Towards logics for neural conceptors

Till Mossakowski
joint work with Razvan Diaconescu and Martin Glauer

Otto-von-Guericke-Universität Magdeburg

AITP 2018, Aussois, March 30, 2018

Overview

Overview

Motivation

Conceptors [Jaeger14]

- Combination of neural networks and logic
- Using a distributed representation like in deep learning and human brain
  - most neural-symbolic integration use localist represtation
  - e.g. logic tensor networks (AITP17): one network for each predicate
- Boolean operators
  - provide concept hierarchy
  - new samples can be added without re-training

Our contribution

- Conceptors obey the laws of fuzzy sets
- Fuzzy logic is the natural logic for conceptors

Reservoir dynamics [Jaeger14]

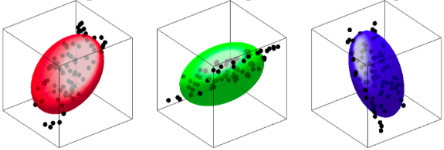
- reservoir = randomly created recurrent neural network
- input signal p drives this network
- for timesteps $n = 0, 1, 2, \ldots L$,

$$x(n+1) = tanh(W\,x(n) + W^{in} p(n+1) + b)$$

  - W: $N \times N$ matrix of reservoir-internal connection weights
  - W^{in}: $N \times 1$ vector of input connection weights
  - b: bias
  - p: input signal (pattern)

W, W^{in} and b are randomly created

Conceptors [Jaeger14]



- collect state vectors $x_0, \ldots x_L$ into $N \times L$ matrix X = cloud of points in the N-dimensional reservoir state space
- reservoir state correlation matrix: $R = XX^T/L$
- conceptor: normalised ellipsoid (inside the unit sphere) representing the cloud of points

$$C = R(R + \alpha^{-2} I)^{-1} \in [0,1]^{N \times N}$$

α: aperture (scaling parameter)

We here use a simplified version where $C = diag(c_1 \ldots c_n)$. The c_i are called conception weights.

Boolean operations on simplified conceptors

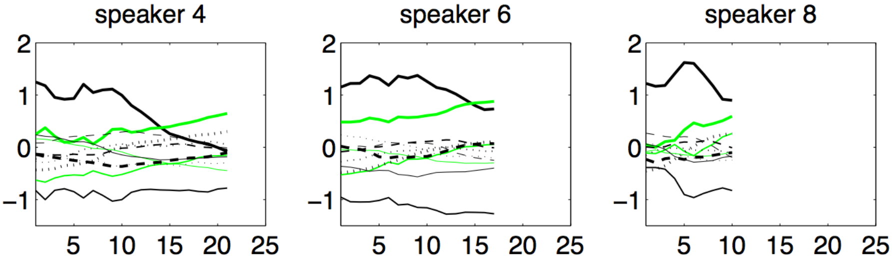
$$
\begin{aligned}
(\neg c)_i &:= 1 - c_i \\
(c \wedge b)_i &:= \begin{cases} c_i b_i/(c_i + b_i - c_i b_i), & \text{if not } c_i = b_i = 0 \\ 0, & \text{if } c_i = b_i = 0 \end{cases} \\
(c \vee b)_i &:= \begin{cases} (c_i + b_i - 2 c_i b_i)/(1 - c_i b_i), & \text{if not } c_i = b_i = 1 \\ 1, & \text{if } c_i = b_i = 1 \end{cases}
\end{aligned}
$$

Aperture adaption

$$
\begin{aligned}
\varphi(c,\gamma)_i &:= c_i/(c_i + \gamma - 2(1 - c_i)) \text{ for } 0 < \gamma < \infty \\
\varphi(c,0)_i &:= \begin{cases} 0, & \text{if } c_i < 1 \\ 1, & \text{if } c_i = 1 \end{cases} \\
\varphi(c,\infty)_i &:= \begin{cases} 1, & \text{if } c_i > 0 \\ 0, & \text{if } c_i = 0 \end{cases}
\end{aligned}
$$

