KNACK v2: Using Analogical Generalization over Qualitative Representations for Quantitative Estimation

Walker Demel, Will Hancock, & Kenneth D. Forbus
Qualitative Reasoning Group, Northwestern University
2233 Tech Drive, Evanston, IL, USA 20208
walker.demel@northwestern.edu, wwhancock@u.northwestern.edu, forbus@northwestern.edu

Abstract

One of the roles of qualitative representations is to provide context for numerical information, making explicit how it is grounded in the world. This supports tasks like quantity estimation, e.g. estimating the cost of a used bicycle by comparing it with similar items. The KNACK model (Paritosh & Klenk, 2006) used analogical retrieval of a fixed number of cases to perform such estimates. This paper describes a new algorithm, KNACK v2, which uses analogical generalization to provide a more robust notion of context for quantitative estimation. We describe how KNACK v2 works and test its performance on a dataset of country information from Wikidata, showing it is competitive with linear regression while providing explanations.

1 Introduction

Quantity estimation is integral to our everyday lives. We may estimate how long it would take to commute home if we stop at the grocery store on the way, whether we have enough fuel to drive to our destination, or how much we should charge for our used bicycle after upgrading to a new one. Solving these estimation problems typically requires some experience with similar examples as well as domain-knowledge about the world.

Before setting the asking price for our old bicycle, we need a contextual sense of what bicycles cost in our environment. We might browse online listings of used bicycles or stop by a used bike shop in our town to get a general idea of the distribution. These serve as reference points for generating our own estimate, or in this scenario, asking price.

During quantity estimation we also regularly use our domain-specific qualitative and quantitative world knowledge. For example, we know that bicycles with a sophisticated multi-gear system are more costly than those without one. The weight of the frame or the thickness of the tires may also be factors that influence our estimate.

The dominant computational model for estimating quantities is multiple linear regression, but this approach has two drawbacks. First, linear regression does not handle qualitative information adeptly. The classic workaround solution is to create one-hot dummy variables that are active when a case has a given feature and inactive when it doesn't. In our bicycle example, the presence or absence of a gear-shifting system would be represented by a 1 or a 0 in a dedicated dimension. This approach can lead to sparsity in feature vectors and subsequent overfitting. The second drawback of pure regression is its lack of explainability. A regression output is simply an intercept and a series of coefficients for associated dimensions. There is no dependency, no higher-level cognitive mechanism that guarantees a reasonable estimate, and no clear explanation for why a given estimate makes sense. Returning to our used bicycle example, negotiations over price often hinge on specific factors (e.g. fancier gear-shifting system versus more wear), so an explainable model would likely give customers more peace of mind that they are getting a fair price.

This paper describes KNACK v2, a model for quantity estimation based on qualitative representations and analogical generalization. We start by discussing relevant background, including the anchoring and adjustment psychological model of quantitative estimation, our analogical processing models, and the construction of qualitative representations of quantities via CARVE (Paritosh 2004). Then we describe the KNACK v2 algorithm, and an experiment using a dataset extracted from Wikidata (Vrandecic & Krotzsch 2014). The experiment provides evidence that KNACK v2 is competitive with linear regression, but with the ability to provide explanations. We close with conclusions and future work.

2 Background

2.1 Anchoring and Adjustment

There has been significant psychological evidence for the heuristic of *anchoring and adjustment* (Tversky and Kahneman, 1974). This method for quantity estimation involves two steps. The first step anchors an estimate by retrieving a relevant example from memory and using its value for that quantity. This retrieval can be a prototypical class instance (subject to the availability heuristic (Tversky and Kahneman 1974)) or a similar example. For instance, when estimating the rent for an apartment, we may start with the rent for apartments of the same configuration (e.g. one bedroom) in the same neighborhood. Using that sample as our estimate would be a type of nearest neighbor sampling, but we can often be more accurate by utilizing *adjustment*. This second step incorporates our intuitive heuristic knowledge of the world to scale up or down our estimate.

