Estimating Numerical Attributes by Bringing Together Fragmentary Clues

Hiroya Takamura and **Jun'ichi Tsujii**
Tokyo Institute of Technology Microsoft Research Asia
takamura@pi.titech.ac.jp jtsujii@microsoft.com

Abstract

This work is an attempt to automatically obtain numerical attributes of physical objects. We propose representing each physical object as a feature vector and representing sizes as linear functions of feature vectors. We train the function in the framework of the combined regression and ranking with many types of fragmentary clues including absolute clues (e.g., A is 30cm long) and relative clues (e.g., A is larger than B).

1 Introduction

We know how large surfboards usually are and also that an inner pocket of any jacket is much smaller than a surfboard. Since we know about these numerical attributes, nobody of sound mind has probably ever tried to vainly put a surfboard into an inner pocket of a jacket. However, computers do not have comprehensive knowledge of this sort. This lack of comprehensive knowledge of the numerical attributes is one obstacle to flexible and natural man-machine communication. This work is an attempt to automatically obtain knowledge on numerical attributes so that computers can use it.

The knowledge on numerical attributes is also very useful on many other occasions. For example, it enables computers to alert their users when the users input incorrect numbers that are outside of the normal range of the attribute. In image recognition, a large red object will unlikely be recognized as a strawberry if the computer knows its normal size. In natural language processing, QA systems can use numerical knowledge to eliminate the out-of-range answer candidates to numerical questions.

A number of attempts similar to the current work have been made in some other fields such as psychology or fuzzy theory. However, such attempts heavily rely on costly experiments such as giving questionnaires to human subjects and have a problem in their scalability. In contrast, the current work attempts to use NLP techniques on large text data both online and offline in order to obtain such knowledge without relying on costly experiments such as questionnaires. A possible criticism of this project is that simply examining an existing knowledge source such as Wikipedia might accomplish this purpose without much effort. Indeed Wikipedia provides numerical information of some physical objects, but not all. For example, the Wikipedia page for *watches* provides descriptions on their function and their history, but no information on their size.

Clues to the numerical attributes are rather scattered over corpora and other linguistic resources. In a corpus, we can find informative descriptions such as "*X is 35cm tall*". We can also find text fragments suggesting an order relation between two physical objects with regard to the size as in "*X is larger than Y*", as well as implicit clues such as "*I put X into Y*", which usually means X is smaller than Y. Holonymy relations (X is a part of Y) in a thesaurus suggest an order relation in size (X is smaller than Y). Glosses in a dictionary also provide subtle clues to the sizes of entry words. Each of these clues alone is not sufficient for precisely determining the size, so we have to bring them together. We have therefore developed a mathematical model that uses these

Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL, pages 1305–1310,
Denver, Colorado, May 31 – June 5, 2015.

clues and determines the sizes of many physical objects simultaneously. The approach consists of two steps: (i) many different types of clues to the numerical attribute are collected from various linguistics resources, and (ii) those collected clues are brought together by a combined regression and ranking.

2 Related Work

Hovy et al. (2002) pointed out the importance of the knowledge on the numerical attributes in question answering. They hand-coded the possible range of a numerical attribute. Akiba et al. (2004), Fujihata et al. (2001), Aramaki et al. (2007), and Bakalov et al. (2011) made similar attempts. Their target, however, is the fixed numerical attributes of the named entities, while our target is the numerical attributes of general physical objects, not restricted to the named entities.

Davidov and Rappoport (2010) collected various types of text fragments indicating values of numerical attributes of physical objects. Our work differs from theirs in that we explore more subtle linguistic clues in addition to those used in the previous work, by using a global mathematical model that brings together all the clues.

Narisawa et al. (2013) tried to determine whether a given amount is large, small, or normal as a size of an object, making good use of clue words such as *only*; The sentence "*This laptop weighs only 0.7kg*" means that laptops are usually heavier than 0.7kg.

