DOES FUSION-ON-THE-FLY REALLY FLY?

Roland Soong
Michelle de Montigny

This paper presents a quick data fusion algorithm (known as predictive isotonic fusion) that is customized on a case-by-case basis. The accuracy of this data fusion for target group ratings was compared against a commercially available syndicated data fusion. The authors found that there was no negative trade-offs from the much faster execution times; in fact, there were significant improvements in some cases. Furthermore, this data fusion method can accommodate many more predictor/matching variables which makes even larger improvements possible.

INTRODUCTION

Data fusion is the practice by which two or more respondent-level databases are brought together to form a single respondent-level database that contains all the previously separate information. Data fusion products are usually produced on a syndicated basis, whereby the fusion database is produced once and for all and issued to all subscribers.

Syndicated data fusion takes a one-size-fits-all approach. There is a sentiment which prefers to have fusions that are customized for specific problems, under the reasonable belief that they might be superior optimal solutions. However, it is also preferred that these customized fusions must be executed rapidly in an interactive environment. Such fast, customized fusions are often referred to as 'fusion on the fly.'

There is no lack of ideas for 'fusion on the fly' but there is not much empirical data on performance. This paper presents an open-source 'fusion on the fly' algorithm, and its performance on target group ratings will be compared against a commercially available syndicated data fusion product.

DESCRIPTION OF SYNDICATED DATA FUSION

In this paper, our interest is in comparing a syndicated data fusion product against a fusion-on-the-fly product on the same database. The most prevalent form of syndicated data fusion is the (TAM+TGI)-like fusion. On one side, we have a television audience measurement (TAM) people meter panel. On the other side, we have a Target Group Index (TGI) consumer survey of media and product usage behavior. The respondents from the TAM and TGI databases are matched to each other based upon the similarity on common variables (such as age, sex, geography, television viewing, etc). The fusion database is a static respondent-level database, where the 'respondents' now carry information from both databases.

There are many ways to conduct (TAM+TGI) fusion. If the objectives are to preserve the TAM and TGI sample sizes and to preserve the media currency values, then there is a well-defined and elegant open-source formulation known as constrained statistical matching (Soong and de Montigny 2001) which is based upon solving the transportation problem in the field of operations research. Syndicated (TAM+TGI) fusion products based upon constrained statistical matching have been produced in Argentina, Brazil, Colombia, Mexico, Puerto Rico and the United States.

The syndicated fusion products are standardized products, so that all subscribers receive the identical fusion databases. They are constructed through the collaboration of the fusion specialists with the original media

research suppliers so that the integrity of the original databases are maintained. Constrained statistical matching will typically take hours to execute, and therefore cannot be executed in an interactive environment. The syndicated fusion product is based upon an omnibus, one-size-fits-all approach, since it is impossible to anticipate all the ways in which the many subscribers might use the database.

DESCRIPTION OF PREDICTIVE ISOTONIC FUSION

There are many variations of 'fusion-on-the-fly' (for example, Czaia 1992, Raimondi and Santini 1997). To distinguish our version from others, we coin the term 'predictive isotonic fusion' here. The terms 'predictive' and 'isotonic' will be clarified in our discussion.

As we see it, here are the requirements:

1. The fusion should be optimized for a specific target group which is defined 'on the fly.' The target group definitions are potentially complex, such as 'young mothers who have purchased non-prescription drugs for their children' or 'professionals/managers who have traveled overseas for business at least three times in the last 12 months' and cannot be pre-listed and processed in advance.
2. The fusion should be executed in sufficiently quick time in an interactive environment. That means not more than a few seconds in elapsed time.
3. The fusion should preserve the media currencies and target group incidences in the original databases.
4. The fusion should preserve the sample sizes of the original databases.

Requirement #1 is a given fact which we cannot change, so it remains for us to devise a fusion algorithm that is fast and accurate. Constrained statistical matching is a computationally hard problem because we are attempting to match people in high-dimensional space (that is, along dozens of common variables). This was deemed necessary because no other choice is apparent. But ideally, we would have preferred to match on the target group information – that is, we match the target group people in the TAM database with the target group people in the TGI database. Unfortunately, the problem was precisely that the TAM database does not have the target group information.

But we can induce the target group information on the TAM database by mapping the high-dimensional space of common variables onto a one-dimensional space of target group propensity score. Such an approach is based upon the abstract construct of fibre bundle topology (Steenrod 1950) and is

used extensively in the low-dimensional visualization of high-dimensional data (Butler and Bryce 1992).

We divide our description into two parts. In the first part, we deploy a predictive model to obtain a predicted score for target group membership for all the cases in both databases. In the second part, we deploy a quick matching algorithm that preserves the order of those predicted scores (hence, 'isotonic'). Our method is similar in spirit, but not identical in details, to works such as Kadane 1978 (reprinted 2001), Rubin 1986, Laaksonen 1999, Moriarity and Scheuren 2001, and Moriarity and Scheuren 2002.