We use two ideas in developing computational models based on anchoring and adjustment. The first is the structure-mapping theory of analogy and similarity (Gentner, 1983) to both find similar examples and compute how they are aligned with the current situation. In structure-mapping, similarity is based on structured representations, including relationships between entities as well as attributes (aka features). There is ample evidence that this model is more psychologically plausible than purely feature-based approaches (e.g. Markman & Gentner, 1993). Returning to the rental example, when trying to estimate the rent for one apartment, we may retrieve another apartment—whose rent we do know—and map the two cases together with their relative parts, comparing configuration with configuration, location with location, price with price, etc. These alignable properties help provide the grist for adjustment: If one apartment is larger than the other, then that suggests its rent might be higher. Qualitative representations provide this kind of causal information needed to drive adjustment. We use *qualitative proportionalities* (Chapter 7, Forbus 2019), which describe how quantities are causally connected with one another. If rent is qualitatively proportional to square footage, then an apartment with more square footage will have a higher rent, all else being equal. Of course, what makes these problems difficult is that all else typically is not equal: A small apartment in a great neighborhood may be more expensive than a huge apartment in an unsafe neighborhood.

This approach is broadly compatible with psychological evidence about component processes. Previous studies have found that relational retrieval improves with domain expertise (Blanchette & Dunbar 2001; Novick 1988; Gentner, Loewenstein, & Thompson 2004). Similarly, the adjustment phase of quantity estimation gets better with expertly tuned heuristics and knowledge of qualitative proportionalities and other quantity relationships (Paritosh & Klenk 2006).

2.2 Analogical Processing

We draw on computational models for three processes involved in analogical learning and reasoning, matching, retrieval, and generalization, discussing each in turn.

Matching is performed by the Structure-Mapping Engine (SME; Forbus et.al. 2017). It takes two cases as input, both structured representations that include both statements about object attributes (e.g. being a bicycle) and relationships (e.g. that the basket is connected to the rear wheel of the bicycle). It constructs one or more *mappings*, each of which consists of three parts. (1) A numerical score indicates the overall quality of the match. This depends on properties such as the nested overlap in relationships, thereby capturing human preferences for arguments and explanations, (2) a set of correspondences, indicating what objects and statements align with each other. Correspondences can be used in supporting how an example is relevant to a situation, among other things. (3) A set of candidate inferences, indicating how non-aligned information in the base or target might be mapped onto the other description, based on the correspondences. These provide conjectures and highlight salient differences between the two descriptions.

Retrieval of cases is modeled by MAC/FAC (Forbus, Gentner & Law 1995). The probe is the case for which a reminding is sought from a case library consisting of structured representations. For scalability, MAC/FAC consists of a two-stage process, both of which use map/reduce. The MAC stage computes a coarse estimate of the probe with every case in the library, in parallel, based on *content vectors*. Content vectors are automatically constructed from structured representations, with the strength of a dimension related to the number of occurrences of each kind of predicate, attribute, or logical function. The dot product of two content vectors is an estimate of SME's structurally grounded similarity score. The best M matches from the MAC stage are passed into FAC, which uses SME for its comparisons, producing the best N matches as outputs.

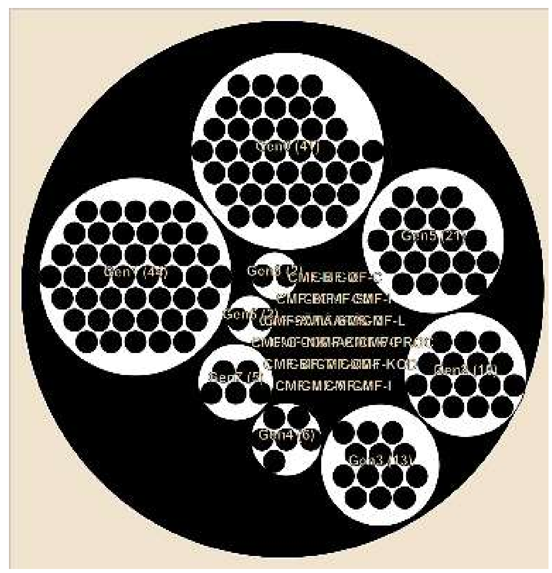

Figure 1:
A Sage Generalization Pool. Generalizations (groups of analogically similar cases) are shown in white circles, with individual constituent cases shown as black dots.

