Real, Complex, and Binary Semantic Vectors

Dominic Widdows
Microsoft Bing

Trevor Cohen
University of Texas School of Biomedical Informatics at Houston

Lance De Vine
Queensland University of Technology

http://semanticvectors.googlecode.com/

“Three Fields” (Numbers and People!)

- Complex Numbers
  - Physics
- Real Numbers
  - Statistics / Machine Learning
- Binary
  - Logic / Linguistics / Computer Science
- Mathematical fragmentation because of community history is silly!
- We have (finally!) taken steps in the SemanticVectors package to fix this. Now you pick your field at runtime!

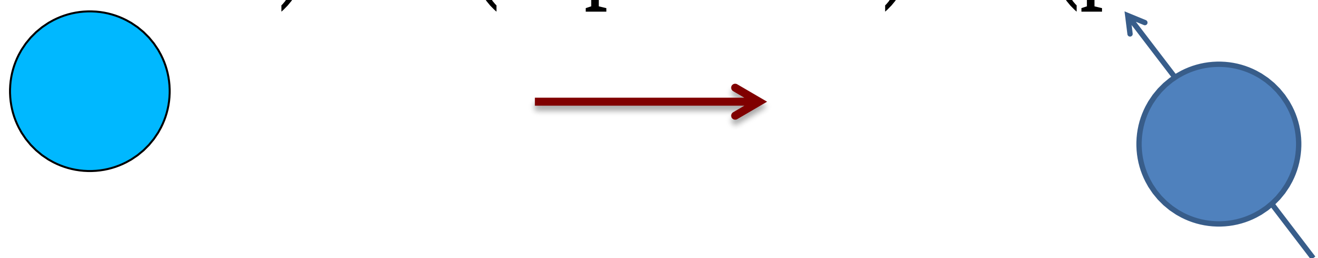
Vector Symbolic Architectures

- Tony Plate, Ross Gayler, Pentti Kanerva
- Vector Space (could be any vector space)
- Generate Random Vector
- Measure Overlap: $a \bullet b$
  - For large dimensions, the overlap measure can be used with even large numbers of randomly generated vectors to distinguish "same" and "different"
- (Weighted) Superposition (bundling): $a + b$
  - Behaves "nicely" with overlap measure
  - If $x \bullet (a + b)$ is significantly bigger than 0, this is strong evidence that $x = a$ or $x = b$
- Binding and release
  - Binding maps a and b to another vector, different from both
  - Release is the inverse of binding: must be possible to recover an ingredient given other ingredients and their product