- *Step 1: Predictive Modeling*. The practice of predictive modeling consists of the following steps (Weiss and Indurkhya 1998). There is one database which contains the outcome variable and some predictor variables. We construct a statistical model that relates the outcome variable with the predictor variables. Then we proceed to apply this statistical model onto another database which contains only the predictor variables to obtain predicted scores for the desired outcome. Predictive modeling is used extensively in database marketing, data mining, direct marketing, credit card solicitation, credit scoring, insurance prospecting, loan approval, magazine subscriber drives, etc.

  In the present context, the TGI database contains the target group information and a list of common variables (e.g. age, sex, television viewing, etc). We construct a statistical model that relates target group membership with the predictor variables. Then we proceed to apply this statistical model to both the TAM and TGI databases, such that every person in both databases has now received a predicted score for target group membership.

- Step 2: Isotonic Matching. The TAM and TGI databases are now sorted by the predicted scores. The two databases are then merged together by a process that preserves the order of these predicted scores. A verbal description may be hard to understand. Instead, we have created an illustrated example in appendix 1. After looking at that example, the ensuing explanation should be easy to understand.

For a standard TAM-TGI setup, the predictive isotonic fusion will take just a few seconds to execute. Therefore, it satisfies the timing requirement.

There is plenty more that we can say about predictive isotonic fusion. In the interest of maintaining the flow of the exposition here, we have relegated our comments to appendix 2 of this paper.

DESCRIPTION OF EMPIRICAL DATABASES

For the empirical portion of this paper, the syndicated fusion product is the NTI-MARS 2002 product. On one side, we have the Nielsen Television Index, consisting of 11,657 adults who were intab in the Nielsen People Meter panel for one or more days during the first 13 weeks of 2002. On the other side, we have 22,097 adults who responded to the MARS OTC/DTC Pharmaceutical Study during the first quarter of 2002.

If our goal is to compare the accuracy of the fusions, then the NTI-MARS fusion itself will be uninformative. The two fusions will sometimes match different people together, but there is no way of deciding which one is more 'accurate.'

Rather, the standard approach in assessing the accuracy of fusions is through a split-sample or foldover analysis of a single source database. The MARS study contains the following relevant data elements:

- *Target group information*: For this comparison, we chose forty ailment conditions (from acid reflux to yeast infection) from the MARS study and the target group variables are defined as presence of these conditions during the past 12 months.
- *Common variables*: There are 21 demographic variables (age, sex, geography, etc) and media variables (average daily television hours, presence of cable/satellite, etc.) that are present in both databases.
- *Television variables*: The MARS study contains 17 television program types, past-seven-day viewing to 34 cable television networks and average viewing hours to 12 dayparts. These variables are not considered to be equal to the people meter data in accuracy or resolution, but they have reasonable similarity in profiles that they can be used as approximate surrogates.

The MARS respondent-level database was sorted by respondent ID and then split into two halves by alternating odd/even cases. One half-sample served the role of the NTI sample (henceforth referred to as the NTI-half-sample), and the other half-sample served the role of the MARS sample (henceforth referred to as the MARS-half-sample).

For the syndicated fusion product, the two half-samples were fused together using the method of constrained statistical matching. This consisted of dividing the samples into 36 mutually exclusive and exhaustive strata defined by age, sex and overall television viewing hours (heavy/medium/light) and then matching with the stepping stone algorithm within each stratum on the

remaining 18 common variables subject to the preservation of weights (and therefore sample size).

For the predictive isotonic fusion, linear regression models were run on the MARS-half-sample. For each target group, a linear regression model was run with that target group variable as the outcome variable. There were 56 predictor variables, which were derived by coding the 21 common variables (namely, the same ones that were used for matching in the syndicated fusion product) as indicator variables. The resulting model equation was applied to both the MARS-half-sample and the NTI-half-sample, so that every person received a predicted score. The two half-samples were then matched together by the isotonic matching method (as illustrated in appendix 1).

For any of these fusions, the result was a respondent-level fusion database. There is only one database for the syndicated fusion but, for predictive isotonic prediction, there were actually 40 such databases since the procedure was implemented separately by target group.

Within each respondent-level fusion database, a 'record' contained the following information:

- the record weight;
- the target group variables, common variables and television surrogate variables from the NTI-half-sample;
- the target group variables, common variables and television surrogate variables from the MARS-half-sample.

The assessment of the accuracy of the various fusions will be based upon comparing the original and fused data within the NTI-half-sample. The next five sections of the paper will deal with different ways of making comparisons. It is important to note that these evaluation criteria are not considered to be equally important or relevant, so the reader should pay careful attention to our discussion (see further discussion in Rässler 2002 and Soong and de Montigny 2001).

Generally speaking, we have great misgivings about using a single split-sample division to document the performance of a particular fusion, due to important issues such as sampling errors and biases. For this paper, those issues are less important since we are interested in comparing two fusions with the same split samples being held constant.

