

Similarity Approach to Defining Basic Level of Concepts Explained from the Utility Viewpoint

Joe Lorkowski¹ and Martin Trnecka²

¹ Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, lorkowski@computer.org

² Department of Computer Science, Palacky University
Olomouc, Czech Republic, martin.trnecka@gmail.com

Abstract. In many practical situations, it is necessary to describe an image in words. From the purely logical viewpoint, to describe the same object, we can use concepts of different levels of abstraction: e.g., when the image includes a dog, we can say that it is a dog, or that it is a mammal, or that it is a German Shepherd. In such situations, humans usually select a concept which, to them, is the most natural; this concept is called the *basic level* concept. However, the notion of a basic level concept is difficult to describe in precise terms; as a result, computer systems for image analysis are not very good in selecting concepts of basic level. At first glance, since the question is how to describe human decisions, we should use notions from a (well-developed) decision theory – such as the notion of utility. However, in practice, a well-founded utility-based approach to selecting basic level concepts is not as efficient as a purely heuristic “similarity” approach. In this paper, we explain this seeming contradiction by showing that the similarity approach can be actually explained in utility terms – if we use a more accurate description of the utility of different alternatives.

1 Formulation of the Problem

What are basic level concepts and why they are important. With the development of new algorithms and faster hardware, computer systems are getting better and better in analyzing images. Computer-based systems are not yet perfect, but in many cases, they can locate human beings in photos, select photos in which a certain person of interest appears, and perform many other practically important tasks.

In general, computer systems are getting better and better in performing well-defined image understanding tasks. However, such systems are much less efficient in more open-ended tasks, e.g., when they need to describe what exactly is described by a photo.

For example, when we present, to a person, a photo of a dog and ask: “What is it?”, most people will say “It is a dog”. This answer comes natural to us, but, somewhat surprisingly, it is very difficult to teach this answer to a computer. The problem is that from the purely logical viewpoint, the same photo can be

characterized on a more abstract level (“an animal”, “a mammal”) or on a more concrete level (“German shepherd”). In most situations, out of many possible concepts characterizing a given object, concepts of different levels of generality, humans select a concept of a certain intermediate level. Such preferred concepts are known as *basic level* concepts.

We need to describe basic level concepts in precise terms. Detecting basic level concepts is very difficult for computers. The main reason for this difficulty is that computers are algorithmic machines. So, to teach computers to recognize basic level concepts, we need to provide explain this notion in precise terms – and we are still gaining this understanding.

Current attempts to describe basic level concepts in precise terms: a brief description. When we see a picture, we make a decision which of the concepts to select to describe this picture. In decision making theory, it is known that a consistent decision making can be described by *utility theory*, in which to each alternative A , we put into correspondence a number $u(A)$ called its *utility* in such a way that a utility of a situation in which we have alternatives A_i with probabilities p_i is equal to $\sum p_i \cdot u(A_i)$; see, e.g., [4, 5, 8, 10, 12].

Naturally, researchers tried to use utility theory to explain the notion of basic level concepts; see, e.g., [3, 7, 14]. In this approach, researchers analyze the effect of different selections on the person’s behavior, and come up with the utility values that describes the resulting effects. The utility-based approach describes the basic level concepts reasonably well, but not perfectly. Somewhat surprisingly, a different approach – called *similarity approach* – seems to be more adequate in describing basic level concepts. The idea behind this approach was proposed in informal terms in [13] and has been described more formally in [11]. Its main idea is that in a hierarchy of concepts characterizing a given object, a basic level concept is the one for which the degree of similarity between elements is much higher than for the more abstract (more general) concepts and slightly smaller than for the more concrete (more specific) concepts. For example, we select a dog as a basic level concept because the degree of similarity between different dogs is much larger than similarity between different mammals – but, on the other hand, the degree of similarity between different German Shepherds is not that much higher than the degree of similarity between dogs of various breeds.

In our papers [1, 2], we transformed somewhat informal psychological ideas into a precise algorithms and showed that the resulting algorithms are indeed good in detecting basic level concepts.