Generalization, the process by which we naturally group similar cases together, is modeled by Sequential Analogical Generalization Engine (SAGE) (Kandaswamy & Forbus 2012). SAGE builds analogical models of concepts incrementally, using structure-mapping as a clustering metric. Each model consists of a *generalization pool*, which can contain both generalizations and outliers (Figure 1). Given a new example of a concept, MAC/FAC is used to retrieve the most similar item, treating the pool as a case library. If the similarity score produced by SME is higher than the *assimilation threshold* for that generalization pool, then the case and the item are assimilated. If the item was an outlier, then a new generalization is formed by merging the corresponding statements.

Generalizations also record summative statistics about constituent cases. For example, in a generalization composed of two countries, facts they share (high population, medium GDP, etc.) will have probabilities of 1, while facts that exist in only one constituent case have probability 0.5. Non-identical entities are replaced by skolems constants called *generalized entities*. If the item was a generalization, the merge process updates the probabilities for the statements based on overlap, and introduces new generalized entities as needed. At any time, the generalizations and outliers in the pool constitute a disjunctive model of that concept given the data so far. It is analogous to k-means clustering with outliers, except that the clustering metric is structure-mapping and the number of clusters is determined automatically based on the data. This ability to handle disjunctive concepts provides a finer-grained notion of context for reasoning, e.g. racing bicycles will likely end up in different clusters from cargo bicycles.

Currently, our analogy stack (MAC/FAC, SME, and SAGE) is not sensitive to quantity; that is, the analogy stack was built primarily for cognitively plausible, qualitative reasoning over relational cases, rather than numerical analyses. In order to make analogy sensitive to attributes on quantitative dimensions, we employ a model called CARVE.

2.3 Qualitative Representation of Quantities

Structure-mapping operations are not sensitive to numerical values. For example, the difference between apartments with 700 and 705 square feet is the same to SME as the difference between apartments with 700 and 1000 square feet. We take this as a job for qualitative representations: In apartments, 5 square feet is a negligible difference. In an engineering analysis of materials needed for an aircraft, five extra square feet can be a considerable difference. Thus we argue that translation to appropriate qualitative values, in a task-specific manner, is a sensible and psychologically plausible way to incorporate such information.

In Qualitative Process (QP) theory (Forbus, 1984), limit points are used to distinguish ranges in numerical values based on when the underlying causal laws change. But what about situations where either it isn't known yet which causal laws are relevant yet, or even what they are? Paritosh (2004) proposed using *distributional limit points*, dividing numerical ranges into a discrete set of values via classic k-means clustering. For example, population might be initially divided into three bins, High, Medium, and Low. Once distributional limit points have been computed, numerical facts can automatically be converted to qualitative statements. For example,

```
(populationOfRegion unitedStatesOfAmerica
  (UnitOfCountFn Person) 331000000)
```

becomes

```
(isa UnitedStatesOfAmerica (CountryTypeFn
  (MediumAmountFn CountryPopulation)))
```

Where the literal value is replaced by the qualitative label (medium) within the broader case library context (all countries). Thus countries that are qualitatively similar in relevant dimensions are more likely to be retrieved. Significant differences in quantities are highlighted via candidate inferences generated during the mapping process.

CARVE (Paritosh 2004), uses k-means clustering to introduce distributional limit points and then used a precursor to SAGE to look for useful partitionings. At the time, the paucity of available data limited experimentation. With modern Semantic Web data sources, that has changed. The experiments described here use the CARVE algorithm with three qualitative values to symbolize quantity.

3 The KNACK v2 Algorithm

KNACK v2 is an algorithm for quantitative estimation using analogical generalization over qualitative representations. It takes a stream of examples to learn analogical models via SAGE, as described in Section 2.2. Figure 2 describes the algorithm for ingestion of new examples, and Figure 3 describes how estimations are made, given the current state of the generalization pool. We discuss each in turn.