Invertible Binding

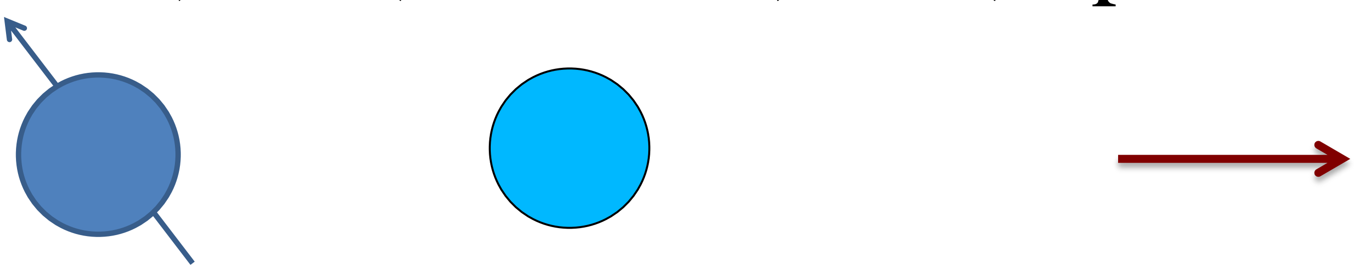
BINDING

$$E(\text{TREATS})\,\text{Ä}\; E(\text{depression}) = S(\text{prozac})$$



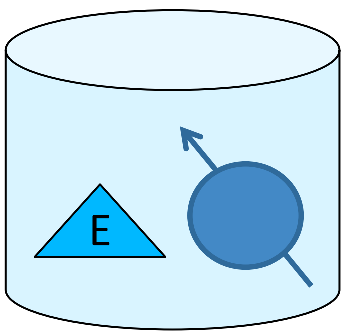
RELEASE

$$S(\text{prozac}) \oslash E(\text{TREATS}) \approx E(\text{depression})$$



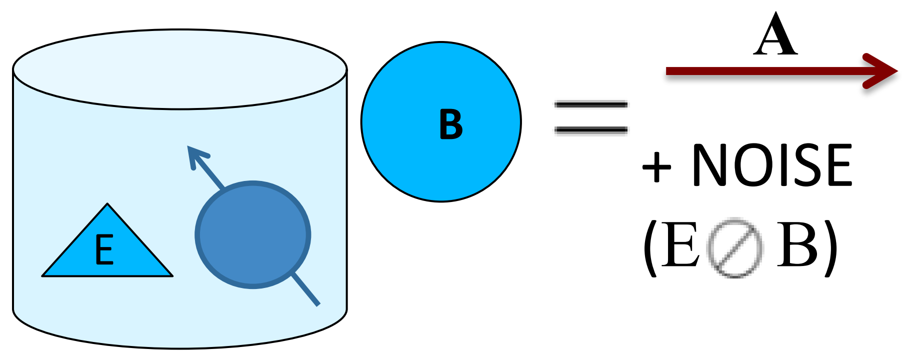
Binding Distributes over Bundling

BUNDLING



$$D = E + (A \otimes B)$$

RELEASE



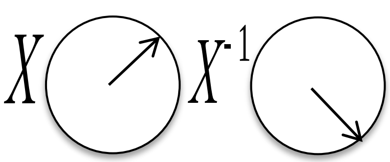
$$D \oslash B \approx A$$

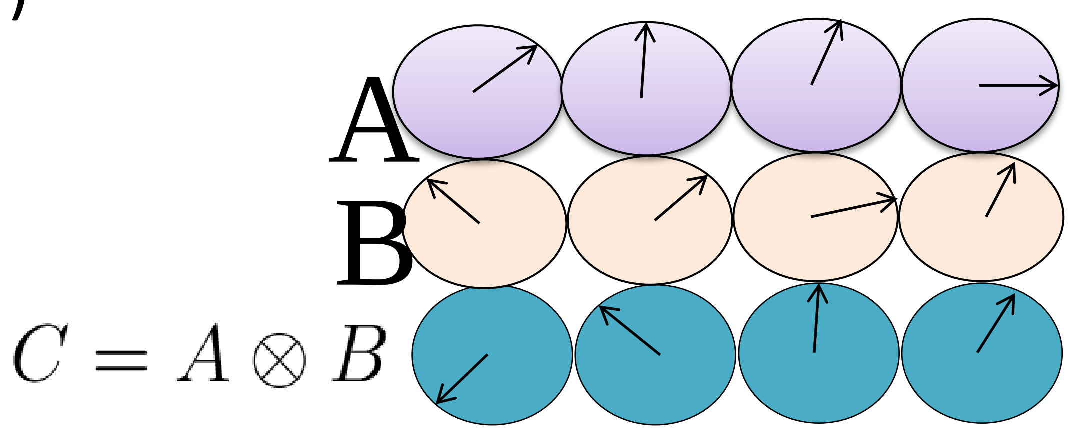
Real Vectors

- Overlap measure – scalar product
- Superposition – linear sum
- Binding – permutation of coordinates then linear sum
- Bundling – subtract inverse permutation

Circular Holographic Reduced Representations (Plate / De Vine)

- Complex vectors (e.g. d=4000)
- Bundling: addition of circular vectors
- Binding via circular convolution
  - Addition of phase angles
  - Pairwise multiplication of component vectors
  - Linear time (no FFT)
  - $C \oslash B = C \otimes B^{-1}$





Binary Spatter Code (Kanerva)

- Binary vectors (e.g. d=32,000), p(1) = p(0) = 0.5
- Bundling: count 1s + 0s, ties broken at random
- Binding : component-wise XOR (self-inverse)

BINDING	RELEASE
A = [0 1 0 0 1]	A = [0 1 0 0 1]
B = [1 0 1 0 1]	C = [1 1 1 0 0]
C = [1 1 1 0 0]	**B = [1 0 1 0 1]**
$C = A \otimes B$	$B = A \oslash C$

Special Binary Features

- Overlap measure is “normalized” Hamming distance
- Maintain a “voting record” to break ties
- This implies “decision” between 1 and 0 for each coordinate at different stages “when we’re done”
- Binary orthogonalization and negation then took advantage of this

Details you may want to ignore

- Elemental and semantic vectors
- Need a sparse representation for elemental vectors
- Binding and bundling cause output to be returned as a dense representation
- Serialization deals with dense representations (so reusing elemental vectors is tricky)
- Complex representation is
- We can do some kind of binary negation
- Many details and instructions at https://semanticvectors.googlecode.com/svn/javadoc/latest-stable/index.html and elsewhere on project site
- Time for demos …

What does this mean for Quantum Interaction?

- Quantum Mechanics or Generalized Quantum?
  - Bell inequalities, Born's rule, self-adjoint operators, entanglement, …
- If quantum mechanics is really key, we would certainly expect complex Hilbert space to be superior
- If generalized quantum properties are really key, we would expect these properties to show up in other models which can claim to be "quantum like"