Challenging question. From the pragmatic viewpoint, that we have an approach that works well is good news. However, from the methodological viewpoint, the fact that a heuristic approach works better than a well-founded approach based on decision theory – which describes rational human behavior – is a challenge.

What we do in this paper: main result. In this paper, we show – on the qualitative level – that the problem disappears if we describe utility more ac-

curately: under this more detailed description of utility, the decision-making approach leads to the above-mentioned similarity approach.

What we do in this paper: auxiliary result. It is usually more or less clear how to define degree of similarity – or, equivalent, degree of dissimilarity (“distance” $d(x, y)$) between two objects. There are several possible approaches to translate this distance between *objects* into distance between *concepts* (classes of objects). We can use worst-case distance $d(A, B)$ defined as the maximum of all the values $d(x, y)$ for all $x \in A$ and $y \in B$. Alternatively, we can use average distance as the arithmetic average of all the corresponding values $d(x, y)$. In [1], we compared these alternatives; it turns out that the average distance leads to the most adequate description of the basic level concepts.

In this paper, we provide a (qualitative) explanation of this empirical fact as well.

2 Analysis of the Problem and the Resulting Solution

What is the utility associated with concepts of different levels of generality. In the ideal world, when we make a decision in a certain situation, we should take into account all the information about this situation, and we should select the best decision based on this situation.

In practice, our ability to process information is limited. As a result, instead of taking into account all possible information about the object, we use a word (concept) to describe this notion, and then we make a decision based only on this word: e.g., a tiger or a dog. Instead of taking into account all the details of the fur and of the face, we decide to run away (if it is a tiger) or to wave in a friendly manner (if it is a dog).

In other words, instead of making an optimal decision for each object, we use the same decision based on an “average” object from the corresponding class. Since we make a decision without using all the information, based only on an approximate information, we thus lose some utility; see, e.g., [9] for a precise description of this loss.

From this viewpoint, the smaller the classes, the less utility we lose. This is what was used in the previous utility-based approaches to selecting basic level concepts.

However, if the classes are too small, we need to store and process too much information – and the need to waste resources (e.g., time) to process all this additional information also decreases utility. For example, instead of coming up with strategies corresponding to a few basic animals, we can develop separate strategies for short tigers, medium size tigers, larger tigers, etc. – but this would take more processing time and use memory resources which may be more useful for other tasks. While this is a concern, we should remember that we have billions of neurons, enough to store and process huge amounts of information, so this concern is rather secondary in comparison with a difference between being eaten alive (if it is a tiger) or not (if it is a dog).

How to transform the above informal description of utility into precise formulas and how this leads to the desired explanations. The main reason for *disutility* (loss of utility) is that in a situation when we actually have an x , we use an approach which is optimal for a similar (but slightly different) object y . For example, instead of making a decision based on observing a very specific dog x , we ignore all the specifics of this dog, and we make a decision based only on the fact that x is a dog, i.e., in effect, we make a decision based on a “typical” dog y .

The larger the distance $d(x, y)$ between the objects x and y , the larger this disutility U . Intuitively, different objects within the corresponding class are similar to each other – otherwise they would not be classified into the same class. Thus, the distance $d(x, y)$ between objects from the same class are small. We can therefore expand the dependence of U on $d(x, y)$ in Taylor series and keep only the first few terms in this dependence. In general, $U = a_0 + a_1 \cdot d + a_2 \cdot d^2 + \dots$. When the distance is 0, i.e., when $x = y$, there is no disutility, so $U = 0$. Thus, $a_0 = 0$ and the first non-zero term in the Taylor expansion is $U \approx a_1 \cdot d(x, y)$.