Algorithm: Ingest Example
Given example E and generalization pool GP,

1. Convert all quantitative values in E to qualitative values
2. Add E to GP via SAGE

Figure 2: KNACK v2 Ingestion Algorithm

The example ingestion process (Figure 2) is straightforward. All statements involving numerical parameters are replaced with qualitative statements, as per the example above. This has the effect of flattening the representation to some degree, since it is replacing relations (e.g. `populationOfRegion`) with attributes (e.g. `(CountryTypeFn (MediumAmountFn CountryPopulation))`), which has the effect of making analogical retrieval sensitive to differences in values, as desired.

Quantity estimation can be viewed as a form of anchor and adjustment. Step 1 in Figure 3 retrieves the anchor. As per Step 1(a), if nothing is retrieved, the average of Q across the examples in the pool is used as a fallback. If the closest anchor is an outlier, then there isn't enough information to build a linear regression model, so the value of Q in the outlier is used instead (Step 2). Step 3 is the interesting case. As noted above, qualitative proportionalities provide the kind of partial causal constraints that can be assembled to form a model for a quantity. We assume the retrieval of relevant qualitative proportionalities (Step 3(a)) is done respecting the constraints of a QP domain theory. Steps 3(b-d) does the adjustment, by constructing and using a linear regression model based on the examples in the retrieved generalization. One subtlety concerns missing data in examples: If an example is missing data, it is thrown out, and if none of the examples in I have relevant data, the marginal average across the pool is used instead as a fallback.

The use of generalization to provide a more focused context is the key innovation of KNACK v2. The original version of KNACK used MAC/FAC over a case library of examples, looking for a hard-coded number of examples—5— to use in model construction. By using analogical generalization instead, we are assured that the cases are all reasonably similar to each other, as opposed to being just the most similar that could be found. Thus this algorithm scales smoothly between low-data situations (e.g. two examples) and high-data situations (e.g. dozens of examples). This does raise the question of what should be done with generalizations that have thousands or even millions of examples. Such situations have never arisen, but if they do, one approach would be incrementally computing more summative statistics rather than keeping everything in the original cases.

Algorithm: Estimate

Given: New example E with quantity Q to be estimated, with respect to generalization pool GP

1. Retrieve closest item I from GP, using MAC/FAC
    a. If no retrieval, return marginal average of Q across all cases in GP
2. If I is an outlier, use the value of Q in I as the estimate.
3. If I is a generalization,
    a. Let qprops = {qualitative proportionalities constraining Q}
    b. Let $a_1,\ldots,a_n$ be the antecedent quantities from qprops.
    c. Construct linear regression model from values for $a_1,\ldots,a_n$ using the cases used to produce I
    d. Produce estimate from linear model, computing Q from data for $a_1,\ldots,a_n$ from E.

Figure 3: KNACK v2 Estimation algorithm

At the time of the original KNACK's publication in 2006, the landscape of open-source datasets was very different. Prior to the machine learning boom of the 2010s, datasets for learning were more often smaller and experiment-specific. The datasets used with the original KNACK algorithm, for example, contained 15 cases (each case representing one basketball player). Datasets have ballooned in size since this time, and access is often easy and free (Forbus & Demel, 2022). Thus to test the scalability of the KNACK v2 algorithm, we generated a new dataset using Wikidata, one of the largest open knowledge graphs available.

[1] NextKB is available at qrg.northwestern.edu, and we will make the country dataset available on the web as well to support replication.

4 Experiment

We generated a dataset describing information about 197 countries, and used KNACK v2 to build models for quantity estimation. We start by summarizing Wikidata and how we translated the data into our representations[1]. Then we describe our experimental method and the results.

4.1 Wikidata

Wikidata is a collaboratively edited knowledge graph hosted by the WikiMedia foundation (Wikipedia, Wiktionary, etc.) Utilizing an extensive distributed community of editors, Wikidata has grown to over 104 billion items at the time of writing.[2] The open-source nature of Wikidata allows it to serve as a downstream aggregate of otherwise siloed data from various sources. For example, Wikidata contains data from the Google Books initiative as well as the Vatican Library, linking common entities across domains. We briefly describe the structure of Wikidata items.

