCEPTS represented by such vectors

. Vectors for "ATOMIC" concepts can be chosen at RANDOM: A, B, C, ...; cos(A,B) = 0

. Vectors for DERIVED concepts by H-D arithmetic

. Similar vectors --> similar meanings

. Demonstrated here with 10,000-bit vectors

H-D ARITHMETIC

ADDITION (+), thresholded sum (of binary vectors)

- Sum SIMILAR to its constituent vectors:
  $A + B + C \cong A$, $\cos(A+B+C, A) = 0.5$

- Good for representing SETS

MULTIPLICATION (*), bitwise Exclusive-Or (XOR)

- Product DISSIMILAR to its constituent vector:
  $A*B \not\cong A$, $\cos(A*B, A) = 0$

- Good for BINDING

Random PERMUTATION (g)

- Makes DISSIMILAR by scrambling:
  $gA \not\cong A$, $\cos(gA, A) = 0$

- Good for representing SEQUENCES

H-D ARITHMETIC (cont'd)

. Multiplication is INVERTIBLE (XOR self-inverse):

```
X*(X*A) = X XOR (X XOR A)
        = (X XOR X) XOR A
        = 0 XOR A
        = A
```

. Multiplication DISTRIBUTES over addition:

```
X*(A + B + C) = X*A + X*B + X*C
```

. Permutation is INVERTIBLE and DISTRIBUTES over both addition and multiplication

. Multiplication and permutation PRESERVE DISTANCE

ENCODING A HOLISTIC RECORD

BINDING: Pairing a variable with a value using multiplication:

. (x = a) is encoded by X*A

. (name = Pat) is encoded by NAME*PAT

HOLISTIC RECORD:

(name = Pat) & (sex = male) & (age = 66)

is encoded by

NAME*PAT + SEX*MALE + AGE*66 = PAT_MALE_60

where NAME, PAT, SEX, MALE, AGE, and 66 are random H-D vectors

DECODING

"What is the age in the holistic record PAT_MALE_66?"

AGE * PAT_MALE_60

= AGE*(NAME*PAT + SEX*MALE + AGE*66)

= AGE*NAME*PAT + AGE*SEX*MALE + AGE*AGE*66

= noise1 + noise2 + 66

= 66′

≅ 66, cos(66′,66) = 0.5

MAPPING WITH H-D VECTORS

- Multiplication by H-D vector maps
  a SET of points to a different part of space
  MAINTAINING RELATIONS (distances) between them

- The mapping vector can be computed from examples

- Opens a way to ...

  - LEARNING from example

  - SUBSTITUTION within structure

  - FIGURATIVE use of language:

    "What is the dollar of Mexico?"

  - ANALOGY

SEQUENCES A B C D E F ...

FLATTENING a sequence with permutation

Basic idea, RECURRENT NET: To each element, add a permuted (g) version of its history

1.	A	=	A
2.	B + gA	=	B + gA
3.	C + g(B + gA)	=	C + gB + ggA
4.	D + g(C + g(B + gA))	=	D + gC + ggB + gggA

. . .

- ggg...g is yet another permutation
- Different permutations keep track of how far back in history a specific pattern appears

SEQUENCES (cont'd)

- ORDER matters:
  Vectors for A B C and B A C are different

- Can find elements from the past:

  "What happened 2 steps ago in A B C?"

  Using h, the inverse permutation of g:

$$
\begin{aligned}
hh(A\ B\ C) &= hh(C + gB + ggA) \\
&= hhC + hhgB + hhggA \\
&= hhC + hB + A \\
&= A' \\
&\cong A, \quad \cos(A',A) = 0.5
\end{aligned}
$$

WHY WOULD SUCH A SYSTEM OF COMPUTING WORK

when so many results are approximate?

E.g., $\cos(A', A) = 0.5$

EXTRAORDINARY PROPERTIES OF HYPER-D SPACES:

DISTANCE vs. VOLUME

Exemplified by 10,000-bit vectors, N = 10,000

Mathematically: Corners of a 10,000-dimensional unit cube - HYPERCUBE

Distance: HAMMING DISTANCE, number of places at which two patterns differ

SYMMETRY: All points alike - every corner of a (hyper)cube is just like every other corner
hence
no concentration anywhere in SPACE
yet
it looks as if VOLUME of space / MASS / DISTANCE were highly concentrated!!

PROPERTIES (cont'd)

- Distance from ANY POINT to all the points of the space is BINOMIAL
- Approximated with the normal distribution
  - . Mean = 5,000, Standard deviation = 50
- Relative distance
  - . Mean = 0.5, Standard deviation = 0.005

```
                         .999999998
 ANY POINT                    /
     /                       ||
    /  1 in a billion        ||
   /     .000000001          ||        .000000001
 :__________________________/  \______________________:
 0                        .47 .53                     1

distance 0-.47        .47-.53        .47-1
      similar        dissimilar     opposite
     correlated     uncorrelated
                    ≅ orthogonal
```

PROPERTIES (cont'd)

. DISTANCE of 0.47 from a point into the space covers a mere BILLIONTH of the space by VOLUME, whereas distance 0.53 covers nearly ALL

. Thus ROBUSTNESS:

Representation in 10,000-bit vectors is VARIATION/NOISE-TOLERANT:

. 4,700 bits away is very CLOSE, quite SIMILAR
. 5,000 bits away is very FAR, DISSIMILAR

This is a GENERAL property of H-D spaces

PROPERTIES (cont'd)

INTERMEDIARY POINTS:

Between any two UNcorrelated points A and B there are many points that are significantly correlated with both

COGNITIVE ANALOG:

Any two unrelated concepts can be connected by one or two intermediate concepts similar to both

Significant intermediaries mostly UNCORRELATED with each other:

man ≅ fisherman ≅ fish ≅ lake
man ≅ plumber ≅ water ≅ lake

plumber ≇ fish

MEMORY

We can build a RAM for 10,000-bit words:

Sparse Distributed Memory (SDM)

- An associative memory
- Addressed by 10,000-bit words
- Stores 10,000-bit words
- Retrieves with NOISY address
- Capacity can be made arbitrarily large

RELATED IDEAS

- Plate: Holographic Reduced Representation (HRR)
- Rachkovskij & Kussul: Context-Dependent Thinning
- Olshausen: Shifter Circuits
- Arathorn: Mapping Circuits
- Gayler: Vector-Symbolic Architecture (VSA)
- My Sparse Distributed Memory, Spatter Code, and Random Indexing

* * *

- Von Neumann, J. (1948). THE COMPUTER AND THE BRAIN. New Haven and London: Yale U Press.

  Refers to possible need for "NEW MATH"

CONCLUDING MESSAGE

- Very-high-dimensional spaces have extraordinary properties

- A new kind of computing can be based on them

---

* "LOOK MOM, NO FIELDS!"

* MAPPING WITH H-D VECTORS,
COMPUTING OF THE MAPPING VECTORS:
the new kid on the block. Watch it grow!

T H A N K   Y O U

Presented at NIPS 2009 workshop on

The Curse of Dimensionality Problem:
How Can the Brain Solve It?

organized by Simon Haykin, Terrence Sejnowski, and Steven Zucker, held at Whistler, BC, Canada on December 11, 2009.

Based on my paper by the same name in COGNITIVE COMPUTATION 1(2):139-159, 2009:

http://www.springerlink.com/content/96615184lg415165/

Please see it for a more complete story and references.