Why Normalized Difference Vegetation Index (NDVI)?

Francisco Zapata[1], Eric Smith[1]
Vladik Kreinovich[1], and Hoang Phuong Nguyen[2]

[1]University of Texas at El Paso, El Paso, Texas 79968, USA
fcozpt@outlook.com, esmith2@utep.edu, vladik@utep.edu
[2]Division Informatics, Math-Informatics Faculty
Thang Long University, Nghiem Xuan Yem Road, Hoang Mai District
Hanoi, Vietnam, nhphuong2008@gmail.com

1. Formulation of the problem

- Plants play a very important role in ecological systems – they transform CO_2 into oxygen.
- It is therefore very important to be able to estimate the overall amount of live green vegetation in a given area.
- The most efficient way to provide such a global analysis is:
  - to use remote sensing,
  - i.e., to use multi-spectral photos taken from satellites, drones, planes, etc.
- At present:
  - one of the most efficient ways to detect, based on remote sensing data, how much live green vegetation an area contains
  - is to compute the value of the normalized difference vegetation index (NDVI):

$$\text{NDVI} \stackrel{\text{def}}{=} \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}.$$

2. Formulation of the problem (cont-d)

- Here Red and NIR are spectral reflectance measurements corresponding to red and near infrared (NIR) parts of the spectrum.
- Why this particular combination of these two values is the most adequate?
- In this talk, we provide a theoretical explanation for the above formula.

3. Different scales

- When we process data, we process numerical values of the physical quantities.
- These numerical values depend not only on the quantity itself, they depends also on:
  - what measuring unit we select, and
  - what starting point we select for measurement.
- For example, to measure temperature, we can use Celsius (C) and Fahrenheit (F) scales.
- These scales use different measuring units – a 1-degree difference in the C scale corresponds to 1.8 degree difference in the F scale.
- These scales also use different starting points: 0 degrees on a C scale correspond to 32 degrees on the F scale.

4. Different scales (cont-d)

- In general:
  - if we replace the original measuring unit by a new unit which is a times smaller,
  - then all numerical values x get multiplied by a: $x \mapsto a \cdot x$.
- For example, if we replace meters by centimeters, then, e.g., the original height of 1.7 m becomes $1.7 \cdot 100 = 170$ cm.
- Similarly:
  - if we replace the original starting point with a new point which is b degrees smaller,
  - then to all numerical values we add this value b: $x \mapsto x + b$.
- Thus, in general, if we change both the measuring unit and the starting point, we get a linear transformation $x \mapsto a \cdot x + b$.

5. Nonlinear rescalings

- In many practical situations, we can also have nonlinear rescalings.
- For example, instead of describing the electric properties of an object, we can use resistance R or we can use conductivity $1/R$.
- Which of such nonlinear transformations are natural?
- In general, linear transformations are natural.
- If we have a natural transformation from scale A to scale B, then the inverse transformation – from scale B to scale A – is also natural.
- Also:
  - if we have a natural transformation from scale A to scale B and another natural transformation from scale B to scale C,
  - then their composition is a natural transformation from scale A to scale C.
- Thus, the class of all natural transformations is closed under composition and under taking an inverse.

6. Nonlinear rescalings (cont-d)

- In mathematical terms, this means that this class should be a *transformation group*.
- At any given moment of time, we can store and process only finitely many values; thus:
  - to effectively deal with different natural transformations,
  - we should require that a natural transformation should be uniquely determined by the values of finitely many parameters.
- In mathematical terms, this means that the class of all natural transformations should be *finite-dimensional*.

7. Nonlinear rescalings (cont-d)

- It is known that for transformations from real numbers to real numbers:
  - each finite-dimensional transformation group containing all linear transformations
  - contains only fractional linear transformations.
- In other words, all natural transformations should be fractional linear, i.e., have the form $x \mapsto \dfrac{a \cdot x + b}{c \cdot x + d}$ for some a, b, c, d.

