Unsupervised learning of VerbNet argument structure

Jesse Mu, Timothy J. O’Donnell, Joshua K. Hartshorne
1Department of Psychology, Boston College
2Brain and Cognitive Sciences, MIT

Background and motivation

Why can we say “She slept in the bed” but not “She slept the bed”? Argument structure acquisition problem (e.g. Levin, 2007)

Competing theories of argument structure (e.g. Pinker, 1989; Goldberg, 2006) agree on hierarchical and overlapping verb classes that predict syntax.

Q1 How well can we recover verb classes from syntactic frames alone?
Q2 Can learned structure predict graded human intuitions about grammar?

Desideratum: Large scale investigation (cf. case studies above)

Research questions

Why can we say “She slept in the bed” but not “She slept the bed”?

Verb classes that predict syntax agree on (e.g. Pinker, 1989; Goldberg, 2006)

Competing theories of argument structure

Research questions

Q1 How well can we recover verb classes from syntactic frames alone?
Q2 Can learned structure predict graded human intuitions about grammar?

Desideratum: Large scale investigation (cf. case studies above)

Research questions

Q1 How well can we recover verb classes from syntactic frames alone?
Q2 Can learned structure predict graded human intuitions about grammar?

Desideratum: Large scale investigation (cf. case studies above)

Cluster analysis

Cluster analysis allows us to systematically learn structure in large datasets

Gold standard: VerbNet (Köppen et al., 2008; hand-constructed)

Binary matrix encoding: 6334 verbs x 7 frames

Options for encoding syntactic frames

• Expand prepositions (PPs) into specific thematic roles
• Add/remove
  • selectional restrictions
  • thematic roles

Bayesian Hierarchical Clustering (BHC)

(Miller and Shanks, 2005)

Advantages over traditional hierarchical clustering (e.g. UPGMA):

• No distance metric required
• Provides intuitive place(s) to cut tree for flat clustering

Method & Encoding V Measure Entanglement Robinson-Sundberg median
BHC 0.92 0.26 499
UPGMA 0.85 0.27 507
BHC + Frame semantics, expand PPs 0.85 0.26 521
BHC + Add theta roles, expand PPs 0.93 0.21 463
Random baseline — 0.67 900

Comparison between cut BHC tree and flat VerbNet clustering (The “crossing number” [CN] of the tanglegram normalized by the worst case tanglegram [Penny and Hendy (1985)] variant).

A1 BHC v-measure high, tree metrics significantly better than chance.

A measure of grammaticality

BHC’s hierarchy defines fine-grained similarity between verbs. For Q2, we want to predict grammaticality judgments on verbs in frames.

“Distance” measure from verb to frame: Geometric mean of cophenetic distances to all classes that take the frame

Intuition: for an imperfect verb + frame, consider coercions to all compatible classes

Hypothesis: for an imperfect verb + frame, consider coercions to all compatible classes

Hierarchy evaluation

Tanglegrams (e.g. Scovascia et al., 2011) enable visual comparison of two hierarchies

BHC (optimal encoding) VerbNet

A1 Good qualitative alignment with VerbNet

Every 5th leaf labeled. Shades of red indicate worse alignment with VerbNet.

Predicting human judgments

Is BHC a good predictor of human grammaticality judgments?

Stimuli
Sample 10 VerbNet frames, 10 verbs for each frame (varying grammaticality)

Procedure
50 MTurk participants rate grammaticality of verb-frame pairs

Grammaticality predictions by BHC

A2 BHC predicts graded acceptability well, VerbNet doesn’t.

Implications:
1. We can explain graded acceptability with learned verb classes
2. Human grammatical intuitions are more fine-grained than VerbNet structure

Next steps

• Factor analysis (e.g. IBP; Wood et al, 2006)
• Finding and comparing other predictors of grammaticality

Conclusions

A1 Clustering on syntax reliably recovers VerbNet’s broader structure

A2 BHC predicts graded acceptability well, VerbNet doesn’t. Implications:
1. We can explain graded acceptability with learned verb classes
2. Human grammar intuitions are more fine-grained than VerbNet structure

Acknowledgements: JM is supported by an undergraduate research fellowship from the Boston College Psychology Department and a Barry M. Goldwater Scholarship. JH is supported by the Ruth L. Kirschstein NRSA program and the Center for Brains, Minds, and Machines at MIT.