The Illumination Model for Nearest Neighbor Classification

Michael E. HOULE
National Institute of Informatics

Why

The disciplines of machine learning and data mining continue to grapple with fundamental issues in the area of knowledge representation. Many important tasks in data analysis, such as similarity search, classification, and clustering, depend on the interplay between data features and similarity measures.

What

This presentation introduces a new model of nearest neighbor classification that starts from the premise that featurization should be scale invariant. Under this premise, the influence of individual training points on the classification can be shown to resemble many physical phenomena, most notably the way in which light sources combine to illuminate objects.

Featurization

NN CLASSIFICATION

- Traditional K-NN classification: K most similar objects influence the classification of the test object.
- The featurization of the data and the similarity measure are both taken as given.
- Open issues:
  - How big should K be?
  - What weighting should be given to the influence of each neighbor?

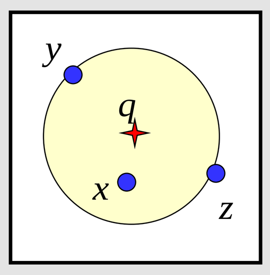
FEATURES & SIMILARITY

- Role of features: competing influences in the valuation of object-to-object similarity.
- Distance (or similarity) functions always implicitly determine a relative weighting among the features.
- Feature set + distance function → designer's policy on relative influence of object attributes.
- Can't change this policy!

SIMILARITY & SCALE

- Relative weighting is important:
  - Ratio of contribution of feature values to similarity.
  - Ratio of contribution of similarity values to classification scores.
- Scaling of feature values:
  - Leads to proportional scaling of the similarity values.
  - In general, no effect on relative contributions of features to similarity.
  - Not accounted for by modelers.

DISCRIMINABILITY



- $d(q,x) < d(q,y) \Rightarrow$ x should have greater influence on y in the classification of q.
- $d(q,y) = d(q,z) \Rightarrow$ q cannot distinguish y from z.
- Scaling of feature values should not affect the classification decision.
  - Note that distributional classification methods are affected by scaling of features.

Illumination Model

BASIC CRITERIA

- Contribution of training point to classification criterion: influence.
- Influence should be isotropic and monotone:

$$d(x,y) = d(a,b) \Rightarrow I(x,y) = I(a,b)$$
$$d(x,y) < d(a,b) \Rightarrow I(x,y) > I(a,b)$$

- Isotropy and monotonicity force influence to change with scaling of distance values (and feature values).

SCALE INVARIANCE

- Relative to q, sets of indistinguishable points form equivalence classes:

$$E(q,r) = \{x \in S \mid d(q,x) = r\}$$

- Equivalence classes can be thought of as an indivisible entity.
- Influence over equivalence classes should then be scale invariant:

$$\oint_{E(q,r)} I(q,x)\,dx = \oint_{E(q,r')} I(q,y)\,dy = G$$

INFLUENCE FORMULA

- Due to isotropy, influence over members of an equivalence class is constant.
- Basic criteria determines the form that influence can take.

$$G = \oint_{E(q,r)} I(q,x)\,dx = I(q,x) \oint_{E(q,r)} dx$$
$$\Rightarrow I(q,x) = G \Big/ \oint_{E(q,r)} dx$$

PHYSICAL INTERPRETATION

- In m-dimensional Euclidean space:

$$\oint_{E(q,r)} dx = \phi(\pi, m) \cdot r^{m-1} \Rightarrow$$
$$I(q,x) = \frac{G}{\phi(\pi,m) \cdot d^{m-1}(q,x)} = \frac{G'}{d^{m-1}(q,x)}$$

- In 3-dimensional space, inverse square law.
- Can also be applied to other spaces, both continuous and discrete.

COMBINING INFLUENCES

- Each influence carries class information with it.
- The greater the number of influencing objects from a given class, the stronger the total influence for that class.
- Influences additive by class:

$$I(q,C) = \sum_{x \in C} I(q,x) = G' \sum_{x \in C} \frac{1}{d^{m-1}(q,x)}$$

- Physical analogue: illumination with training objects as light sources.
- Majority vote: class with largest total influence wins. (All items participate!)
- Other combination strategies are possible.

EFFECT OF DIMENSIONALITY

- Consider ordered list of objects ranked in terms of distance to q:

$$q: \quad x_0, x_1, x_2, \ldots, x_i, \ldots$$

- Can normalize influences by dividing each by the influence of the 1-NN object.

$$I(q,\bullet): \quad \frac{G'}{d^{m-1}(q,x_0)}, \ldots, \frac{G'}{d^{m-1}(q,x_i)}, \ldots$$
$$\Downarrow$$
$$I_0(q,\bullet): \quad 1, \ldots, \left(\frac{d(q,x_0)}{d(q,x_i)}\right)^{m-1}, \ldots$$

- For $m > 1$, the influences of farther neighbors diminishes faster as the dimensionality increases.
- When the normalized influences are sufficiently small, further neighbors need not be evaluated.

INTRINSIC DIMENSIONALITY

- As m tends to infinity, the illumination model decision tends to that of 1-NN classification.
- Farther neighbors generally have significant (normalized) influence only when the dimension is small.
- In the vicinity of a given test point q, only a subset of the features can be considered relevant.
- Idea: substitute m by the intrinsic dimension in the vicinity of q, as estimated from the training set data.
- Can be estimated in many ways:
  - PCA.
  - Generalized expansion dimension.
  - Others.

SUMMARY

- A NN classification variant that automatically determines (based on distance values and dimensionality):
  - an appropriate neighborhood size.
  - a natural distance-based weighting scheme.
- Takes effects of dimensionality into account in a natural way.
- Derivable from principles of isotropy, monotonicity, equivalence, and scale invariance.

NII 連絡先： Michael E. HOULE（フール　マイケル）／ 国立情報学研究所　客員教授
TEL : 03-4212-2538　　Email : meh@nii.ac.jp