3 Fragmentary clues to sizes

3.1 Physical objects

We first collect physical objects, i.e., objects for which the size can be defined. However, the numerical attribute of a word depends on the sense in which the word is being used. We will therefore determine the size of each sense instead of each word. Specifically, we determine the size of each noun synset in the Japanese WordNet (Bond et al., 2009). We basically regard as physical objects the synsets that are descendants of the synset corresponding to "physical objects" (00002684-n). We filter out the physical objects that are descendants of any of the following synsets (09334396-n, 00027167-n, 09239740-n, 09287968-n, 09277686-n, 09335240-n, 04564698-n, and 03670849-n), since their sizes would be hard to define (e.g., earth, location, soil).

We further filter out approximately 400 synsets for various reasons such as ambiguity.[1]

3.2 Collecting absolute clues

We collect *absolute clues*, which indicate a value of a physical object without reference to other physical objects.

We used a search engine[2] with a query such as ' "*the size of A*" AND *meter*' (AND represents a logical conjunction) and decompose the retrieved snippets into text fragments with "..." as a delimiter. We used only the first 1,000 pages (the maximum amount allowed by the terms of use for API users) for the query comprising a pattern ("*the size of A*", "*the length of A*", or "*the height of A*") and a length unit (millimeter, centimeter, meter, or kilometer).

Note that absolute clues are corpus-based.

3.3 Collecting relative clues

We also collect *relative clues*, which suggest a numerical order relation between two physical objects, i.e., A should be larger than B. Note that holonymy and comparative sentences below are explicit relative clues as opposed to implicit relative clues that follow. Note also that holonymy is WordNet-based while comparative sentences and implicit relative clues are corpus-based.

3.3.1 Holonymy

If A is a part of B, it usually means that A is smaller than B. We can obtain such part-of (holonymy) relations from the WordNet. Specifically, for each physical object obtained in Section 3.1, we retrieve its holonymy synsets. If a synset is a holonym of another synset, it suggests that the former is larger than the latter.

3.3.2 Comparative sentences

The sentence "*the middle finger is longer than the ring finger*" suggests that the relation 'middle finger $>$ ring finger' holds for the size attribute. We collect such comparative sentences. Specifically, we search

[1] The list of those synsets and textual patterns and Japanese search keywords used in this work are available from `http://www.lr.pi.titech.ac.jp/~takamura/core9.html`.

[2] Yahoo!JAPAN API.

an n-gram corpus (Kudo and Kazawa, 2007) for the textual patterns including "*A is longer than B*".[3]

3.3.3 Implicit relative clues

People tend not to write out clues explicitly when most readers are expected to have the relevant knowledge. Since absolute clues and comparative sentences are explicit, we cannot expect a sufficient amount of such clues.

We argue that people unintentionally put many pieces of common knowledge into some specific textual patterns. The sentence "*I put my wallet into the pocket*" suggests that 'pocket > wallet' holds for the numerical attribute. We collect from the n-gram corpus such textual patterns (*A in B, put A in B, take A out of B, store A in B, put A on B, drop A from B, A go into B, and A go out of B*).

4 Bringing together the clues

4.1 Feature representation and linear model

To bring together the clues introduced in Section 3, we first represent physical objects with feature vectors and employ a linear model in which the size $f(\boldsymbol{w}, \boldsymbol{x})$ is represented as the inner product $\boldsymbol{w} \cdot \boldsymbol{x}$ of feature vector $\boldsymbol{x}$ and weight vector $\boldsymbol{w}$.

We use the following features: the synsets that are ancestors of the target synset (i.e., synsets that can be found by traversing up through hypernym-hyponym relations or instance-of relations), the synsets that the target synset is a member of (hmem in WordNet), the hypernym synsets of the target synset (hype in WordNet), the synsets that the target synset is an instance of (inst in WordNet), the synsets that the target synset has as a component (mprt in WordNet), the synsets that the target synset is a component of (hprt in WordNet), the head word in the gloss in a dictionary, and the synonyms in the target synset.

4.2 Formalization

We discuss how to estimate weight vector $\boldsymbol{w}$.

Some physical objects are given absolute clues. If multiple absolute clues are found for an object, we regard their average (actually, its logarithm) as the approximate size used for training. Since the size is a real number, the machine learning framework to be employed should be regression. Additionally, relative clues are incorporated into the training by means of ranking framework. We henceforth use the combined regression and ranking.

