

How Large Are Lions? Inducing Distributions over Quantitative Attributes

Yanai Elazar*
Bar Ilan Univesity
yanaiela@gmail.com

Abhijit Mahabal†
Pinterest
amahabal@gmail.com

Deepak Ramachandran
Google Research
ramachandrand@google.com

Tania Bedrax-Weiss
Google Research
tbedrax@google.com

Dan Roth
University of Pennsylvania
danroth@seas.upenn.edu

1 Supplementary Material

In Figure 1 we present length distributions for the word ‘bat’ and two of its senses, showing a clear distinction between, what we believe to be, baseball bats and cricket bats. In Figure 2, we show the DoQ-induced and real average temperatures of countries. Figure 3 shows some additional measurement distributions from DoQ for various objects types and dimensions, as well as some examples of systematic errors caused by the collection/aggregation process.

In Table 1 we present specific example sentences used to extract correct measurements for objects along the ten most informative dimensions, while Table 2 shows examples where incorrect measurement attributions were extracted. Finally, Table 3 describes the different measurements and domains used in the new dataset we built, along with inter-annotator agreements, a representative example and the labeled relation.

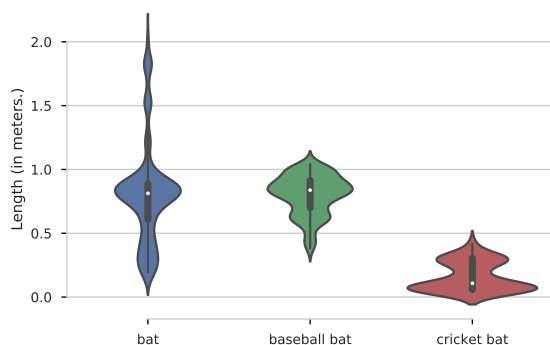


Figure 1: Induced distributions for the various polysemous meanings of ‘bat’.

* Work carried out during an internship at Google.

† Work carried out during employment at Google.

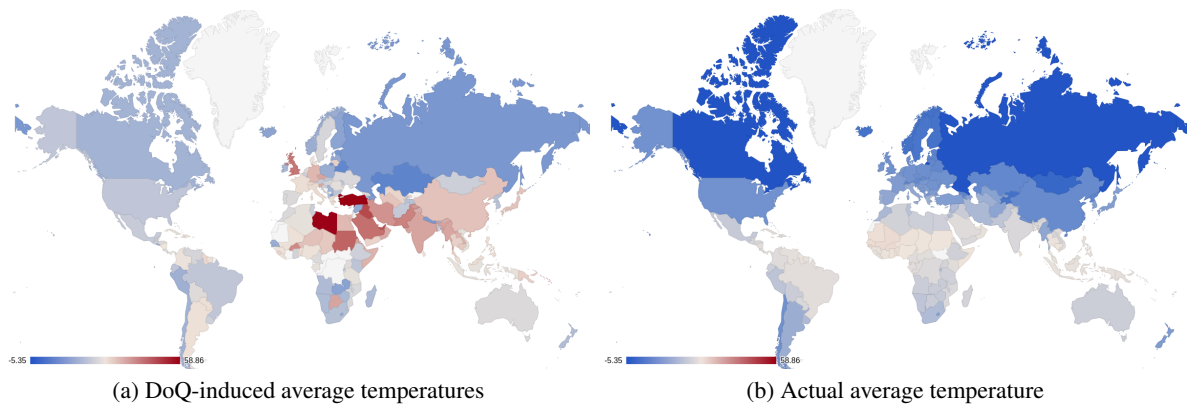


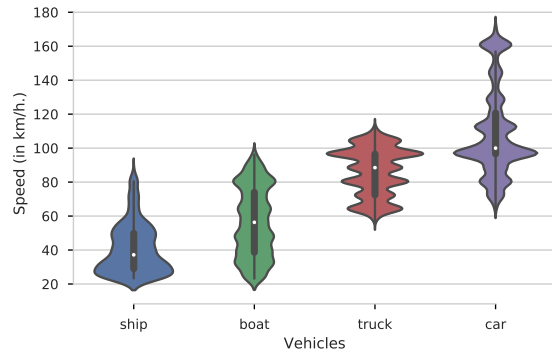
Figure 2: Average world temperatures, predicted by DOQ (Figure 2a), and the ground truth (Figure 2b, taken from https://en.wikipedia.org/wiki/List_of_countries_by_average_yearly_temperature). Interestingly, predicted temperatures are consistently higher than true ones, with the two hottest countries being Turkey (58.8°C) and Libya (57.7°C).

Dimension	Example
TIME	On 24 September drivers struck between 7.30 am and 8.30 am, the middle of the morning rush hour .
CURRENCY	McKay picked up \$2,000 (the biggest winning cheque of her career) for winning the competition.
LENGTH	The railway line has 140 bridges and 25 km of tunnels
AREA	With over 36,000 acres, Smoky Hill is the Air National Guard’s largest weapons range.
VOLUME	The traditional Bordeaux barrique is an oak barrel with a capacity of 225 litres.
MASS	The men ’s 73 kg competition of the 2014 World Judo Championships was held on 27 August.
TEMPERATURE	Places as far north as New York City reached 70 F (21 C) on Christmas Eve.
DURATION	Formerly the market lasted for 14 days.
SPEED	Dora was the strongest storm of the year, peaking at 155 mph, just short of Category 5 status.
VOLTAGE	The motorcycle used a 24 volt electric starter motor from a Douglas A-4B fighter plane.