Wikidata items are entities with a unique identifier (QID) and a set of statements concerning them. Each statement is a key-value pair, with the key being a property (associated with a unique property ID, or PID) and the value being some value—a quantity, another item, or multimedia like a photo. This structure is effectively a series of triples of the form <subject, predicate, attribute>. This RDF structure makes all of Wikidata queryable from a SPARQL endpoint.[3] For example, say one wants to find the capital of the United States. The United States is an item in Wikidata with the QID Q30. There is a *capital* property with the PID P36. Then all we have to do is query for the statement <Q30, P36 ?X> in SPARQL, giving us another entity, Washington D.C. (Q61).

But statements can be more sophisticated than linking multiple items. Some predicates, like *area* (P2046), link an item to a quantity, margin of error, and a unit. (According to Wikidata, the United States (Q30) has an area (P2046) of 9,826,675±1 square kilometers.) Other facts have qualifiers attached—between 1785 and 1790, the capital of the United States was New York City. Similarly, the value for a population statement is constrained by the year when it holds true. Finally, most facts in Wikidata can be traced back to their source through citations or provenance information, increasing the trustworthy of the data available.

We queried Wikidata for 197 countries and their associated statements. We gathered both qualitative data, like:

- Bordering Countries
- Continent Membership
- Currency
- Bordering Bodies of Water
- International Organization Membership
- Language Spoken

Along with quantitative data, such as

- Area
- Population
- Human Development Index (HDI)

[2] For up-to-date statistics on items, edits, and users, visit https://www.wikidata.org/wiki/Special:Statistics

[3] query.wikidata.org

- Development Index
- Gross Domestic Product (GDP)
- GDP Per Capita
- Literacy Rate
- Fertility Rate
- Life Expectancy
- Median Income
- Democracy Index.

Due to the crowdsourced nature of Wikidata[4], not all cases are complete with every dimension. For example, no literacy rate was found for Mexico, and no median income found for Mauritius. Wikidata had only 4 of 11 possible quantitative facts for Monaco: area, population, GDP, and GDP per capita. This makes our estimation task more difficult but is inevitable in real-world situations.

To build our dataset, facts retrieved via SPARQL queries were automatically translated into the OpenCyc ontology used in NextKB, our knowledge base. For example, a population fact in Wikidata looks like

```
<U.S. (Q30), Population (P1082), ~331Million>
```

This is translated to this CycL sentence:

```
(populationOfRegion unitedStatesOfAmerica
  (UnitOfCountFn Person) 331000000)
```

Since this dataset will be used for analogical estimation, we need to have an understanding of what it means for two countries to be analogically similar to one another. There are qualitative similarities: if they are a part of the same continent, in the same international organizations, use the same currency, or share cultural similarities like the language spoken. There are also quantitative similarities. They may have similar populations or areas, or their Human Development Indices may both be between 0.8 and 0.9. Consequently, we used CARVE to generate qualitative representations of quantitative dimension facts using three qualitative distinctions to generate facts like

```
(isa Poland
  (CountryTypeFn (LowAmountFn Area)))
(isa Spain
  (CountryTypeFn
    (MediumAmountFn CountryGDP)))
(isa UnitedStatesOfAmerica
  (CountryTypeFn
    (HighAmountFn CountryGDP)))
```

4.4 Experimental Method

The dataset we built contains 197 country cases, each consisting of 2 to 91 facts, with a mean of 38. Each experimental fold consisted of holding out 19 or 20 cases for testing while the model learned (generalized) the remaining ones. The assimilation threshold for SAGE during the learning phase of KNACK v2 was set at 0.8, requiring strong match strength between a test case and a given generalization. The dimensions to be estimated were GDP, Human Development Index (HDI), and Democracy Index (DI). The qualitative proportionalities involving them are shown in Table 1.

DIMENSION	DEPENDS ON
GDP	Population
HDI	Life Expectancy
DI	HDI

Table 1. World knowledge is built into the model of dimensional dependence. Dimensions in the right column were used as independent variables when regressing on a generalization.