Our formalization is similar to the combined regression and ranking model developed by Sculley (2010). Let D_a and D_r denote respectively the training datasets consisting of absolute clues and relative clues. Each element in D_a is represented as a pair of a feature vector $\boldsymbol{x}$ and its average size y. Each element in D_r is represented as a tuple of feature vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and the order relation z; z indicates whether $\boldsymbol{x}_1$ is larger ($z = +1$), or $\boldsymbol{x}_2$ is larger ($z = -1$). We minimize the following function:

$$(1-\alpha)L_a(\boldsymbol{w}, D_a) + \alpha L_r(\boldsymbol{w}, D_r) + \frac{\lambda}{2}||\boldsymbol{w}||^2, \quad (1)$$

where α is a trade-off parameter between regression loss $L_a(\boldsymbol{w}, D_a)$ and pairwise loss $L_r(\boldsymbol{w}, D_r)$. $(\lambda/2)||\boldsymbol{w}||^2$ is the regularization term.

The regression loss $L_a(\boldsymbol{w}, D_a)$ is decomposed as

$$\frac{1}{|D_a|} \sum_{(\boldsymbol{x},y)\in D_a} l_a(y, f(\boldsymbol{w}, \boldsymbol{x})), \quad (2)$$

where $l_a(y, f(\boldsymbol{w}, \boldsymbol{x}))$ is the loss of the pair $(\boldsymbol{x}, y)$ under the model $\boldsymbol{w}$, and is represented by *squared loss* $(y - f(\boldsymbol{w}, \boldsymbol{x}))^2$, indicating the difference between the target value and the model output.

The pairwise loss $L_r(\boldsymbol{w}, D_r)$, is decomposed as

$$\frac{1}{|D_r|} \sum_{(\boldsymbol{x}_1,\boldsymbol{x}_2,z)\in D_r} l_r(z, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{w}), \quad (3)$$

where $l_r(z, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{w})$ is the loss of the tuple $(\boldsymbol{x}_1, \boldsymbol{x}_2, z)$ under the model $\boldsymbol{w}$, and is represented by *hinge loss*, $\max(0, 1 - z \cdot f(\boldsymbol{w}, \boldsymbol{x}_1 - \boldsymbol{x}_2))$.

While a single type of loss function was used for the regression loss and the pairwise loss in the previous work (Sculley, 2010), the current framework relies on two different types of loss functions, i.e., squared loss and hinge loss, so that both absolute and relative clues can be used in the model.

5 Experiments

5.1 Experimental setting

We followed the process in Section 3.1 and eliminated infrequent ones from the obtained synsets. For

[3] We also used "shorter", "larger", and "smaller".

the remaining synsets, we performed a search for absolute and relative clues and obtained 1,329 absolute clues and 7,335 relative clues. This set of relative clues stems from 848 WordNet-based clues and 6,496 corpus-based clues with a small overlap. We note that fewer than 1% of these 6,496 corpus-based clues are explicit. The synsets for which no clues are found are removed from the following process, leaving 3,598 synsets. Thoroughly using the web data might provide a larger overall amount, but the current result suggests that there are fewer absolute clues than relative ones and fewer explicit clues than implicit ones.

We evaluate the methods in two different ways. One is the difference: the sizes of the 262 randomly sampled synsets without absolute clues are manually determined, and we calculated the difference between the estimated size and the manually determined size for each of those synsets. The other is the order relation classification: the size relations of 1,152 randomly sampled pairs of synsets are manually annotated, and we employ as an evaluation metric the accuracy indicating how many of those relations are correctly predicted.

We implemented our combined regression and ranking method by modifying a package.[4] We used the logarithms of sizes as the target value. We tuned λ in Equation (1) to the value that optimizes the accuracy out of 11 values[5]: 10^{-7}, 10^{-6}, $\cdots$, 10^{3}.

We tested different numbers of absolute clues in training (namely, 300, 500, 800, 1,000) in order to examine its effect.

5.2 Results

Figure 1 shows how the average difference for each number of absolute clues changes as α in Equation (1) is varied. All types of clues and features are used for Figure 1 (a), while the clues and features extracted from WordNet except for glosses are excluded for Figure 1 (b). The latter emulates the situation where the dictionary is available, but the large-scale thesaurus such as WordNet is not. The left-most point ($\alpha = 0$) for each figure corresponds to simple regression. The curves show that the difference can be reduced by using the combined regression and ranking. The improvement is more remarkable when fewer absolute clues are used.

