Integrating Experiential and Distributional Data to Learn Semantic Representations

Mark Andrews, Gabriella Vigliocco, David P. Vinson

Cognitive, Perceptual & Brain Science, University College London

{m.andrews,g.vigliocco,d.vinson}@ucl.ac.uk

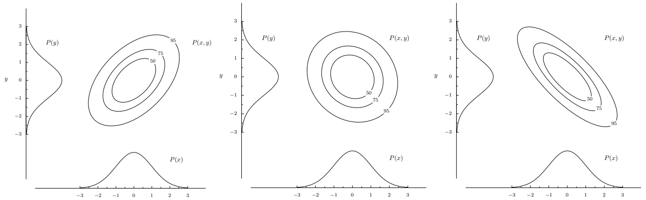
THERE are two major types of statistical data from which we can learn semantic representations:

1. *Experiential* data is derived by way of our experience with the physical world and comprises the sensory-motor data obtained through our sense-receptors.
2. *Distributional data*, by contrast, describes the statistical distribution of words across spoken and written language.

In previous literature, the roles of these data-types have been considered indepedently and in a mutually exclusive manner, e.g. McRae, de Sa, and Seidenberg (1997); Vigliocco, Vinson, Lewis, and Garrett (2004); Landauer and Dumais (1997); Griffiths, Steyvers, and Tenenbaum (2007). Our theoretical proposal is that human semantic representations are derived from an statistical combination of these two data types.

The Consequences of Combining Data Types

By learning semantic representations from the joint distribution of experiential and distributional data, more semantic knowledge may be gained from the available data than is possible using one source exclusively, or using both independently. This is a consequence of the elementary statistical fact that all the information in a joint probability distribution can not be known by reference to its marginal distributions.



We can see this above where the joint distribution $P(x, y)$ varies across the sub-figures, but both marginal distributions, $P(x)$ and $P(y)$, remain unchanged. By a direct analogy, all the information from which semantic knowledge can be attained is given by the joint distribution over both experiential and distributional data. It is only by treating the data as a single joint data-set can all the available information be utilized.

THE probabilistic models we use are based on the Latent Dirichlet Allocation (LDA) model introduced by Blei, Ng, and Jordan (2003) and also used, for example, in Griffiths et al. (2007). For more details, see our accompanying appendix.

Semantic knowledge derived from an LDA model based on experiential data alone is represented as a set of clusters of sensory-motor features, e.g.

juice	fur	speak	wheel	mix	construct	leg
yellow	4-legs	word	transport	rotate	build	fast
red	tail	voice	passenger	spoon	new	exercise
round	pet	talk	gas	turn	fix	feet
grow	big	mouth	automobile	utensil	work	slow
sweet	small	language	drive	dance	create	body
sour	bark	sound	metal	hand	building	intentional

In a LDA model using distributional data alone, semantic knowledge corresponds to a set of discourse-topics, e.g.

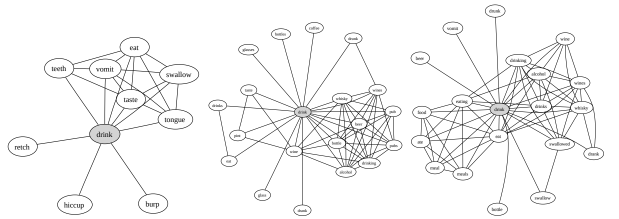
league	prison	rate	pub	market	railway	air
cup	years	cent	guinness	stock	train	aircraft
season	sentence	inflation	beer	exchange	station	flying
team	jail	recession	drink	demand	steam	flight
game	home	recovery	bar	share	rail	plane
match	prisoner	economy	drinking	group	locomotive	airport
division	serving	cut	alcohol	news	class	pilot

In a LDA model using both experiential and distributional data in combination, semantic knowledge corresponds to sets of coupled feature-clusters and discourse-topics, e.g.

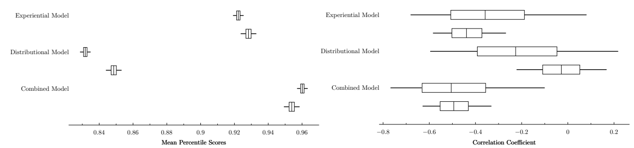
Features	*mouth*	*teach*	*drive*	*body*	*food*	*need*
	liquid	*learn*	*wheel*	*hand*	*cook*	*give*
	consume	*instruct*	*engine*	*joint*	*kitchen*	*money*
	food	*guide*	*gas*	*move*	*pot*	*purchase*
	swallow	*school*	*move*	*arm*	*heat*	*own*
	ingest	*talk*	*passenger*	*humans*	*hot*	*trade*
	enjoy	*idea*	*steer*	*connect*	*eat*	*return*
Words	food	course	car	arms	add	bank
	eat	students	road	arm	cook	exchange
	drink	english	drive	fingers	oil	loan
	eating	language	driving	side	minutes	loans
	wine	education	cars	hands	chopped	lend
	drinking	college	driver	shoulder	heat	mortgage
	drinks	university	drove	body	serve	borrow

THE similarity between the semantic representations of any pair of words, in any model, can then be measured by the distance between their distributions over the model's latent variables. Below we show examples of the neighbourhoods of *drink*.

Below we show examples of the neighbourhoods of *drink* according to the (from left to right) experiential, distributional and combined models.



According to our hypotheses, the semantic similarities in the combined model should more closely resemble human semantic representations. Below left, we show the correspondence between each model and the Nelson and EAT association norms. Below right, we show the correlation between lexical decision reaction times and neighbor closeness in each model.



References

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993-1022.

Griffiths, T., Steyvers, M., & Tenenbaum, J. (2007). Topics in semantic representation. *Psychological Review*, *114*, 211-244.

Landauer, T., & Dumais, S. (1997). A solutions to Plato's problem: The Latent Semantic Analyis theory of acquistion, induction and representation of knowledge. *Psychological Review*, *104*, 211-240.

McRae, K., de Sa, V., & Seidenberg, M. (1997). On the nature and scope of featural representation of word meaning. *Journal of Experimental Psychology: General*, *126*, 99-130.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422-488.