EVAULATION: GOODNESS-OF-FIT MEASURES

The first step of the predictive isotonic fusion is a multiple linear regression. We used the same set of predictor variables for all the target group variables, and we would expect that the results to be better for some than for others. For this type of method, there are some standard measures of goodness-of-fit.

The correlation coefficient reflects the relationship between the predicted score and the actual target group value. Across the 40 target groups, the average correlation coefficient is 0.250, with a range from 0.132 to 0.550. The R^2-measure reflects the proportion of variance that is accounted for by the regression. Across the 40 target group variables, the average R^2 is 0.069. These look like small numbers, but they are statistically significantly greater than 0.000 due to the large sample sizes. This range of R^2 values is typical for media and product usage variables.

Measures such as correlation coefficients and R^2 do not provide directly relevant information about the variables upon which the application revolves. Predictive modelers have a more appealing visual approach. In their terminology, the MARS-half-sample is a training sample from which the predictive model is constructed. The model is then applied to each case in the training sample to derive a predicted score. The training sample is then sorted into deciles (10%-tiles) based upon these predicted scores, and then the target group incidences are calculated by decile.

Figure 1
INCIDENCE INDEX BY PREDICTED SCORE DECILES
(APPLIED TO TRAINING SAMPLE 'MARS-HALF-SAMPLE')

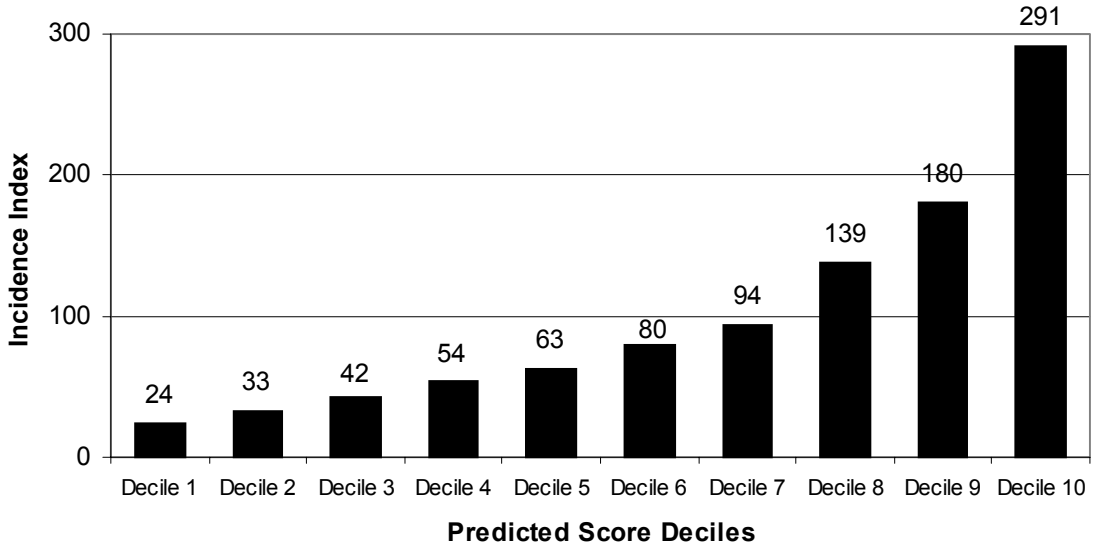

Figure 1 shows the incidence indexed to the total incidence level by these predicted score deciles. If the predictive model were totally ineffective, the indices would be around 100 everywhere. If the predictive model was effective, then the top deciles would have considerably higher incidences, with a declining trend going down the deciles. This is indeed the observed situation in figure 1.

The top decile in figure 1 has an index of 291, which translates to 29.1% of all target group people. This number is the average across 40 target groups, of which the smallest index is 156 and the largest index is 795. The top three deciles have an average index of 203, which translates to 61.0% of all target group people, with a range from a minimum of 42.2% to a maximum of 97.1% across the 40 target groups.

The use of a many-parameter predictive model will result in overfitting of the data. This means that the performance measures from training samples may be inflated. The predictive modeler will usually run an independent validation of the model. In their terminology, the NTI-half-sample is a validation (or hold-out) sample. The predictive model derived from the training sample (that is, the MARS-half-sample) is applied to each case in the validation sample to derive a predicted score. The validation sample is then sorted into deciles based upon these predicted scores, and then the target group incidences are calculated by decile.