8. Natural functions of two variables

- For a function $z = f(x, y)$ of two variables:
  - if we fix x, then we get a transformation $y \mapsto f(x, y)$, and
  - if we fix y, then we get a transformation $x \mapsto f(x, y)$.
- It is reasonable to say that a function $f(x, y)$ is natural if all these transformations are natural – and thus, fractional linear.
- Let us characterize all such functions.

9. First Result: Characterizing All Natural Functions of Two Variables

- Let us recall that a function $f(x, y)$ is called *bilinear* if it has the form

$$f(x, y) = a_0 + a_1 \cdot x + a_2 \cdot y + a_{12} \cdot x_1 \cdot x_2.$$

- In these terms, we have the following result.
- We say that a function $f(x, y)$ is *natural* if the following two conditions are satisfied:
  - for every x, the mapping $y \mapsto f(x, y)$ is fractional linear, and
  - for every y, the mapping $x \mapsto f(x, y)$ is fractional linear.
- **Proposition.** *A function $f(x, y)$ is natural if and only if it is a ratio of two bilinear functions.*

10. Proof

- If we fix the value of one of the variables, then a bilinear function becomes linear.
- So, the ratio of two bilinear functions becomes fractional linear.
- Thus, a ratio of two bilinear functions is natural in the sense of our Definition.
- So, to prove the proposition, it is sufficient to prove that every natural transformation is a ration of two bilinear functions.
- Indeed, naturalness means, in particular, that for every y, there exists values $a(y)$, $b(y)$, $c(y)$, and $d(y)$ for which

$$f(x,y)=\frac{a(y)\cdot x+b(y)}{c(y)\cdot x+d(y)}.$$

11. Proof (cont-d)

- In the generic case, $d(y)\ne 0$, so we can divide both numerator and denominator by $d(y)$ and get a simpler expression

$$f(x,y)=\frac{A(y)\cdot x+B(y)}{C(y)\cdot x+1},\ \text{where}$$

$$A(y)\stackrel{\rm def}{=}\frac{a(y)}{d(y)},\quad B(y)\stackrel{\rm def}{=}\frac{b(y)}{d(y)},\ \text{and}\ C(y)\stackrel{\rm def}{=}\frac{c(y)}{d(y)}.$$

- The case when $d(y)\equiv 0$ can be treated similarly.
- Since the function $f(x,y)$ is natural, for each x, the expression $f(x,y)$ is a fractional linear function of y.
- In particular, if we select three generic different values x_1, x_2, and x_3, we conclude that

$$f_i(y)=\frac{A(y)\cdot x_i+B(y)}{C(y)\cdot x_i+1},\ \text{where}\ f_i(y)\stackrel{\rm def}{=}f(x_i,y)\ \text{is fractional linear.}$$

12. Proof (cont-d)

- Multiplying both sides of this formula by the denominator, we conclude that

$$C(y)\cdot x_i\cdot f_i(x)+f_i(y)=A(y)\cdot x_i+B(y),\ \text{i.e., that}$$

$$x_i\cdot A(y)+B(y)-f_i(y)\cdot x_i\cdot C(y)=-f_i(y).$$

- So, for each y, we have a system of three linear equations to determine the three unknowns $A(y)$, $B(y)$, and $C(y)$.
- According to Cramer's rule, the solution to a system of linear equations is a ratio of two determinants.
- The determinants are polynomials in terms of the coefficients.
- In mathematics, ratios of polynomials are called rational functions.
- So, the solution to a system of linear equations is a rational function of all the coefficients.
- In our case, x_i, 1, and -1 are constants, and $f_i(y)\cdot x_i$ is a rational function.

