

OPTIMIZING CHEMICAL SENSOR ARRAY SIZES

G. C. Osbourn, R. F. Martinez, J. W. Bartholomew, W. G. Yelton, A. J. Ricco*

Sandia National Laboratories, Albuquerque, NM 87185-1423, USA

*ACLARA Biosciences, Inc., 3906 Trust Way, Hayward, CA 94545-3716, USA

RECEIVED
JUN 09 1999
OSTI

ABSTRACT

Optimal selection of array sensors for a chemical sensing application is a nontrivial task. It is commonly believed that "more is better" when choosing the number of sensors required to achieve good chemical selectivity. However, cost and system complexity issues point towards the choice of small arrays. A quantitative array optimization is carried out to explore the selectivity of arrays of partially-selective chemical sensors as a function of array size. It is shown that modest numbers (dozens) of target analytes are completely distinguished with a range of arrays sizes. However, the array selectivity and the robustness against sensor sensitivity variability are significantly degraded if the array size is increased above a certain number of sensors, so that relatively small arrays provide the best performance. The results also suggest that data analyses for very large arrays of partially-selective sensors will be optimized by separately analyzing small sensor subsets.

BACKGROUND

Optimal selection of sensors for array applications is desirable to minimize cost and to maximize performance. An unambiguous way to carry out such optimization is to compare the numerical pattern recognition scores of alternative arrays to directly discover those combinations of sensors that provide the best selectivity. The visual empirical region of influence (VERI) pattern recognition technique, used in a leave-one-out mode, has recently been shown to be particularly well suited for this combinatorial array optimization task (1,2). These studies provide direct performance metrics that enable appropriate design decisions. Previous applications of VERI to array optimization have examined relatively small numbers of sensors. The present study extends this previous work to a larger set of sensors to study the selectivity of partially-selective sensor arrays as a function of array size. Array optimization is carried out here for 29 chemical sensors. Specific optimization results are presented for an array intended to distinguish: water, 16 individual volatile organic compounds (VOCs) and one VOC mixture (kerosene). The sensors are surface acoustic wave (SAW) devices coated with a variety of films that are intended to exhibit differing affinities for the VOCs. These films are described elsewhere (3,4). Training data were obtained for a wide range of concentrations for each analyte.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

UPPER BAND ARRAY OPTIMIZATION METHOD

For the present case of 29 available sensors, a full combinatorial comparison is not feasible because there are too many sensor combinations to examine. The number of array combinations is smallest at the limits of small arrays (using only a few sensors) and large arrays (using all but a few sensors), and is largest for array sizes that use half of the available sensors. It is always feasible to exactly examine all array combinations for the smallest (in the present case, arrays with 5 or fewer sensors) and the largest arrays (here, 25 or more sensors). For arrays of intermediate size, a technique called the "upper band" optimization is used to find small groups of candidate arrays that give the best performance. Good candidate arrays with $N+1$ sensors can be developed from an "upper band" of best arrays of size N (e.g. the top 100 arrays) simply by adding each of the remaining sensors to each of the upper band array combinations. The exact results for arrays with 5 sensors provide the basis for a good candidate set of arrays with 6 sensors. The chemical selectivities of these candidate arrays with $N+1$ sensors are then evaluated using VERI, and the upper band of these arrays are then used to build arrays with $N+2$ sensors, and so on. In this way, good arrays with 6,7,...,14 sensors are successively discovered. Similarly, large candidate arrays with $M-1$ sensors can be developed from the upper band of arrays with M sensors simply by removing each of the sensors, one at a time, from the upper band array combinations. In this way, good arrays with 24,23,...,15 sensors are discovered based on the exact results from the 25 sensor arrays.

The appropriateness of this method is supported in two ways. First, previous VERI optimization studies (unpublished) of smaller sensor sets have directly compared the upper band optimization results to the exact combinatorial results. The true optimal arrays were found by the upper band method for all array sizes in the comparison. Second, arrays with essentially perfect selectivity are discovered in the present study.