Figure 2
INCIDENCE INDEX BY PREDICTED SCORE DECILES
(APPLIED TO TRAINING SAMPLE 'MARS-HALF-SAMPLE')

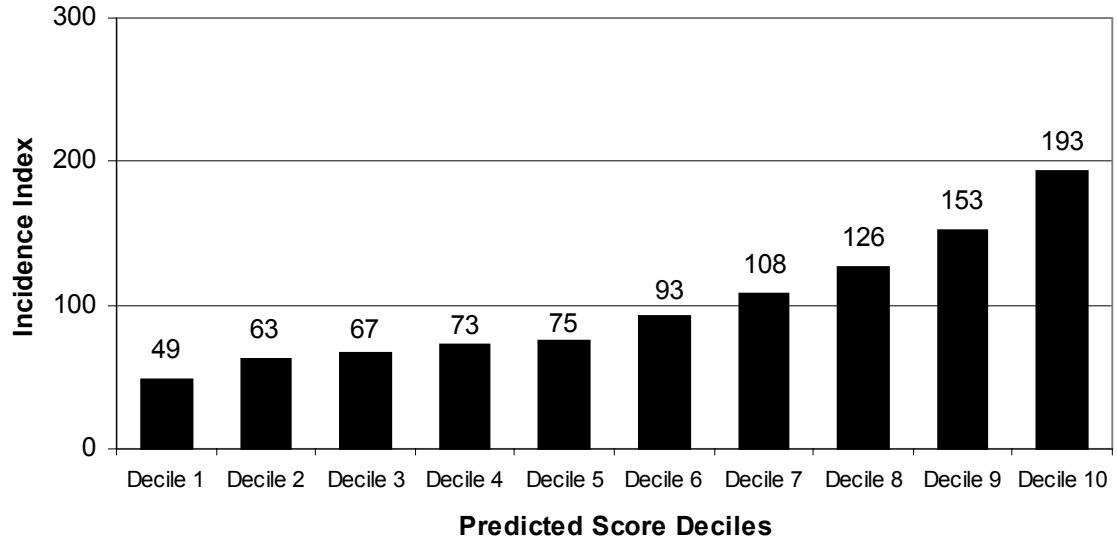

Figure 2 shows the incidence indices by predicted score deciles within the validation sample. By comparing against figure 1, we can see that there is a general pullback (known as regression-to-the-mean) in the top deciles. These are now realistic measures of the performance of the predictive model. The top decile in figure 2 has an index of 197, which translates to 19.7% of all target group people. The range of the index runs from a minimum of 97 to a maximum of 497 across the 40 target groups. There are a couple of cases in which the predictive model was not very effective. The top three deciles have an average index of 157, which translates to 47.2% of all target group people, ranging from a minimum of 31.0% to a maximum of 92.6%.

It is easy to see that the predictive model has generated a powerful sorting of the cases for most target groups. However, the exact impact on the accuracy of fusion will have to be addressed by some other type of evaluation criteria.

EVALUATION: MATCHING SUCCESS RATES

In the fusion database, each record contains two values for the same common variable (e.g. age), one coming from the NTI-half-sample and the other from the MARS-half-sample. A measure of the closeness of the fusion is the percent of time in which the two values coincide with each other across all records. (See table 1.)

Table 1 shows the percentages of records in which the matching was successful by the two fusion methods for the twenty-one common variables.

The syndicated fusion is designed to maximize the matching success rates subject to the constraints of preserving weights and therefore sample sizes. Therefore, those matching success rates are theoretically the best that can be achieved under the particular choice of importance weights assigned to the matching variables.

The predictive isotonic fusion concentrates solely on matching the predicted scores and gives no direct consideration to matching the common variables. From table 1 we clearly see that the matching success rates are much lower under predictive isotonic fusion.

Kadane's (1978, reprinted 2001) proposal was to run a predictive model and then match on both the common variables and the predicted scores simultaneously. While this would yield better matching success rates on the common variables, the computational load is the same as that of the syndicated fusion and therefore takes it out of the realm of interactive fusion-on-the-fly.

Table 1
SUCCESS RATES IN MATCHING COMMON VARIABLES

Variable	*Syndicated Fusion*	*Predictive Isotonic Fusion*	*Index*
Gender	100	57	57
Age	97	26	27
Race	70	65	94
Presence of Child <2	91	85	93
Presence of Child 2-5	88	79	89
Presence of Child 6-11	88	74	84
Presence of Child 12-17	84	70	83
Household Income	57	20	35
HOH Age	75	34	45
HOH Education	55	29	53
HOH Occupation	54	29	54
Household Size	90	79	87
Territory	45	19	41
County Size	65	30	46
Working woman	85	57	67
Cable TV	72	54	75
Satellite TV	75	73	97
TV Weekday 6am-9am	78	52	67
TV Weekday 9am-4pm	81	54	67
TV Weekday 11pm-1am	80	58	73
TV viewing deciles	55	12	21

Realistically, we are faced with the fact that predictive isotonic fusion can result in what appears to be poor success matching rates on the common variables. But that fact by itself does not have any direct bearing on the accuracy of the fusions for the intended application. Here, the most important consideration is about the target group information and its relationship to

television viewing behavior, and the common variables are merely intermediate devices.

EVALUATION: INCIDENCES

Both fusion methods aim to preserve sample weights (and hence sample sizes). Television variables are perfectly preserved in the fusion databases, including average ratings, duplications, reaches and exposure frequency distributions. Target group incidences are almost perfectly preserved, with very small discrepancies (less than two parts per thousand) due to slight structural differences between the two databases. The two fusion methods are therefore equally good with respect to this criterion.

