Approximating discontinous functions

Egon Elbre, Olga Agen
Department of Computer Science
University of Tartu
Tartu

January 3, 2013

Abstract

Calculating discontinous functions can be difficult and computationally expensive. By combining using approximation functions we can lower the computational expensiveness at the cost of errors. Defining good approximation function can be a hard to get right. However, defining an approximation that works for some value can be simpler. We show how combining several simple approximations can give us a better approximation and reduce computational requirements.

1 Introduction

Outline

2 Theory

We are trying to approximate a function f:

$$f : X \mapsto \mathbb{R}$$

Although we explain the idea in $\mathbb{R}$ set, the results also apply to any set that has transitive relation $<$ and $+-$ operation (TODO: find the correct algebraic structure). Using $<$ we can define min and max for that set.

$$\min(x, y) := \begin{cases} x & \text{if } x \leq y, \\ y & \text{otherwise.} \end{cases}$$

$$\max(x, y) := \begin{cases} y & \text{if } x \leq y, \\ x & \text{otherwise.} \end{cases}$$

X can be any set.

2.1 Exact approximation

Let's assume we are interested in range $R \subseteq X$. Let's assume we have functions α and β such that

$$\alpha(x) \leq f(x), \forall x \in R$$
$$\beta(x) \leq f(x), \forall x \in R$$

Now we can define ϵ function.

$$\alpha(x) = f(x) + \varepsilon_\alpha, \varepsilon_\alpha \geq 0$$
$$\beta(x) = f(x) + \varepsilon_\beta, \varepsilon_\alpha \geq 0$$

It is trivial to derive function γ where ε_γ is smaller than ε_α and ε_β.

$$\begin{aligned} \gamma(x) &= max(\alpha(x), \beta(x)) \\ &= max(f(x) - \varepsilon_\alpha(x), f(x) - \varepsilon_\beta(x)) \\ &= f(x) - min(\varepsilon_\alpha(x), \varepsilon_\beta(x)). \end{aligned}$$

This also gives us a usefulness requirement for α and β:

$$\exists x \in R,\ \alpha(x) < \beta(x)$$
$$\exists x \in R,\ \beta(x) < \alpha(x)$$

This means that function α and β must be complementary. For some inputs one should give better approximations than the other.

2.2 Probablisitic approximation

Having such hard boundary severly restricts the possible functions we can use. If we allow some mistake in the boundary we can get better precision, but we won't know to which side we will make errors.

Depending of our use case it may be acceptable.

3 Distance approximation

The same idea can be used in distance measures.

$$f(x, y) = max(g(x, y), h(x, y))$$

Since most distance measures are required to calculate between multiple elements we can first do a transformation into a metric space and then use the same prinicple to extract the distance.

$$f(x, y) = max(dist_1(t_1(x), t_1(y)), dist_2(t_2(x), t_2(y)))$$

This way we can reuse computation of t_1 or even do additional operations on it. For example it can allow us to do range querys.

$$query(x, R) = \{y | dist_1(t_1(x), t_1(y)) <= R\} \cap \{y | dist_2(t_2(x), t_2(y)) <= R\}$$

Where w is some function to adjust for the transformations. If our distance function is simple, such as Manhattan or Euclidian, we can use already existing range query algorithms and data structures.

4 Example: Levenshtein on Nucleotide Sequences

Although the theory is straightforward the complexity arises from finding the appropriate functions to combine.

As the basis we use hamming, fourier, haar transformation.

One interesting aspect of these transformations is that are linear. This means we can use it on two chromosomes and expect similar measure. Although this also means we can't attach cost to cross-overs.

$$F(af + bg) = aF(f) + bF(g)$$