SENSOR SENSITIVITY VARIABILITY

A key ingredient to array optimization studies is often the inclusion of some form of response variability into the data set. This is especially necessary when many arrays can provide 100% selectivity for the raw data. Addition of variability to the data spreads out the array response patterns for each analyte, and can mimic changes in sensor response strengths over the life of the array. Increasing the variability ultimately causes response patterns of differing analytes to overlap and become indistinguishable. The most robust arrays are those that retain the greatest performance for large, added response variability to all sensors. Because such arrays provide the greatest separation of the analytes in the multivariate space of array response vectors, they also are expected to minimize false alarms due to overlap of an unknown analyte with one of the training set analytes. When many arrays exhibit perfect selectivity for low sensor variability, the optimal arrays are the highest performers found at the response variability which is just large enough to reduce all array selectivities below 100%.

The variability model used here is the same as in a previous study (1). Sensors are assumed to lose sensitivity with time, so that the future responses of the sensor are related to the initial response by a multiplicative constant (independent of analyte). A simulated data set is created for each analyte by adding random values of sensitivity loss to each sensor. The sensitivity loss is bounded by a chosen ceiling value (e.g. up to 10% loss) for all analytes in the data set. This loss ceiling is used to control overall array variability.

RESULTS

The main results are summarized in Fig. 1. The two "upper band" optimizations shown in Fig. 1 were carried out with response sensitivity loss ceilings of 20% and 35%, respectively. The symbols give the largest percent correct obtained from the VERI leave-one-out analysis as a function of the number of sensors in the array. The results for up to 20% sensitivity loss show that many array sizes can yield 100% chemical selectivity for this example. It is clearly necessary to have an array that is not too small for this case.

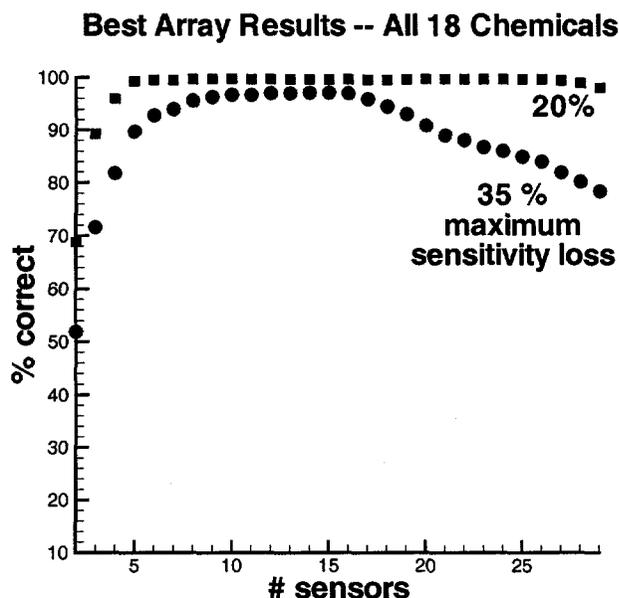


Fig. 1 Combined chemical recognition performance on the 18 analytes for the best sensor array combination as a function of array size. The two results are for sensitivity losses up to (20%, 35%) for any sensor, respectively.

The larger value of response variability was added as described above to identify the arrays that provide the most distinguishable patterns of responses for these analytes. All of the arrays now fall below 100% selectivity, and those with the largest scores provide the best chemical selectivity. Arrays in the size range of 8-15 sensors are best for the present set of analytes. At larger array sizes, the added effect of variability in the added response channels dominates, so that chemical selectivity decreases with array size and potential false alarm rates are expected to increase.

Further details of the 35% sensitivity loss ceiling results are shown in Fig. 2. This figure shows the percent correct for each analyte as a separate line for different array sizes. The loss of performance can now be seen to occur in stages as different groups of analytes begin to overlap with increasing array size. The three chemicals that overlap at the smallest array sizes are all of the aliphatic hydrocarbons in this analyte set. These chemically similar analytes give array response patterns that are closest together in the multivariate space of array response vectors. A second group of analytes exhibits overlap at somewhat larger array sizes. These chemicals are all of the chlorinated hydrocarbons in this analyte set. Increasing the array size beyond the optimal range is seen here to cause a loss in "resolution" of the analyte patterns for large sensor variabilities, so that the selectivity among the most chemically similar analytes is impacted first.