EVALUATION: TARGET GROUP MATCHING

These individual-level analyses are described in detail by Soong and de Montigny (2001, Section 9). Within the fusion database, we cross-classified the records by their NTI-half-sample and MARS-half-sample target group variables. This results in the following 2x2 contingency table known as the confusion matrix.

Table 2
DEFINITION OF CONFUSION MATRIX

	MARS-half-sample: Yes	***MARS-half-sample: No***
NTI-half-sample: Yes	True positive	False positive
NTI-half-sample: No	False negative	True negative

From this confusion matrix, some common statistics are derived:

$$\text{Accuracy} = 100 \times \frac{(\text{Number of true positive}) + (\text{Number of true negatives})}{(\text{Total number of cases})}$$

$$\text{Sensitivity} = 100 \times \frac{(\text{Number of true positives})}{(\text{Number of true positives}) + (\text{Number of false negatives})}.$$

$$\text{Specificity} = 100 \times \frac{(\text{Number of true negatives})}{(\text{Number of true negatives}) + (\text{Number of false positives})}.$$

$$\text{Precision} = 100 \times \frac{(\text{Number of true positives})}{(\text{Number of true positives}) + (\text{Number of false positives})}.$$

Accuracy is the percent of correctly classified cases, but its weakness is that it lumps the true positives and true negatives together when we are more interested in the true positives. Sensitivity addresses the question: "Of the target group people, what percent of them of were classified as such?" Specificity addresses the question, "Of the people who were not in the target group, what percent of them were classified as being in the target group?" Precision addresses the question, "Of the people who were classified as being in the target group, what percent of them were really that?" In the current context, the sensitivity measure appears to be the most relevant to the situation.

Table 3
TARGET GROUP MATCHING MEASURES

Statistic	*Syndicated Fusion*	*Predictive Isotonic Fusion*	*Index*
Accuracy	79.5	79.8	100
Sensitivity	17.0	17.7	104
Specificity	87.5	87.6	100
Precision	17.0	17.5	103

In table 3, we show these summary measures averaged across the 40 target groups. For the sensitivity and precision measures, there were 4% improvements with the predictive isotonic fusion. The accuracy and specificity measures were essentially the same. Thus, customization by target group yielded a modest improvement on the average.

In detail for the sensitivity measure, 14 of the 40 target groups did not show improvement, 9 improved by 1% - 5%, 5 improved by 6%-10%, 3 improved by 11% - 15%, 1 improved by 16% - 20% and 8 improved by 21% or more. So while the average improvement was modest, there were large gains for individual target groups.

Measures such as accuracy and sensitivity are easy to describe and calculate. But we do not believe that they or any other statistics derived from the confusion matrix directly address the question of the quality of a fusion in the present context. Our application is about target group ratings – to emphasize the point, this is about target groups *and* television ratings. The confusion matrix contains no reference whatsoever to any television viewing information and is therefore not directly relevant to the application. The question about the

accuracy of target group ratings needs to be answered in terms of target group ratings themselves.

EVALUATION: TARGET GROUP RATINGS

Within the TAM-half-sample, we have information for 40 target groups and 63 television entities (17 program types, 34 cable networks and 12 dayparts). The original target group rating (TGR) is defined as:

$$\text{Original TGR} = 100 \times \frac{\text{(\# of original TV viewing product users)}}{\text{(\# of original product users)}}$$

After the fusion, the TAM-half-sample received fused target group information. The fused target group rating (TGR) is defined as:

$$\text{Fused TGR} = 100 \times \frac{\text{(\# of fused product users who are original TV viewers)}}{\text{(\# of fused product users)}}$$

The goodness-of-fit measure is:

$$\text{TGR index} = 100 \times (\text{Fused TGR}) / (\text{Original TGR})$$

Under perfect fusion, the TGR index is 100, and large deviations from 100 are taken to be bad.

Table 4
SUMMARY OF TARGET GROUP RATING (TGR) INDICES BY FUSION METHOD

Summary Measure	*Syndicated Fusion*	*Predictive Isotonic Fusion*
Minimum	44.1	48.3
10%-tile	80.8	79.8
25%-tile	87.6	86.5
Mean	93.7	92.9
Median (50%-tile)	94.3	93.5
75%-tile	99.9	99.0
90%-tile	106.4	107.5
Maximum	232.6	281.6

In table 4 we show the summary measures of the 40 x 63 = 2,520 TGR indices by fusion method. We remind the reader that these numbers should not be taken as absolute indicators of the goodness of the fusions as there are issues associated with the sampling errors and biases from a single split-sample. Rather, the attention should be focused on the relationships between the two fusions.

In table 4, the two types of fusion yield about the same TGR indices in terms of means and spreads, so there is little to choose among the two methods. Out of the 2,520 TGR indices, the syndicated fusion is closer to 100 than the predictive isotonic fusion in 55% of the cases. Out of the 40 target groups, the syndicated fusion is on the average closer to 100 in 27 cases.

Under this evaluation criterion, which we consider to be the most directly relevant one, the syndicated fusion comes out just slightly better than the predictive isotonic fusion. This is in fact good news for predictive isotonic fusion, since there are little negative trade-offs for the quick execution time.

