Bilinear Models from System Approach Justified for Classification, with Potential Applications to Bioinformatics

Richard Aló[1], François Modave[2]
Vladik Kreinovich[2], David Herrera[2], Xiaojing Wang[2]

[1]Center for Computational Sciences & Advanced Distributed Simulation,
University of Houston-Downtown, One Main Street,
Houston, TX 77002, USA, RAlo@uh.edu
[2]Department of Computer Science, University of Texas at El Paso,
El Paso, TX 79968, USA, vladik@utep.edu

1. System Approach: In Brief

- *Problem:* often, we do not know the exact dynamics $\dot{x}_i = f_i(x_1, \ldots, x_n)$.
- *Solution:* build the expressions for f_i based on common sense (von Bertalanffy, Forrester, Meadows).
- *Case 1:* increase in x_i slows down the growth of x_j.
- *Suggestion:* f_j includes a term $-k \cdot x_i$ with $k > 0$.
- *Case 2:* x_j and x_k, when combined, enhance the growth of x_i.
- *Suggestion:* f_i includes $+k \cdot x_j \cdot x_k$.
- *Comment:* values of k are determined empirically.
- Common sense rarely goes beyond simple interaction, so we have *bilinear* f_i.
- *Successes:* good *qualitative* predictions.

2. Systems Approach in Classification

- *Examples:*
  - separate stocks with a good growth potential from the risky ones, or
  - separate cancerous cells from the normal ones.
- *Idea:* use a discrimination function $f(x_1, \ldots, x_n)$:
  - objects with $f > 0$ belong to the first class, and
  - objects with $f < 0$ belong to the second class.
- *Common sense:* leads to bilinear f.
- *Unexpected phenomenon:*
  - for system dynamics, we had *qualitative* predictions;
  - for classification, we have good *quantitative* fit.
- *Our objective:* explain this mystery.

3. Approximations of Different Order of Accuracy: General Idea

- *Assumption:* the actual (unknown) discrimination function $f(x_1, \ldots, x_n)$ is smooth.
- *Conclusion:* in the neighborhood U of a point $\widetilde{x} = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$, we keep only lower order Taylor terms.
- *Different points* $\widetilde{x}$*:*
  - if $\widetilde{x}$ is in class 1, then U is in class 1;
  - if $\widetilde{x}$ is in class 2, then U is in class 2;
  - *conclusion:* the problem is interesting only when $\widetilde{x}$ is on the border, i.e., $f(\widetilde{x}_1, \ldots, \widetilde{x}_n) = 0$.
- *Simplification:*
  - *Idea:* take new coordinates $x_i \to x_i - \widetilde{x}_i$ in which the starting point is $(0, \ldots, 0)$.
  - *Result:* starting point is 0, and $f(0, \ldots, 0) = 0$.

4. First and Second Approximations

- *General idea:* $f(0, \ldots, 0) = 0$.
- *First approximation:* linear discrimination function
$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i \cdot x_i.$$
- *Second approximation:* quadratic function
$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i \cdot x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \cdot x_i \cdot x_j.$$
- *Beyond bilinear:* this general expression:
  - has linear terms $a_i \cdot x_i$;
  - has bilinear terms $a_{ij} \cdot x_i \cdot x_j$ for $i \neq j$;
  - also has purely quadratic (not bilinear) terms $a_{ii} \cdot x_i^2$ (corresponding to $i = j$).

5. For System Dynamics, Bilinear Functions Provide a Rather Crude Approximation

- *Fact:* in linear approximation, we ignore quadratic (and higher order) terms.
- *Accuracy of linear approximation:* quadratic in x_i.
- *Accuracy of quadratic approximation:* cubic in x_i.
- *Accuracy of bilinear approximation:*
  - *fact:* we ignore quadratic terms $a_{ii} \cdot x_i^2$;
  - *conclusion:* accuracy is quadratic in x_i.
- *Bad news:* same asymptotic as linear.
- *Good news:*
  - *linear approximation:* we ignore n^2 terms $a_{ij} \cdot x_i \cdot x_j$, so accuracy is $n^2 \cdot \delta$;
  - *bilinear approximation:* we ignore n terms $a_{ii} \cdot x_i^2$, so accuracy is $n \cdot \delta \ll n^2 \cdot \delta$.

6. The Fact that We Are Interested in Classification Applications Allows Further Simplifications

- *In dynamics applications:* the function $f(x_1, \ldots, x_n)$ can be determined from observations.
- *In the classification applications:* we only observe the *signs* of $f(x_1, \ldots, x_n)$.
- *Conclusion:* a new function $f'(x_1, \ldots, x_n)$ with the same signs leads to the same classification:

$$f(x_1, \ldots, x_n) > 0 \text{ if and only if } f'(x_1, \ldots, x_n) > 0;$$

$$f(x_1, \ldots, x_n) > 0 \text{ if and only if } f'(x_1, \ldots, x_n) > 0.$$

- *What we plan to do:* we prove that
  - for every quadratic function $f(x_1, \ldots, x_n)$,
  - there is a bilinear function $f'(x_1, \ldots, x_n)$ with the same signs.