Overview

"Japanese Vowels" Pattern Recognition [Jaeger14]



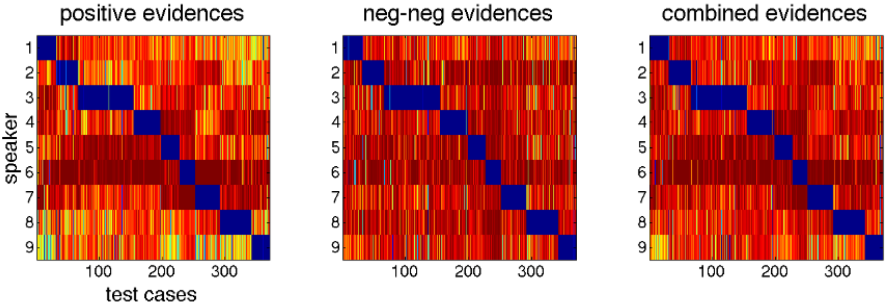
- Data: 12-channel recordings of short utterance of 9 male Japanese speakers
- 270 training recordings, 370 test recordings
- Task: train speaker recognizer on training data, test on test data

Conceptors at work: Japanese vowels [Jaeger14]

- for each speaker j, build a conceptor C_j
- for a test pattern p, compute the reservoir response signal r
- positive classification for speaker j: use $\frac{1}{L}(r^T C_j r)$
- negative classification for speaker j, using Boolean conceptor logic

$$\frac{1}{L}(r^T \neg(C_1 \vee \cdots \vee C_{j-1} \vee C_{j+1} \vee \cdots \vee C_n) r)$$

Result of Japanese vowel classification [Jaeger14]



- with 10-neuron reservoirs, mean (50 trials with fresh reservoirs) test errors for 370 tests:
  8.4 (positive ev.) / 5.9 (neg-neg ev.) / 3.4 (combined)
- incremental model extension possible, again enabled by Boolean logic

Overview

Conceptors are fuzzy

Central thesis:

> *Conceptors and conception vectors behave like fuzzy sets, and their logic should be a fuzzy logic.*

Proposition

Conceptors form a (generalised) de Morgan triplet, i.e. a t-norm, a t-conorm and a negation that interact usefully.

Laws for a deMorgan triplet

- $\neg 0 = 1, \neg 1 = 0$
- $x < y$ implies $\neg x > \neg y$ (strict anti-monotonicity)
- $\neg\neg x = x$ (involution)
- T1: $x \wedge 1 = x$ (identity)
- T2: $x \wedge y = y \wedge x$ (commutativity)
- T3: $x \wedge (y \wedge z) = (x \wedge y) \wedge z$ (associativity)
- T4: If $x \leq u$ and $y \leq v$ then $x \wedge y \leq u \wedge v$ (monotonicity)
- S1: $x \vee 0 = x$ (identity)
- T2: $x \vee y = y \vee x$ (commutativity)
- T3: $x \vee (y \vee z) = (x \vee y) \vee z$ (associativity)
- T4: If $x \leq u$ and $y \leq v$ then $x \vee y \leq u \vee v$ (monotonicity)
- $x \vee y = \neg(\neg x \wedge \neg y)$ (de Morgan)

Further algebraic laws

- $\wedge$ and $\vee$ do not form a lattice, so De Morgan algebras, residuated lattices, BL-agebras, MV-algebras, MTL-algebras etc. do not apply

[Jaeger14] lists:

- $C \vee C = \varphi(C, \sqrt{2})$
- $C \wedge C = \varphi(C, \sqrt{\frac{1}{2}})$
- $A \leq B$ iff $\exists C. A \vee C = B$
- $A \leq B$ iff $\exists C. A = B \wedge C$