EXTENSIONS

Much of the practical experience of predictive modeling suggests that the choice of method will make little or no difference, and it is much more important to have the right predictor variables. This piece of wisdom certainly applies to data fusion as well. Given the same set of predictor variables, our two fusion methods seemed close on the most important criterion.

We can improve the performance of data fusion in general by introducing better matching/predictor variables, especially ones that are related to television viewing in this case. Here, there is a difference in the ability to accommodate additional variables. For syndicated fusion, the high-dimensional matching problem is already stressful. Adding a large group of TV-related matching variables will simply mean that their matching success rates will be poor; in addition, the matching rates of the existing common variables will deteriorate. This phenomenon is illustrated in appendix 2 of Soong and de Montigny (2001).

By contrast, predictive isotonic fusion is not subjected to this limitation. The existing framework here consists of fitting a multiple linear regression of 56 predictor variables for a total sample size of more than 11,000 cases. Adding another few dozen more predictor variables will not stress the system.

To illustrate this point, we run another predictive isotonic fusion, this time adding the 63 TV-related variables as predictor variables. We find the following results.

Figure 3
INCIDENCE INDEX BY PREDICTED SCORE DECILES WITH ADDITIONAL TV VARIABLES (APPLIED TO TRAINING SAMPLE 'MARS-HALF-SAMPLE')

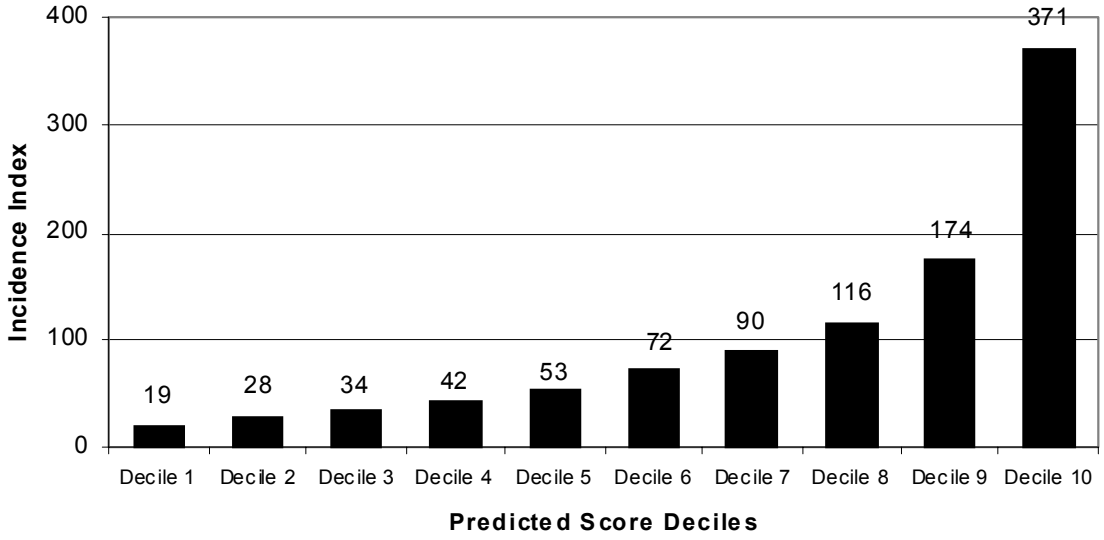

Figure 4
INCIDENCE INDEX BY PREDICTED SCORE DECILES WITH ADDITIONAL TV VARIABLES (APPLIED TO VALIDATION SAMPLE 'NTI-HALF-SAMPLE')