13. Proof (cont-d)

- So, each of the coefficients $A(y)$, $B(y)$, and $C(y)$ is a rational function of a rational function – and thus, a rational function itself.
- In other words, for some polynomials $P_A(y)$, $Q_A(y)$, $P_B(y)$, $Q_B(y)$, $P_C(y)$, and $Q_C(y)$, we have:

$$A(y) = \frac{P_A(y)}{Q_A(y)}, \quad B(y) = \frac{P_B(y)}{Q_B(y)}, \quad C(y) = \frac{P_C(y)}{Q_C(y)}.$$

- Substituting these expressions into the formula for $f(x, y)$, we get:

$$f(x, y) = \frac{\dfrac{P_A(y)}{Q_A(y)} \cdot x + \dfrac{P_B(y)}{Q_B(y)}}{\dfrac{P_C(y)}{Q_C(y)} \cdot x + 1}.$$

14. Proof (cont-d)

- By using the formulas for addition and division of fractions, we get:

$$f(x,y) = \frac{\dfrac{P_a(y)\cdot Q_B(y) + x\cdot P_B(y)\cdot Q_A(y)}{Q_A(y)\cdot Q_B(y)}}{\dfrac{x\cdot P_C(y) + Q_C(y)}{Q_C(y)}} =$$

$$\frac{P_A(y)\cdot Q_B(y)\cdot Q_C(y) + x\cdot P_B(y)\cdot Q_A(y)\cdot Q_C(y)}{x\cdot Q_A(y)\cdot Q_B(y)\cdot P_C(y) + Q_A(y)\cdot Q_B(y)\cdot Q_C(y)}.$$

- In other words, the function $f(x,y)$ is a ratio of two polynomials which are linear in x.
- These two polynomials may have a non-constant common divisor, which is either linear in x or does not depend on x at all.
- Dividing both numerator and denominator by the greatest common divisor of these two polynomials, we conclude that

  $$f(x,y) = \frac{a_0(y) + x\cdot b_0(y)}{x\cdot c_0(y) + d_0(y)}$$ for some polynomials $a_0(y), b_0(y), c_0(y), d_0(y)$.

15. Proof (cont-d)

- Similarly, from the fact that for each y, the mapping $x \mapsto f(x,y)$ is fractional linear, we conclude that

$f(x,y) = \dfrac{A_0(x) + y \cdot B_0(x)}{y \cdot C_0(x) + D_0(y)}$ for some polynomials $A_0(x), B_0(x), C_0(x), D_0(x)$.

- Equating the right-hand sides of the two formulas for $f(x,y)$, we conclude that

$$\frac{a_0(y) + x \cdot b_0(y)}{x \cdot c_0(y) + d_0(y)} = \frac{A_0(x) + y \cdot B_0(x)}{y \cdot C_0(x) + D_0(x)}.$$

- Multiplying both sides of this equality by both denominators, we conclude that

$$(a_0(y)+x\cdot b_0(y))\cdot(y\cdot C_0(x)+D_0(x)) = (x\cdot c_0(y)+d_0(y))\cdot(A_0(x)+y\cdot B_0(x)).$$

- The left-hand side of this equality is divisible by $a_0(y) + x \cdot b_0(y)$.
- This means that the right-hand side must be divisible by the same expression.

16. Proof (cont-d)

- By our construction of the sum $a_0(y) + x \cdot b_0(y)$, this sum has no common factors with $x \cdot c_0(y) + d_0(y)$.
- Thus, it must divide $A_0(x) + y \cdot B_0(x)$.
- Similarly, the right-hand side of the above equality is divisible by $A_0(x) + y \cdot B_0(x)$, and thus, it must divide $a_0(y) + x \cdot b_0(y)$.
- So, the polynomials $a_0(y) + x \cdot b_0(y)$ and $A_0(x) + y \cdot B_0(x)$ must divide each other.
- Thus, they should differ only by a multiplicative constant c:

$$a_0(y) + x \cdot b_0(y) = c \cdot A_0(x) + y \cdot c \cdot B_0(x) \text{ for all } x \text{ and } y.$$

- In particular, for two different values x_1 and x_2, we conclude that

$$a_0(y) + x_i \cdot b_0(y) = c \cdot A_0(x_i) + y \cdot c \cdot B_0(x_i).$$

- So, we have two linear equations with constant coefficients for determining two unknowns $a_0(y)$ and $b_0(y)$.