Measuring accuracy is subtle given the varying nature of these quantities. Gross Domestic Product is an unbounded quantity that ranged from 39,000 to 19 trillion US Dollars. Accuracy for GDP was measured by distance away from ground-truth values, scaled by the magnitude of the ground truth itself.

$$|truth - estimate| \,/\, truth$$

where *truth* is the ground-truth fact, and estimate is the output generated by KNACK v2. This was done according to Weber's law (Fechner 1966), which states that perceived similarity of quantities is measured by a ratio between them, i.e. although 1,000 and 1,001 are the same distance apart as 1 and 2, the former pair is judged to be closer together because the ratio of the two is closer to 1 than the ratio of the latter pair.

Accuracy for Human Development Index and Democracy Index were measured in mean squared error since they are bounded quantities. Since HDI is measured on a 0 to 1 scale, a 0.2 estimate for 0.3 would be considered less accurate than a 0.7 estimate for 0.8.

For all three test dimensions, 10 folds were generated that contained 19 or 20 held-out test cases. Accuracy was averaged across every predicted case in every fold.

We also generated a baseline linear regression model across all cases. The linear regression estimator is run using the implementation in Python's sklearn module, using default parameters. This requires vectorizing structured knowledge from the country cases by generating a set of features from the structured facts. This was accomplished by manually creating a mapping, where each quantity type is considered a feature, and each unique qualitative attribute (e.g. currency, international association membership) is represented by a one-hot vector. Missing quantities are imputed using Python's impute function in the scipy module. This results in 883 features across the 197 country cases.

4.5 Results

Table 2 shows the results from KNACK v2 against those generated by pure linear regression. The first run of our experiment recorded accuracy only for those cases suited especially well for analogy; they mapped to a generalization and used regression to generate an estimate. The second run of our

[4] Wikidata editors often have conflicting views of correct representations. The label for *Czech Republic* (Q213) has alternated between *Czechia* and *Czech Republic* multiple times in 2023.

experiment included accuracy for cases that mapped outside of generalizations—either to an outlier or to nothing—that fell back to a baseline of sampling within a generalization or using the marginal average. This was necessary for anywhere from 0 to 4 (with an average of 0.9 cases per fold) of the 20 test cases for a cross validation fold.

Pure KNACK v2 (with thrown away estimations) performed better than Pure Regression for 2 of the 3 testing dimensions, but not significantly ($P > 0.05$). P-values are shown in the right-most column of Table 2.

	Pure KNACKv2 Accuracy	KNACKv2 + sampling Accuracy	Pure Regression Accuracy	p-val
HDI	**0.004003988**	0.00459702	0.00708158	0.19
DI	2.367816143	2.63343881	**2.08532303**	0.64
GDP*	**15.47324807**	15.2617797	60.7963312	0.18

Table 2. KNACK results compared with KNACK and sampling for a complete set of estimations.
**GDP accuracy is normalized* (|*truth – estimate*| / *truth*)

4.6 Explainability

One of the advantages of our methods as opposed to pure quantitative regression is the explainability of our models. The primary mechanism that provides this capability is the summative statistics generated by SAGE. Recall that each generalization will yield a unique linear regression during our estimation procedure, so overarching information about the generalization can help explain unique trend lines. For example, when our system predicts the HDI of Belarus, we retrieve a generalization made up of Tajikstan, Kyrgyzstan, Armenia, Pakistan, Uzbekistan, and Kazakhstan. SAGE tells us these are all located on the Asian continent, have low (as labeled by CARVE) democracy indices, land area, and GDPs. Five of the six have low populations. Four of the six are members of the Central Asian Cooperation Organization. Being able to identify these trends and patterns is insight that other tools for quantitative estimation lack.

5 Discussion & Future Work

The results show that KNACK v2 is competitive with pure linear regression. It's interesting to note that falling back to sampling an anchor country or the marginal average made the results slightly less accurate for HDI and DI, but made the prediction for GDP *more* accurate. This could be explained by cases that do not get mapped to generalizations tending to have GDPs close to the marginal average of all countries.