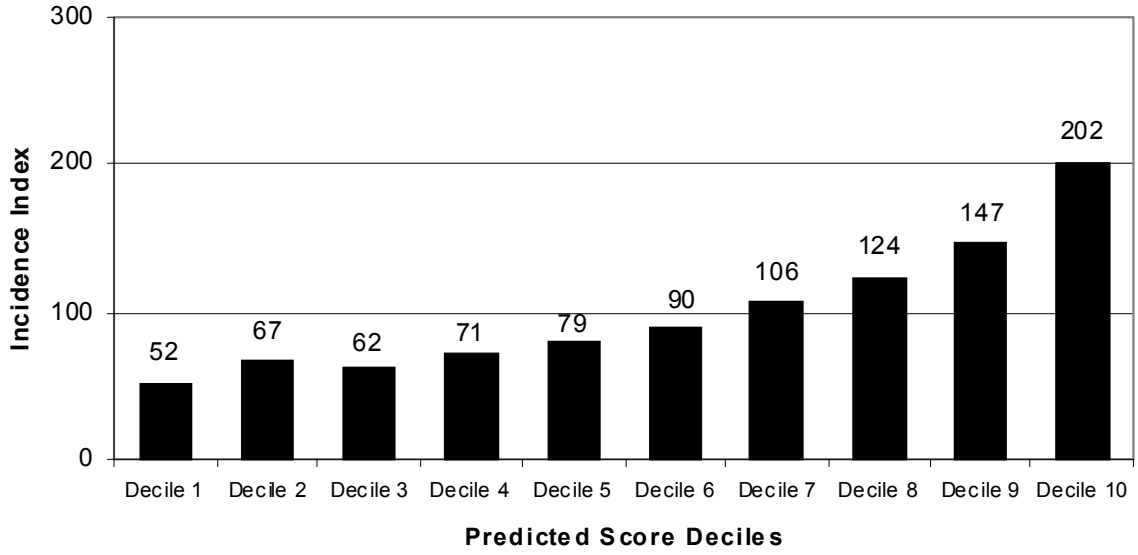

Figure 3 shows the incidence indexed to the total incidence level by these predicted score deciles. Compared to figure 1, the addition of more predictor variables has increased the incidences in the top deciles. Figure 4 shows the incidence indices from the validation sample. This is about the same as in figure 2, so that there was no damage from model-overfitting.

On the target group matching, the mean sensitivity of the TV-enhanced predictive isotonic fusion was 10% better than the syndicated fusion, compared to 4% for the regular predictive isotonic fusion. In detail, 9 of the 40 target groups did not show improvement, 9 improved by 1% - 5%, 4 improved by 6% - 10%, 2 improved by 11% - 15%, 4 improved by 16% - 20% and 12 improved by 21% or more.

On the most important criterion, the average TGR index for the TV-enhanced predictor isotonic fusion rose to 99.0 compared to the 94.3 for syndicated fusion and 93.5 for the regular predictive isotonic fusion. Out of the 2,520 TGR indices, the syndicated fusion is closer to 100 than the TV-enhanced predictive isotonic fusion in only 38% of the cases. Out of the 40 target groups, the syndicated fusion is on the average closer to 100 in just 6 cases. The details are shown in table 5. The advantage clearly goes to the TV-enhanced predictive isotonic fusion.

Table 5
SUMMARY OF TARGET GROUP RATING (TGR) INDICES BY FUSION METHOD

Summary Measure	*Syndicated Fusion*	*Predictive Isotonic Fusion*	*TV-enhanced Predictive Isotonic Fusion*
Minimum	44.1	48.3	43.3
10%-tile	80.8	79.8	87.4
25%-tile	87.6	86.5	93.2
Mean	93.7	92.9	99.0
Median (50%-tile)	94.3	93.5	98.3
75%-tile	99.9	99.0	103.7
90%-tile	106.4	107.5	111.0
Maximum	232.6	281.6	188.0

CONCLUSIONS

The first contribution of this paper is to describe an open-source algorithm that is based upon the well-understood practice of predictive modeling followed by a quick sort-and-match. When we compared the performance of this predictive isotonic fusion algorithm against a syndicated data fusion, we found that this method is about the same. In a few cases, though, there appeared to be significant improvements.

This is not surprising at all. Given the same databases with the same variables, different fusions can be thought of as just variations in ways of prioritizing variables and arranging matches. Two very large-scale studies (Soong and de Montigny 2001, appendix 1; and Okauchi 2002), which explored many, many different ways of prioritizing variables via genetic algorithms showed that the best solutions are not distinctly superior to average solutions. This phenomenon is known as the 'flat maximum effect' or 'the curse of insensitivity.'

This is in fact good news for predictive isotonic fusion, since we have shown that we have a fast algorithm that does not suffer any loss in accuracy. The more valuable observation is that predictive isotonic fusion has the ability to accommodate many more predictor variables and is versatile to the point of even importing predictive models from outside (see discussion in appendix 2).

The story on predictive isotonic fusion is only half-complete. In this paper, we have focused only on target group ratings. So far, the results have been favorable. Our next step is to examine the accuracy of multimedia applications, including variations that may enhance those types of fusion.

REFERENCES

Baron, R. (2001). A new practical approach to data fusion. *ARF Week of Workshops*, Chicago, IL.

Butler, D. and Bryson, S. (1992). Vector-bundle classes form powerful tool for scientific visualization. *Computers in Physics*, 6(6), 576-584.

Czaia, U. (1993). Interactive fusion: step two. In *Sixth Worldwide Readership Research Symposium*, San Francisco, 489-493.

DeGroot, M.H., Feder, P.I. and Goel, P.K. (1971) Matchmaking. *Annals of Mathematical Statistics*, 42, 578-593.

Goel, P.K. and Ramalingam, T. (1989). The matching methodology: some statistical properties. Lecture Notes in Statistics, Volume 52. Springer-Verlag New York.

Kadane, J.B. (1978). Some statistical problems in merging data files. In 1978 Compendium of Tax Research, Office of Tax Analysis, Department of the Treasury, 159-171. Washington, DC: U.S. Government Printing Office. Reprinted in Journal of Official Statistics (2001), 17, 423-433.

Kadane, J.B. (2001) Some statistical problems in merging data files. *Journal of Official Statistics*, 17, 423-433.