17. Proof (cont-d)

- Thus, the solution is a linear combination of the right-hand sides.
- Right-hand sides are linear functions of y, so both $a_0(y)$ and $b_0(y)$ are linear functions of y:

$$a_0(y) = a_{00} + a_{01} \cdot y, b_0(y) = a_{10} + a_{11} \cdot y, \text{ for some } a_{ij}.$$

- Thus, the numerator of the expression for $f(x, y)$ has the form

$$a_0(y) + x \cdot b_0(y) = a_{00} + a_{10} \cdot x + a_{01} \cdot y + a_{11} \cdot x \cdot y.$$

- In other words, this numerator is a bilinear function.
- Similarly, we can conclude that the denominator of the expression for $f(x, y)$ is a bilinear function:

$$x \cdot c_0(y) + d_0(y) = c_{00} + c_{10} \cdot x + c_{01} \cdot y + c_{11} \cdot x \cdot y \text{ for some } c_{ij}.$$

- Thus, the function $f(x, y)$ is indeed equal to the ratio of two bilinear functions:

$$f(x, y) = \frac{a_{00} + a_{10} \cdot x + a_{01} \cdot y + a_{11} \cdot x \cdot y}{c_{00} + c_{10} \cdot x + c_{01} \cdot y + c_{11} \cdot x \cdot y}. \text{ Q.E.D.}$$

18. Comment

- A similar result can be similarly proven for a function of several variables.
- Let us recall that a function $f(x_1, \ldots, x_n)$ is called *multi-linear* if it has the form

$$f(x_1, \ldots, x_n) = \sum_{S \subseteq \{1, \ldots, n\}} a_S \cdot \prod_{i \in S} x_i.$$

- In these terms, we have the following result.
- We say that a function $f(x_1, \ldots, x_n)$ is *natural* if for all i and for all possible values $x_1, \ldots, x_{i-1}, \ldots, x_{i+1}, \ldots, x_n$, the mapping

$$x_i \mapsto f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \text{ is fractional linear.}$$

- **Proposition.** *A function $f(x_1, \ldots, x_n)$ is natural if and only if it is a ratio of two multi-linear functions.*

19. Scale invariance

- The values Red and NDVI depend on the solar angle.
- For a different angle, the same situation leads to the values Red and NDVI both multiplied by the same constant.
- We want a characteristic that does not depend on the solar angle.
- So, we want to select a characteristic $f(\text{NIR}, \text{Red})$ that should not change if we simply change the angle:

$$f(c \cdot x, c \cdot y) = f(x, y) \text{ for all } x, y, c.$$

- Such functions are known as *scale-invariant.*
- The following result characterizes all scale-invariant natural functions.
- **Proposition.** *A natural function $f(x, y)$ which is not identically constant is scale-invariant if and only if it has the form*

$$f(x, y) = \frac{a_{10} \cdot x + a_{01} \cdot y}{c_{10} \cdot x + c_{01} \cdot y}.$$

20. Proof

- One can easily prove that each expression of this type is scale-invariant.
- Vice versa, let us assume that a natural function is scale-invariant.
- We know, a general expression for all natural functions.
- Thus, scale invariance means that for all x and y, we have

$$f(x,y)=\frac{a_{00}+a_{10}\cdot x+a_{01}\cdot y+a_{11}\cdot x\cdot y}{c_{00}+c_{10}\cdot x+c_{01}\cdot y+c_{11}\cdot x\cdot y}=$$

$$\frac{a_{00}+c\cdot a_{10}\cdot x+c\cdot a_{01}\cdot y+c^2\cdot a_{11}\cdot x\cdot y}{c_{00}+c\cdot c_{10}\cdot x+c\cdot c_{01}\cdot y+c^2\cdot c_{11}\cdot x\cdot y}.$$

- To prove the proposition, we need to prove that $a_{00}=c_{00}=0$ and that $a_{11}=c_{11}=0$.
- Let us prove that $a_{00}=c_{00}=0$.