Size (cm)	Synset	Example word
1.35×10^{-1}	11678768-n	ovum
2.68×10^{0}	02312744-n	silkworm
3.26×10^{0}	02206856-n	bee
7.16×10^{0}	04453037-n	tooth of gear
9.09×10^{0}	03209910-n	floppy disk
1.14×10^{1}	03378442-n	foot
3.01×10^{1}	04586225-n	wind chime
3.35×10^{1}	03485794-n	hand towel
4.57×10^{1}	04590553-n	windshield
1.56×10^{2}	09189157-n	nest of hawk or eagle
1.65×10^{2}	04152829-n	screen
4.31×10^{4}	02687992-n	airport

Table 1: Sample of the estimated sizes

Similarly, Figure 2 shows how the accuracy of the order relation classification for each number of absolute clues changes as α is varied. The accuracy of the order relation classification was around 70 to 80 %. The benefit of using combined regression and ranking is more remarkable in Figure 2 (b), i.e., when the thesaurus is not available.

Table 1 shows a sample of physical objects and their estimated sizes. We can see that the overall trend of the size has been successfully captured.

We also examine some features with small or large weights in Table 2. Very small weights are given to, for example, elementary particles in the field of particle physics, hydrons, and bacteria.[6]

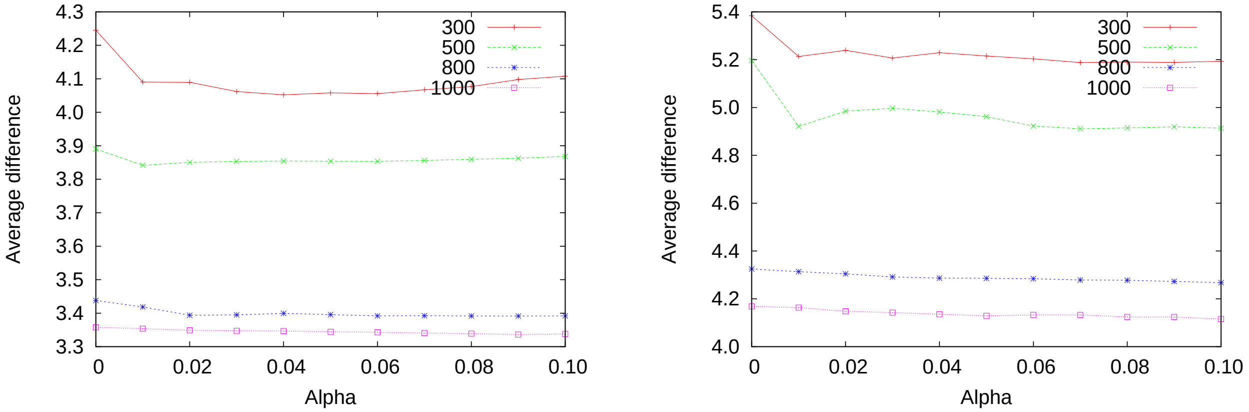
Feature	Weight
Synset for baryon, as ancestor	-7.75
Hydron as synonym	-7.75
Synset for fermion, as ancestor	-7.13
Electron, as synonym	-6.06
Bacteria, as synonym	-6.06
Bell tower, hprt feature	+7.15
Railroad as synonym,	+8.16
Means of transportation as ancestor	+8.38

Table 2: Features with large absolute weights. Note that baryon is a heavy particle in the field of particle physics.

[4] http://code.google.com/p/sofia-ml/

[5] In the actual application, we would be able to use development data for tuning.

