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

Joe Lorkowski and Martin Trnecka

Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
lorkowski@computer.org and martin.trnecka@gmail.com

COPROD'2013

Basic Level Concepts: an idea.

- There has been progress in well-defined computerized tasks like finding similar images.
- There is not as much progress at more open-ended tasks like describing exactly what is described by an image.
- Example:
  - Show a person an image of a German Shepherd and ask, "What is it?"
  - Most people naturally say, "It's a dog."
  - People naturally prefer "dog" to "mammal" or "German Shepherd."
- This basic level concept is difficult to describe in precise terms.

Which level of concept is the best to label an object?

- There are usually many levels of concept (levels of generality) characterizing an object.
- Levels of concept are usually arranged hierarchically:
  - there are levels with broad generality (e.g., Animal)
  - there are levels with less generality (e.g., Dog)
  - and levels with even less generality (e.g., German Shepherd)
- People naturally select one of the intermediate levels to describe as the *basic level*.
- Current computer programs cannot select such human-natural levels.
- So, we need to describe basic level concepts in precise terms.

Towards Use of Utility Theory to describe Basic Concept

- Utility theory is used in Decision Making to describe rational human behavior.
  - to each alternative A, we assign a number $u(A)$ called its *utility*
  - the utility of a situation with alternatives A_i, with probabilities p_i, is equal to $\sum p_i \cdot u(A_i)$
- A natural idea is to try to describe basic concepts using utility theory.
- This approach describes basic level concepts reasonably well but far from perfectly.
- A different approach – called *similarity approach* – seems more adequate.

Use of Similarity Approach to describe Basic Concept

- The main idea of this approach is to use an object's hierarchy of concepts and choose one level for which:
  - the level on one side has more abstract concepts with a *much lower* degree of similarity between elements and
  - the level on the other side has a *slightly higher* degree of similarity between elements.
- In our example:
  - the elements of "mammal" (including dogs and dolphins) have a *much lower* degree of similarity than those of "dog,"
  - the elements of "dog" have a *slightly lower* degree of similarity than those of "German Shepherd."
- So, we have a heuristic that works better than existing utility-based approach to describe the basic level concept.
- Since rational decision making is described by utilities, how can we describe this heuristic in terms of utility?

Auxiliary Result

- Similarity can be described in terms of degree of *dissimilarity*; distance $d(x, y)$ between objects.
- Degree of *dissimilarity* between concepts can be defined as:
  - worst-case distance $d(X, Y) = \max_{x \in X, y \in Y} d(x, y)$

    OR
  - average distance $d_{AVG}(X, Y) = \frac{1}{|X| \cdot |Y|} \cdot \sum_{x \in X} \sum_{y \in Y} d(x, y)$.
- It turns out that the average distance leads to more adequate description of the Basic Level Concept.

Humans Make Non-Optimal Decisions

- In practice, humans' decision-making abilities are limited (knowledge, ability, time).
- Some decisions favor using a concept rather than complete information (size, temperament, diet, ...)
  - we quickly assess the danger level of the concept "tiger" as very high, and,
  - we quickly assess the danger level of the concept "dog" as low.

  So, a good (non-optimal) decision can be made based on the "average" object from the corresponding class.
- But, non-optimal decisions imply some loss of utility.
- But, optimal decisions can require too much resources.
- How much of non-optimality is good enough?

Notion of Disutility Needed

- Disutility is the loss of utility.
- Disutility emerges when we have an object x and use an approach which is optimal for a *similar* object y.
- Determining complete information for a specific tiger enables a wealth of information:
  - but you could get eaten while making the determination, so,
  - we ignore the specifics of the particular tiger, and,
  - we assess danger based on a typical tiger (x) rather than the specific tiger (y).
- In this case, disutility is proportional to the distance $d(x, y)$ between the objects x and y.
  - The smaller the distance $d(x, y)$, the smaller the disutility U.

Describe Similarity Approach using Disutility U

- The disutility U of similar objects in the same class is very small.
- The disutility U is dependent on the distance between the objects $d(x, y)$.
- We can expand the dependence in a Taylor series and keep the first few terms
  - In general, $U = a_0 + a_1 \cdot d + a_2 \cdot d^2 + ...$
  - When $x = y$, the distance is zero so the disutility is 0 and $a_0 = 0$.
  - Thus, the first non-zero term is $U \approx a_1 \cdot d(x, y)$.

Describe Similarity Approach using Disutility U

- Once we select a class label ("concept")
  - we don't know the exact object within the class or the probability
  - we only know the class to which the object belongs
  - so the disutility of selecting a class is the *average distance* $d_{AVG}(x, y)$ between the objects of the class.
- This explains why average distance works better then worst-case.

Describe Similarity Approach using Disutility U

- When we go from a more abstract concept to a more specific concept:
  - the average distance between objects decreases
  - so the main part of disutility U_{main} decreases
  - but a secondary (smaller) part of utility $u_{sec} \ll U_{main}$ *increases*
- If the more specific concept (with U_{main}) has a drastically smaller average distance than the more abstract concept (with U'_{main})
  - there is a drastic decrease in disutility ($U_{main} - U'_{main} \gg 0$) and
  - the decrease overwhelms the (inevitable) increase in utility ($u'_{sec} - u_{sec}$) in the secondary part.
- However, if the decrease in the average distance is small ($U_{main} \approx U'_{main}$)
  - there is a small decrease in disutility and
  - the decrease is over-staged by the increase in utility.

Result

- A basic level concept U
  - has a much larger degree of similarity than that of the more general concept U' on one side and
  - has a slightly smaller degree of similarity than the more specific level U'' on its other side.
- In terms of disutility, $U'_{main} \gg U_{main} \approx U''_{main}$
- This explains the similarity approach in utility terms.

Conclusion

- When an optimal decision is not possible, a satisfactory decision *may* be possible based on a basic level concept.
- A description of basic level concept using utility terms is not sufficient.
- A new method, similarity approach, describes basic level concept very well.
- Similarity approach can be described in terms of utility.