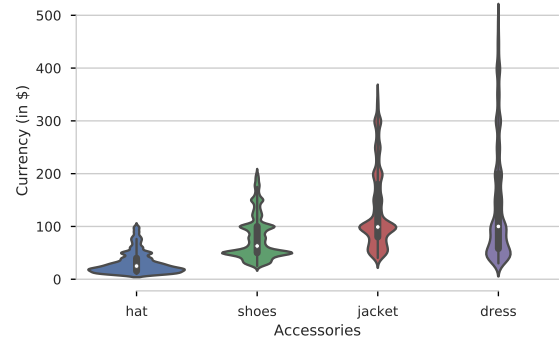
Table 1: Examples of sentences from the web corpus used to infer a correct measurement of the corresponding object for different dimensions.

Measurement	Example
TIME	The largest aftershock measured 5.6 and struck nine minutes later at 1.51 pm.
CURRENCY	Tuncay’s move was confirmed the following day , with Stoke paying 5 million for the Turkish player.
LENGTH	Everest from nine miles away in Rongbuk valley.
AREA	The plant is spread over 25 acres of land.
VOLUME	It was created for cars with engine size below 2 Litres.
MASS	In recent years, the reported average annual peanut production lies around 828,000 tons (95% for oil).
TEMPERATURE	The corresponding phosphorine and benzene undergo the analogous reaction at 100 C and 200 C, respectively.
DURATION	Its surface remains frozen for around 155 days per year.
SPEED	Both trains covered the 466 mile route at an average pace of 49 mph.
VOLTAGE	The value of the shunt ensures a potential difference of 0.5 volt across it at the maximum generator load.

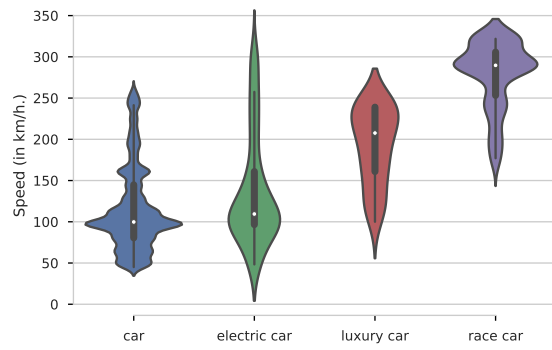
Table 2: Examples of sentences from the web corpus from which an incorrect measurement was extracted, due to attribution/reporting bias or other reasons.



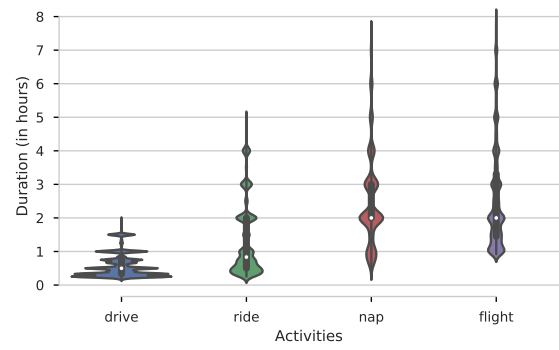
(a)



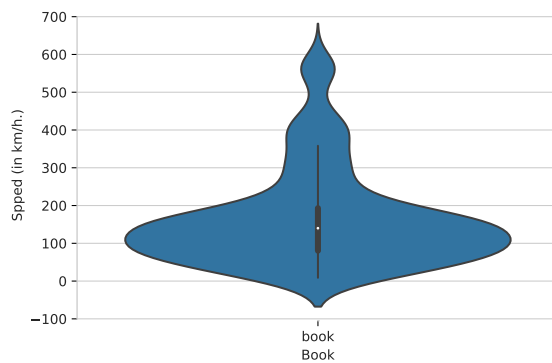
(b)



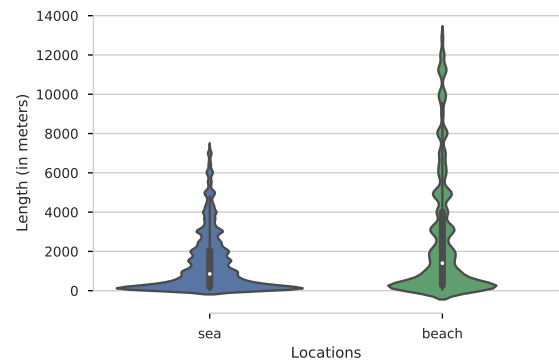
(c)



(d)



(e)



(f)

Figure 3: Comparisons of measurements extracted from DoQ for different categories of objects along their most salient dimensions. Note that (e) and (f) show spurious measurements for the speed of books (caused by attribution bias) and the lengths of seas and beaches (caused by confusion with sea levels).

Measurement	Domain	Count	Inter-Annotator Agreement	Example	Relation
Mass	Vegetables & Fruits	500	76.9	(alfalfa, watermelon)	<
	Animals	501	86.0	(skunk, lemur)	<
	Technology	500	80.2	(modem, microphone)	=
Speed	Animals	500	83.3	(doll, dragon)	N.A.
	Vehicles	500	78.8	(tractor, truck)	<
	Activities	500	70.3	(barreling, traversing)	>
Currency	Dishes	500	74.3	(tiramisu, gnocchi)	>
	Accessories	500	68.9	(bra, bikini)	<
	Technology	500	76.2	(hose, sensor)	>
Length	Places	500	76.8	(bungalow, arena)	N.A.
	House Object	500	72.6	(terrace, railing)	N.A.
	Organs	499	73.2	(arm, ovary)	>
All	All	6,000	77.1	-	-

Table 3: Statistics for the new dataset, with the number of examples in each domain, their inter-annotator agreement, a representative example and the majority label assigned by annotators.