Implication

In a De Morgan triplet, we can define implication as

$$x \to y = \neg x \vee y$$

Alternative: residual implication

$$R(x, y) = \sup\{t \mid x \wedge t \leq y\}$$

But: has the unpleasant property that

$$R(c, 0)_i = \begin{cases} 0, & \text{if } c_i > 0 \\ 1, & \text{if } c_i = 0 \end{cases}$$

while our implication behaves more smoothly: $(c \to 0)_i = 1 - c_i$

Fuzzy conceptor logic

- parameterised over dimension $N \in \mathbb{N}$
- two sorts: individuals and conception vectors
- Signatures: constants for individuals and for conception vectors
- Models: interpret constants as N-dimensional vectors in $[0,1]^N$
  - individuals are interpretated as feature vectors, conceptor terms as conception vectors
- Conceptor terms:

$$C ::= c \mid x \mid 0 \mid 1 \mid \neg x \mid C_1 \vee C_2 \mid C_1 \wedge C_2 \mid \varphi(C, r)$$

- Atomic formulas:
  1. ordering relations between conceptor terms
  2. memberships of individual constants in conception vectors

Fuzzy conceptor logic: semantics

A formula yields a fuzzy truth value in $[0,1]$:

$$[\![C_1 \leq C_2]\!] = \min_{j=1\ldots N}([\![C_1]\!]_j \to [\![C_2]\!]_j)$$
$$[\![i \in C]\!] = \frac{1}{N}[\![i]\!]^T diag([\![C]\!])[\![i]\!]$$

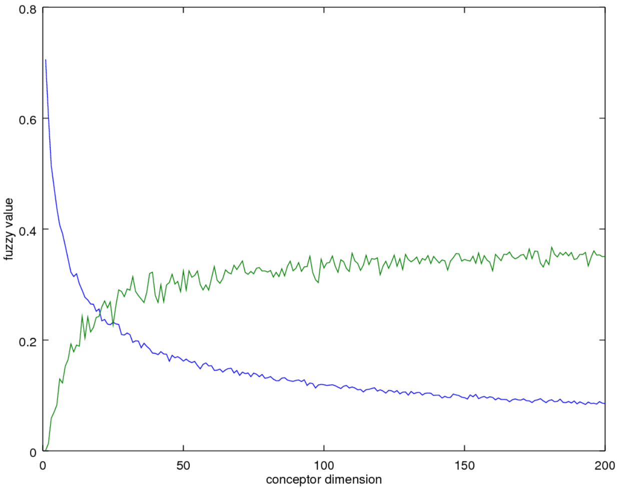
Complex formulas like in FOL:

$$F ::= i \in C \,|\, C_1 \leq C_2 \,|\, \neg F \,|\, F_1 \vee F_2 \,|\, F_1 \wedge F_2 \,|\, \forall x^i.F \,|\, \forall x^c.F \,|\, \exists x^i.F \,|\, \exists x^c.F$$

- x^i: variable ranging over individuals
- x^c: variable ranging over conception vectors.
- Interpretation of formulas like in fuzzy FOL:
  - infimum for universal quantification
  - supremum for existential quantification

Fuzzy conceptor logic: subset relations

Consider $C \leq D$ versus $\forall x^i.(x^i \in C \rightarrow x^i \in D)$:



Fuzzy conceptor logic in action

Suppose we have two sets of speakers, call them Dialect_1 and Dialect_2. Using disjunction, we can build conceptors C_1 and C_2 for these sets. Then we can ask:

- how far is Dialect_1 similar to Dialect_2? ($C_1 \leq C_2 \wedge C_2 \leq C_1$)
- how much is Dialect_1 a sub-dialect of Dialect_2? ($C_1 \leq C_2$)

If we have an ontology of dialects, we can

- test the ontology by checking how far it follows from speaker data
- infer new consequences by (fuzzy/crisp) reasoning in the (fuzzy/crisp) ontology

Overview

Conclusions

- Defined a new fuzzy logic for conceptors
  - In his conceptor report, Jaeger only defines two crisp logics
- Can be basis for neural-symbolic integration
  - Crisp and fuzzy reasoning about ontologies of concepts
  - Learning and classification using conceptors

Future work

- Fuzzy conceptor logic
  - suitable algebraisation
  - proof calculus
  - automated theorem proving
- Work out details of integrated reasoning