The results show that under the right circumstances, KNACK v2 might be a more accurate model than pure linear regression. And unlike traditional linear regression, the cases that were used to form the estimate can be traced back to their source, increasing the explainability of, and potentially trust in, its results.

We see four directions for future work. First, we need to test KNACK v2 over more datasets. For example, Wikidata provides copious information about movies and their releases, with qualitative and quantitative information that appears promising for analogical estimation. Second, we plan to experiment with ways that systems using KNACK v2 can tune it to produce more relevant results. For example, agricultural models of a country might focus on different aspects than models of its overall economy or educational system. This could be handled with different case construction strategies and accumulating models in separate generalization pools. Third, we plan to investigate the effects of incrementality on estimation, e.g. how rapidly do estimates improve? Fourth, we plan to use KNACK v2 in a number of tasks using the Companion cognitive architecture (Forbus & Hinrichs 2017), such as back of the envelope reasoning (Paritosh & Forbus, 2007; Bundy et al. 2013) but also in metacognitive reasoning within the architecture itself, e.g. estimating effort and utility of tasks.

Acknowledgments

This research was sponsored by the US Air Force Office of Scientific Research under award number FA95550-20-1-0091.

References

Blanchette, Isabelle & Dunbar, Kevin. (2001). Analogy use in Naturalistic settings: The influence of audience, emotion and goals. Memory & cognition. 29. 730-5. 10.3758/BF03200475.

Bundy, A., Sasnauskas, G., Chan, M. (2013). Solving Guesstimation Problems using the Semantic Web: Four Lessons from an Application. *Semantic Web Journal*, doi:10.3233/SW-130127.

Fechner, Gustav Theodor (1966) [First published 1860]. Howes, D H; Boring, E G (eds.). *Elements of psychophysics* [*Elemente der Psychophysik*]. Vol. 1. Translated by Adler, H E. United States of America: Holt, Rinehart and Winston.

Forbus, K. (1984). Qualitative process theory. *Artificial Intelligence*, 24, 85-168.

Forbus, K. D. (2019). *Qualitative Representations: How People Reason and Learn about the Continuous World.* MIT Press.

Forbus, K. D., & Demel, W. (2022). Integrating QR Quantity Representations with the Semantic Web: A Progress Report. *Proceedings of QR 2022*.

Forbus, K. D., Ferguson, R. W., Lovett, A., & Gentner, D. (2017). Extending SME to Handle Large-Scale Cognitive Modeling. *Cognitive Science*, 41(5), 1152–1201.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19(2), 141–205.

Forbus, K., & Hinrichs, T. (2017) Analogy and Qualitative Representations in the Companion Cognitive Architecture. *AI Magazine*.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155–170.

Gentner, Dedre & Loewenstein, Jeffrey & Thompson, Leigh. (2004). Learning and Transfer: A General Role for Analogical Encoding. J Educ Psychol. 95. 10.1037/0022-0663.95.2.393.

Kandaswamy, S. and Forbus, K. (2012). Modeling Learning of Relational Abstractions via Structural Alignment. *Proceedings of the 34th Annual Conference of the Cognitive Science Society (CogSci).* Sapporo, Japan

Markman, A., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25, 431–467

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(3), 510–520.

Paritosh, P. K. (2004). Symbolizing Quantity. *Proceedings of 2004 Annual Meeting of the Cognitive Science Society*

Paritosh, P.K. and Forbus, K.D., (2005). Analysis of Strategic Knowledge in Back of the Envelope Reasoning, In *Proceedings of the 20th National Conference on Artificial Intelligence.*

Paritosh, P. K., & Klenk, M. E. (2006). Cognitive Processes in Quantitative Estimation: Analogical Anchors and Causal Adjustment. *The Proceedings of the 28th Annual Conference of the Cognitive Science Society, Vancouver*.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases, *Science*, 185, pp 1124-1131.

Vrandecic, D. & Krotzsch, M. (2014). Wikidata: A Free *Collaborative Knowledgebase. Communications of the ACM*, 57(10):78-85, October, doi:10.1145/2629489