Once we act based on the class label (“concept”), we only know that an object belongs to the class, we do not know the exact object within the class. We may have different objects from this class with different probabilities. By the above property of utility, the resulting disutility of selecting a class is equal to the *average* value of the disutility – and is, thus proportional to the *average distance* $d(x, y)$ between objects from a given class. *This explains why average distance works better than the worst-case distance.*

When we go from a more abstract concept (i.e., from a larger class) to a more specific concept (i.e., to a smaller class of objects), the average distance decreases – and thus, the main part U_m of disutility decreases: $U'_m < U_m$. However, as we have mentioned, in addition to this main part of disutility U_m , there is also an additional secondary (smaller) part of utility $U_s \ll U_m$, which increases when we go to a more specific concept: $U'_s > U_s$.

On the qualitative level, this means the following: if the less general level has a much smaller degree of similarity (i.e., a drastically smaller average distance between the objects on this level), then selecting a concept on this less general level drastically decreases the disutility $U'_m \ll U_m$, and this decrease $U_m - U'_m \gg 0$ overwhelms the (inevitable) increase $U'_s - U_s$ in the secondary part of disutility, so that $U' = U'_m + U'_s < U_m + U_s = U$. On the other hand, if the decrease in degree of similarity is small (i.e., $U'_m \approx U_m$), the increase in the secondary part of disutility $U'_s - U_s$ can over-stage the small decrease $U'_m - U_m$.

A basic level concept is a concept for which disutility U' is smaller than for a more general concept U and than for a more specific concept U'' . In view of the above, this means that there should be a drastic difference between the degree of similarity U'_m at this level and the degree of similarity U_m at the more general level – otherwise, on the current level, we would not have smaller disutility. Similarly, there should be a small difference between the degree of similarity at the current level U'_m and the degree of similarity U''_m at the more specific

level – otherwise, on the current level, we would not have smaller disutility. *This explains the similarity approach in utility terms.*

Acknowledgments. We acknowledge support by the Operational Program Education for Competitiveness Project No. CZ.1.07/2.3.00/20.0060 co-financed by the European Social Fund and Czech Ministry of Education; it was partly performed when Martin Trnecka was visiting University of Texas at El Paso.

References

1. R. Belohlavek and M. Trnecka, “Basic level of concepts in formal concept analysis”, In: *Proceedings of ICFCA '12*, Springer Verlag, Berlin, Heidelberg, 2012, pp. 28–44.
2. R. Belohlavek and M. Trnecka, “Basic level in formal concept analysis: interesting concepts and psychological ramifications”, *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence IJCAI'2013*, Beijing, China, 2013, pp. 1233–1239.
3. J. E. Corter and M. A. Gluck, “Explaining basic categories: Feature predictability and information”, *Psychological Bulletin*, 1992, Vol. 111, No. 2, pp. 291–303.
4. Fishburn P C. *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
5. Fishburn P C. *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.
6. D. H. Fisher, “Knowledge acquisition via incremental conceptual clustering”, *Machine Learning*, 1987, Vol. 2, No. 2, pp. 139–172.
7. F. Gosselin and P. G. Schyns, “Why do we SLIP to the basic level? Computational constraints and their implementation”, *Psychological Review*, 2001, Vol. 108, No. 4, pp. 735–758.
8. Keeney R L and Raiffa H. *Decisions with Multiple Objectives*, John Wiley and Sons, New York, 1976.
9. J. Lorkowski and V. Kreinovich, “Likert-Scale Fuzzy Uncertainty from a Traditional Decision Making Viewpoint: It Incorporates Both Subjective Probabilities and Utility Information”, *Proceedings of the Joint World Congress of the International Fuzzy Systems Association and Annual Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS'2013*, Edmonton, Canada, June 24–28, 2013, pp. 525–530.
10. Luce R D and Raiffa R. *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
11. G. L. Murphy, *The Big Book of Concepts*, MIT Press, Cambridge, Massachusetts, 2002.
12. Raiffa H. *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.
13. E. Rosch, “Principles of categorization”, In: E. Rosch and B. B. Lloyd, editors, *Cognition and Categorization*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1978, pp. 27–48.
14. M. D. Zeigenfuse and M. D. Lee, “A comparison of three measures of the association between a feature and a concept”, In: L. Carlson, C. Holscher, and T. F. Shipley, editors, *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, Austin, Texas, 2011, pp. 243–248.