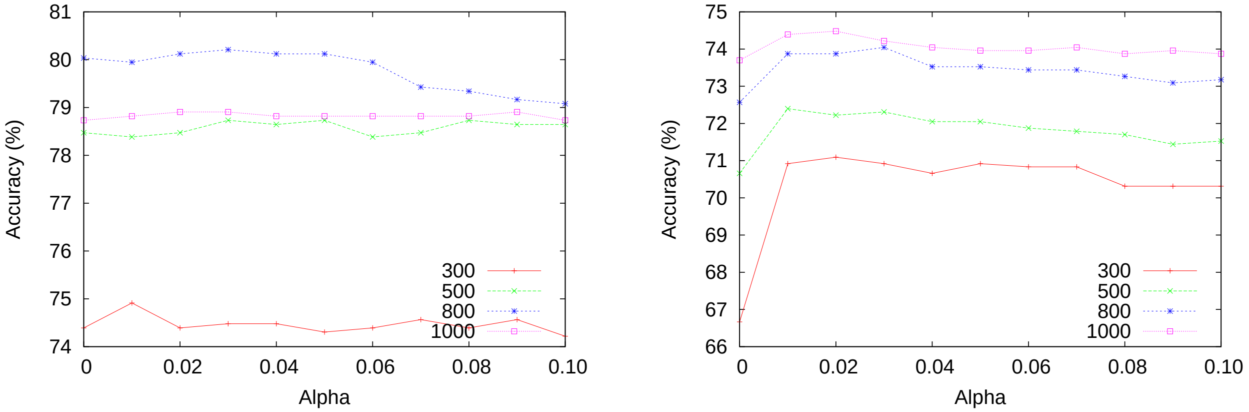
[6] More comprehensive results are available from http://www.lr.pi.titech.ac.jp/˜takamura/core9.html .



(a) All types of clues and features are used

(b) WordNet-based clues and features are excluded except for the glosses

Figure 1: Average difference between the estimated size and the manually determined size (log of centimeter).



(a) All types of clues and features are used

(b) WordNet-based clues and features are excluded except for the glosses

Figure 2: Accuracy of order relation classification

6 Conclusion

We addressed the task of automatically extracting numerical attributes of physical objects. We propose representing the sizes of objects using a linear function. We used the combined regression and ranking model with both absolute and relative clues.

Currently, many features are extracted from a thesaurus WordNet. If we can extract effective features from other resources, we would be able to apply our method to the objects that are not in the thesaurus. Future work also includes the following:

· more accurately collecting physical objects,
· sense disambiguation of words in clues,
· use of superlative sentences,
· filtering out descriptions of rare events,
· a more effective way of using glosses,
· application to other attributes, e.g., weight,
· handling idioms.

Acknowledgment

This work was partially supported by Microsoft Research (CORE Project 9).

References

Tomoyosi Akiba, Katunobu Itou, and Atsushi Fujii. 2004. Question answering using common sense and utility maximization principle. In *Proceedings of the Fourth NTCIR Workshop on Research in Information Access Technologies Information Retrieval, Question Answering and Summarization*, pages 297–303.

Eiji Aramaki, Takeshi Imai, Kengo Miyo, and Kazuhiko Ohe. 2007. UTH: SVM-based semantic relation classification using physical sizes. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 464–467, Stroudsburg, PA, USA. Association for Computational Linguistics.

Anton Bakalov, Ariel Fuxman, Partha Pratim Talukdar, and Soumen Chakrabarti. 2011. SCAD: Collective discovery of attribute values. In *Proceedings of the 20th International Conference on World Wide Web (WWW'11)*, pages 447–456.

Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. 2009. Enhancing the japanese wordnet. In *Proceedings of the 7th Workshop on Asian Language Resources (in conjunction with ACL-IJCNLP 2009).*

Dmitry Davidov and Ari Rappoport. 2010. Extraction and approximation of numerical attributes from the web. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1308–1317.

Katsuyuki Fujihata, Masahiro Shiga, and Tatsuro Mori. 2001. Extraction of numerical expressions by constraints and default rules of dependency structure. In *Special Interest Group of Information Processing Society of Japan, 2001-NL-145 (in Japanese).*

Eduard Hovy, Ulf Hermjakob, Chin-Yew Lin, and Deepak Ravichandran. 2002. Using knowledge to facilitate factoid answer pinpointing. In *Proceedings of the 19th International Conference on Computational Linguistics*, pages 369–375.

Taku Kudo and Hideto Kazawa. 2007. Japanese web n-gram corpus, version 1.

Katsuma Narisawa, Yotaro Watanabe, Junta Mizuno, Naoaki Okazaki, and Kentaro Inui. 2013. Is a 204 cm man tall or small? acquisition of numerical common sense from the web. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pages 382–391.

David Sculley. 2010. Combined regression and ranking. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 979–988, New York, NY, USA. ACM.