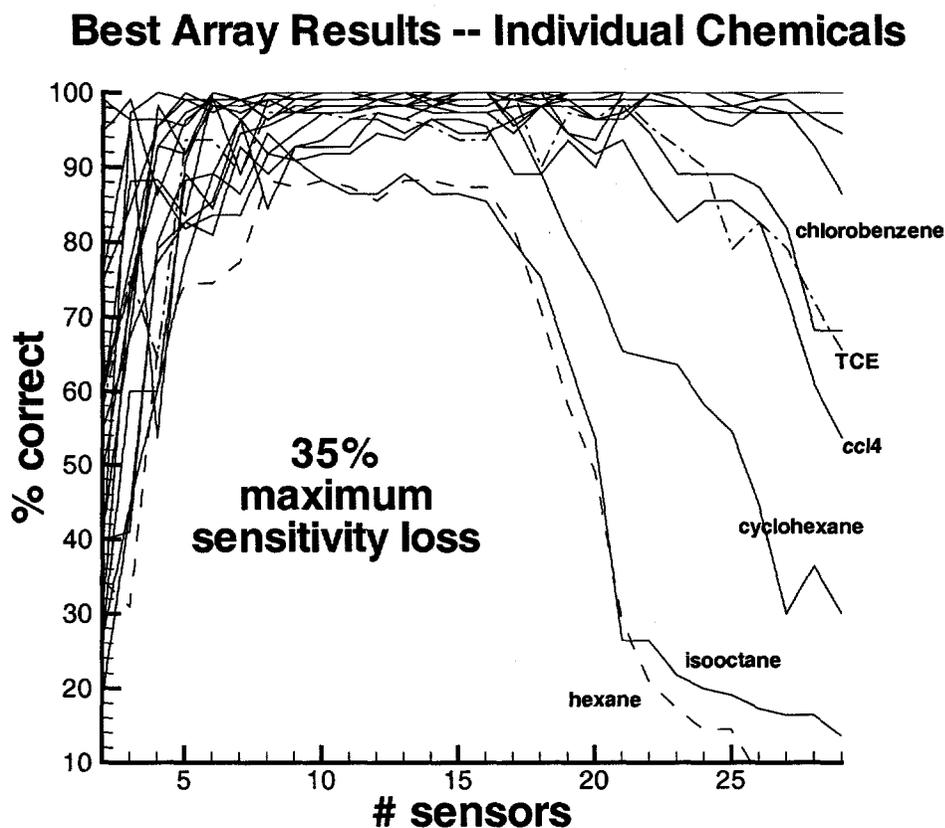


Fig. 2 Chemical recognition performance on each of the individual 18 analytes for the best sensor array combination as a function of array size. Results are for the case of sensitivity losses up to 35% for any sensor.

It is clear for this set of analytes that a smaller array gives better selectivity as well as offering lower system costs. For much larger analyte sets that might require many more sensors, these results argue against analyzing all sensors as a single, large group. Alternatively, data analyses for such cases should combine the results from multiple, small subsets of sensors in the larger set. For example, the analytes could be analyzed hierarchically, with small subarrays analyzed first to distinguish broader analyte classes (e.g., distinguish alcohols from aliphatic hydrocarbons), followed by analysis of other small subarrays to distinguish the members of these classes (e.g. distinguish hexane from isooctane).

CONCLUSIONS

A quantitative array optimization has been carried out to explore the selectivity of arrays of partially-selective chemical sensors as a function of array size. These results for volatile organic compounds indicate that only applications with very large numbers of analytes should consider the use of large (more than 15 sensors) arrays. The results also suggest that very large analyte sets are best analyzed by grouping a large set of sensors into small, optimal subsets.

ACKNOWLEDGMENT

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

REFERENCES

1. G. C. Osbourn, J. W. Bartholomew, A. J. Ricco and G. C. Frye, *Acc. Chem. Res.* 31,297 (1998). See also <http://www.sandia.gov/1100/1155Web/users.htm>.
2. G. C. Osbourn and R. F. Martinez, *Pattern Recog.* 28, 1793 (1995).
3. A. J. Ricco, R. M. Crooks and G. C. Osbourn, *Acc. Chem. Res.* 31, 289 (1998).
4. W.G. Yelton, A.J. Ricco, and A.W. Station, 1998 Sensor and Actuator Workshop, Hilton Head Island, SC; also see proceedings of this ECS mtg.