7. Explanation of Bilinear Functions

- *Idea:* $f'(x_1,\ldots,x_n)=f(x_1,\ldots,x_n)\cdot\left(1+\sum_{j=1}^n b_j\cdot x_j\right)$.
- *Good news:* for $x\approx 0$, $\left|\sum_{i=j}^n b_j\cdot x_j\right|\ll 1$, hence $f(x_1,\ldots,x_n)$ and $f'(x_1,\ldots,x_n)$ have the same sign.
- *General case:* $f(x_1,\ldots,x_n)=\sum_{i=1}^n a_i\cdot x_i+\sum_{i=1}^n\sum_{j=1}^n a_{ij}\cdot x_i\cdot x_j$, hence

$$f'(x_1,\ldots,x_n)=f(x_1,\ldots,x_n)+\left(\sum_{i=1}^n a_i\cdot x_i\right)\cdot\left(\sum_{j=1}^n b_j\cdot x_j\right).$$

- *Conclusion:* for $b_i=-\dfrac{a_{ii}}{a_i}$, we have a bilinear f'.
- *Comment:* this is only possible in the *generic* case, when $a_i\neq 0$ for all i.

8. Discussion: A Similar Simplification Is Not Always Possible for Higher Order Models

- *We proved:* quadratic f can be reduced to bilinear f'.
- *Natural question:* can we reduce cubic f to trilinear

$$f' = \sum_i a_i \cdot x_i + \sum_{i,j} a_{ij} \cdot x_i \cdot x_j + \sum_{i,j,k} a_{ijk} \cdot x_i \cdot x_j \cdot x_k?$$

- *Answer:* no, even for $n = 3$:
  - to describe a general trilinear function, we need 7 parameters a_1, a_2, a_3, a_{12}, a_{13}, a_{23}, and a_{123};
  - a general cubic f can be described by a discriminating curve $x_3 = F(x_1, x_2)$, where a general cubic

$$F(x_1, x_2) = b_1 \cdot x_1 + b_2 \cdot x_2 + b_{11} \cdot x_1^2 + b_{12} \cdot x_1 \cdot x_2 + b_{22} \cdot x_2^2 +$$

$$b_{111} \cdot x_1^3 + b_{112} \cdot x_1^2 \cdot x_2 + b_{122} \cdot x_1 \cdot x_2^2 + b_{222} \cdot x_2^3$$

    requires $9 > 7$ parameters.

9. Auxiliary Question: Can we Get a Further Reduction from Bilinear Functions?

- *To describe a bilinear function:* of n variables, we need n coefficients a_i and $\dfrac{n \cdot (n-1)}{2}$ coefficients a_{ij} $(i \neq j)$.
- *Total number of parameters:* $\dfrac{1}{2} \cdot (n^2 + n)$.
- *Generic second-order classification:* separating surface $x_n = F(x_1, \ldots, x_{n-1})$, with

$$F(x_1, \ldots, x_{n-1}) = \sum_{i=1}^{n} b_i \cdot x_i + \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} b_{ij} \cdot x_i \cdot x_j.$$

- *To describe F:* we need $2 \cdot (n-1)$ parameters b_i and b_{ii} and $\dfrac{(n-1) \cdot (n-2)}{2}$ parameters b_{ij} $(i \neq j)$.
- *Total number of parameters:* $\dfrac{1}{2} \cdot (n^2 + n) - 1$.

10. Auxiliary Question: Can we Get a Further Reduction from Bilinear Functions? (cont-d)

- *To describe a bilinear function:* of n variables, we need $\frac{1}{2} \cdot (n^2 + n)$ parameters.
- *To describe a general quadratic discrimination:* we need $\frac{1}{2} \cdot (n^2 + n) - 1$ parameters.
- *Conclusion:* there is only one extra parameter in the bilinear expression.
- *Reduction:* $a_1 = \pm 1$, by taking

$$f'(x_1, \ldots, x_n) = \frac{1}{|a_1|} \cdot f(x_1, \ldots, x_n).$$

- *Comment:* clearly, $f(x_1, \ldots, x_n)$ and $f'(x_1, \ldots, x_n)$ have the same signs.
- *Conclusion:* no further reduction is possible.

11. Acknowledgments

This work was supported in part by:

- NASA under cooperative agreement NCC5-209,
- NSF grants EAR-0225670 and DMS-0532645,
- Star Award from the University of Texas System, and
- Texas Department of Transportation grant No. 0-5453.