21. Proof (cont-d)

- Indeed, if at least of these two coefficients is different from 0, then:
  - for $c \to 0$, the right-hand side of the above equality becomes a constant a_{00}/c_{00},
  - so we conclude that the function $f(x, y)$ is constant.
- Similarly, if at least one of the coefficients a_{11} and c_{11} is different from 0, then:
  - in the limit $c \to \infty$, the right-hand side of the above equality becomes a constant a_{00}/c_{00},
  - so we conclude that the function $f(x, y)$ is constant.
- So, if the function $f(x, y)$ is not constant, all these four coefficients should be equal to 0.
- Thus, the Proposition is proven.

22. Discussion

- We have almost got the desired NDVI expression.
- To get even closer to the expression (1), let us take into account that the NDVI index changes from -1 to 1.
- If we require that the range of the function is exactly $[-1, 1]$, then we get the following result.
- **Proposition.** *A natural scale-invariant function $f(x, y)$ whose range of values for $x, y \geq 0$ is $[-1, 1]$ has the form*

$$f(x, y) = \frac{x - c \cdot y}{x + c \cdot y} \text{ or } f(x, y) = -\frac{x - c \cdot x}{x + c \cdot y}.$$

23. Proof

- If we divide both the numerator and the denominator of the general scale-invariant expression by x, we get
$$f(x,y) = F(z) \stackrel{\text{def}}{=} \frac{a_{10} + a_{01} \cdot z}{c_{10} + c_{11} \cdot z}, \text{ where } a = \frac{y}{x}.$$
- The fact that the range of this function is between -1 and 1 means that this function is defined for all z.
- In this case, the fractional linear function is monotonic.
- So its range when $z \in [0, \infty]$ is simply an interval bounded by the values of this function at the endpoints $z = 0$ and $z = \infty$.
- So, the fact that the range is equal to $[-1, 1]$ means that one of the values $F(0)$ and $F(\infty)$ should be equal to 1, and another value to -1.
- If $F(0) = -1$, then for its opposite $G(z) \stackrel{\text{def}}{=} -F(z) = -f(x,y)$, we have $G(0) = -F(0) = 1$.
- So, without losing generality, we can consider the case when $F(0) = 1$.

24. Proof (cont-d)

- Substituting the value $z = 0$ into the right-hand side of the above formula, we conclude that $a_{10}/c_{10} = 1$, i.e., that $a_{10} = c_{10}$.
- Dividing both the numerator and the denominator by this common value $a_{10} = c_{10}$, we conclude that

$$F(z) = \frac{1 + a \cdot z}{1 + c \cdot z}, \text{ where } a = \frac{a_{01}}{a_{10}} \text{ and } c = \frac{c_{01}}{c_{10}}.$$

- This ratio has to be defined for all z, so we must have $c \geq 0$ – otherwise, if $c < 0$, this expression would not be defined for $z = -1/c$.
- For $z = \infty$, we have $F(\infty) = -1$, so $a/c = -1$, thus $a = -c$.
- The proposition is proven.

25. Conclusion

- We (almost) explained why NDVI is a relevant characteristics.
- Similar argument can – partly – explain the effectiveness of fractional-linear similarity coefficients like Jaccard index

$$s(A,B)=\frac{\mu(A\cap B)}{\mu(A)+\mu(B)-\mu(A\cap B)},\ \text{where } \mu(S) \text{ is the measure of the set } S.$$

26. Acknowledgments

- This work was supported in part by the National Science Foundation grants:
  - 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and
  - HRD-1834620 and HRD-2034030 (CAHSI Includes).
- It was also supported by the AT&T Fellowship in Information Technology.
- It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.