Communicative modes in parent—child conversations

Anonymous

Abstract

Keywords: language input, language acquisition, child-directed speech, corpus analysis, word2vec

Words are amenable to storage, tabulation, and manipulation using common computational tools. Perhaps due to their ease of use, words have received outsize attention in modeling language development.

Studying children's communicative acts using a computational model also allows us to extract communicative patterns across many children with less a priori specification of what those patterns are. Traditionally, studies of communicative acts among children have brought frameworks from adult communication, such as Speech Act theory (Austin, 1962, p. @searle_speech_1969) and Conversation Analysis (Sacks, Schegloff, & Jefferson, 1974), to bear on children's conversations. While concepts like 'directive', 'expressive', 'accusation' and 'justification' can be useful to characterize children's conversations, young children may not have the same communicative needs as adults—these may not be the right joints at which to carve.

Here, we characterize children's growing repertoire of communicative acts using a Hidden Topic Markov Model. This model observes utterances produced by parents and children and attempts to infer common underlying processes—topics—that produced them.

We first show that without top-down specification, this model extracts several communicative acts analogous to those observed in close case studies of children's communication. We then show that use of these acts has a developmental trajectory in line with findings from close studies of children's communication. ...

Corpus

We use transcripts of conversations from the Child Language Data Exchange System (CHILDES), a database of child conversation corpora (MacWhinney, 2000). These corpora predominantly record spontaneous conversations between children and their family members, often in the home. We filtered the transcripts to include only utterances spoken by the target child or their parents. Overall, our training data included 4,043 transcripts from 411 children three months to 60 months old.

Model

We use a Hidden Topic Markov Model (Gruber, Weiss, & Rosen-Zvi, 2007) to extract communicative modes from parent—child conversations. Topic models represent documents as mixtures of topics, and topics as mixtures of words. For instance, a simple topic model trained on news articles may extract a topic whose distinctive words are "fire", "flood", and "aid" and another whose distinctive words are "speech", "legislation", and "administration". Based on its distribution of words, an article about politicians' provision of disaster relief may be correctly inferred to feature these two topics, among others. Intuitively, the goal of a topic model is to recover the underlying sources—topics—from which the words in a document spring.

The Hidden Topic Markov Model (HTMM) differs from a simple topic model in that it takes into account the sequential utterance structure of a document, not just its static distribution of words. The HTMM assumes that words within an utterance are of the same topic, and that sequential utterances may be more likely to be of the same topic. It represents topic transitions between utterances in a coarse-grained way: either switch or stay. Gruber et al. (2007) develop this model and use it to segment machine learning conference papers, showing that the model can distinguish instances of the word "support" in mathematical contexts (describing support vectors) from those in the context of acknowledgements. Typically, function words are removed from corpora before training topic models to aid detection of thematic content. Here, we aim to classify communicative modes underlying utterances rather than thematic topics. We expect function words to be diagnostic of these modes, so include them in our training data.

After training, the model produces a set of topics with associated probability distributions over words and a probability distribution over topics for each utterance in the corpus. The most distinctive words of each topic in our 15–topic model, as measured by difference between a word's probability within a topic and its probability in the corpus as a whole, are shown in Figure 1.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
i	a	two	not	you	to	of	some	you	the		yeah	a	the	your
you	that	one	it's	it	we	they	you	what	in	mommy	no		and	on
what	what's	dis	good	me	go	we	want	are	there	daddy	oh	one	he	my
know	is	three	i'm	to	and	have	i	do	put	hi	okay	this	was	put
don't	this	four	you're	can	you	like	more	look	here	down	hm	is	his	his
it	that's	de	a	want	going	the	have	see	it	baby	uh	that's	said	her
do	what	dat	that's	okay	did	i	eat	with	go	bye	huh	blue	in	hair
that	the	five	very	i	when	and	milk	at	on	ball	ah	red	she	off
did	look	duh	be	up	at	them	juice	play	right	n	mhm	green	of	in
think	who's	six	he	get	school	are	your	those	this	tape	mm	make	to	head

References

Austin, J. L. (1962). *How to do things with words*. Cambridge: Harvard University Press.

Gruber, A., Weiss, Y., & Rosen-Zvi, M. (2007). Hidden Topic Markov Models. In *Artificial Intelligence and Statistics* (pp. 163–170). PMLR. Retrieved from http://proceedings.mlr.press/v2/gruber07a.html

MacWhinney, B. (2000). *The CHILDES Project: The database*. Psychology Press.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735. http://doi.org/10.2307/412243

Searle, J. R., & Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

age	speaker_role	topic	gloss
19	Mother	5	let's do it hold it together
19	Mother	5	naima hold it too
19	Mother	10	mommy and naima can put it back together
19	Mother	10	there
19	Mother	10	that's the spinosaurus
19	Mother	2	what's that
19	Target_Child	10	spinosaurus
19	Mother	10	that's the styracosaurus
19	Target_Child	10	lemme put the styracosaurus back
19	Mother	1	that's the right yes you did it all by yourself
38	Mother	1	do you know what that is
38	Target_Child	5	need to get this thing off here
38	Mother	1	what do you think it is
38	Target_Child	15	a nails
38	Mother	10	i think they're screws to screw and unscrew
38	Mother	5	can you take one off
38	Target_Child	5	no
38	Mother	5	lemme see it
38	Mother	10	bring it over here
38	Target_Child	12	okay
43	Mother	5	i'll fold it up and put it in this envelope how about that
43	Target_Child	1	no i made that one for grace
43	Mother	12	yeah
43	Mother	5	we'll give it to grace
43	Mother	7	but she can have two envelopes because i already sealed the other one see
43	Mother	7	she can have two right
43	Target_Child	13	mom
43	Mother	6	see the bird cape is for who
43	Mother	6	who
43	Target_Child	6	it's for