Laaksonen, S. (1999) How to find the best imputation technique. Tests with three methods. *International Conference on Nonresponse*, Portland OR.

Moriarity, C. and Scheuren, F. (2001) Statistical matching: a paradigm for assessing the uncertainty in the procedure. *Journal of Official Statistics*, 17, 407-422.

Moriarity, C. and Scheuren, F. (2003) A note on Rubin's statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics*, 21, 65-73.

Okauchi, S. (2002) The Japanese fusion experience (validation of data fusion by means of ACR data splitting). *ARF Week of Workships*. October 10, 2002, New York City.

Raimondi, D. and Santini, G. (1997) Just-in-time data modelling. *Proceedings of the Vancouver Worldwide Readership Symposium*.

Rässler, S. (2002) *Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*. Springer-Verlag New York: New York.

Rubin, D.B. (1986) Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics*, 4, 87-94.

Soong, R. (2002) Quick vs. optimal algorithms in data fusion. *Zona Latina*, January 2002. (http://www.zonalatina.com/Zldata215.htm)

Soong, R. and de Montigny, M. (2001). An anatomy of data fusion. *Paper for the Worldwide Readership Research Symposium*, Venice (Italy), 87-109. ()

Steenrod, N. (1950). *The Topology of Fibre Bundles*. Princeton, NJ: Princeton University Press.

Weiss, S.M. and Indurkhya, N. (1998) *Predictive Data Mining: A Practical Guide*. San Francisco, California: Morgan Kaufmann Publishers Inc.

THE AUTHORS

Roland Soong is Chief Technical Officer, Kantar Media Research, United States.

Michelle de Montigny is Executive Vice President, Kantar Media Research, United States.

APPENDIX 1
ILLUSTRATED EXAMPLE OF ISOTONIC MATCHING

After the predictive modeling step, all the cases in the TAM-half-sample and the TGI-half-sample have received predictive scores. The goal now is to create a matching. According to DeGroot, Feder and Goel (1971), the maximum likelihood pairing is to sort the predicted scores and then to link the corresponding pairs (that is, the largest values together, the second largest values together, and so on). Since this pairing preserves the order of the predicted scores, Goel and Ramalingam (1989, Section 3.1.1, p.76-78) named it 'isotonic matching.'

When survey weights are present on databases of unequal sample sizes, the method must be adapted. We will illustrate with a small example. In table A1 the two databases are each sorted in order of these predicted scores. It is noted that one database contains four cases and the other database contains five cases, and they both sum to the same projected weight of 2,000.

Table A1
EXAMPLE OF TAM-SAMPLE AND TGI-SAMPLE WITH PREDICTED SCORES.

TAM-sample ID	*Weight*	*Predicted Score*		*TGI-sample ID*	*Weight*	*Predicted Score*
TAM-1	600	0.75		***TGI-1***	300	0.80
TAM-2	400	0.50		***TGI-2***	400	0.60
TAM-3	300	0.25		***TGI-3***	200	0.30
TAM-4	700	0.10		***TGI-4***	500	0.20
Total	2000			***TGI-5***	600	0.05
				Total	2000	

The fusion database is shown in table A2.

Table A2
FUSION DATABASE

Fused ID	*TAM-sample ID*	*TGI-sample ID*	*Weight*	*TAM Predicted Score*	*TGI Predicted Score*
Fused-1	TAM-1	TGI-1	300	0.75	0.80
Fused-2	TAM-1	TGI-2	300	0.75	0.60
Fused-3	TAM-2	TGI-2	100	0.50	0.60
Fused-4	TAM-2	TGI-3	200	0.50	0.30
Fused-5	TAM-2	TGI-4	100	0.50	0.20
Fused-6	TAM-3	TGI-4	300	0.25	0.20
Fused-7	TAM-4	TGI-4	100	0.10	0.20
Fused-8	TAM-4	TGI-5	600	0.10	0.05
Total			2000		

Isotonic matching works by marching down the two half-samples from the top to the bottom, one record at a time. At first, we look at the first records (TAM-1 and TGI-1). We write into the fusion database a record corresponding to TAM-1 and TGI-1 and the smaller of the two weights. Then we subtract these weights from the two original databases. Thus, TAM-1 is still present in the TAM-half-sample, but with a reduced weight of 600-300 = 300 whereas TGI-1 is completely removed from the TGI-half-sample.

We repeat the process on the revised half-samples. So the next fusion record to be written out is (TAM-1 and TGI-2), after which TAM-1 is completely accounted for and TGI-2 is reduced to 400-300 = 100. This process is continued and will eventually terminate with everyone accounted for.

Isotonic matching is equivalent to the northwest-corner rule that is sometimes used to jumpstart the stepping stone algorithm for the transportation problem (Soong 2002). Therefore, it has the unimodularity property of creating a fusion database whose sample size is no more than the sum of the two input databases minus one. The computational complexity is linear in the sample sizes, and the execution is therefore instantaneous on (TAM+TGI)-like databases.

In the fusion database, the sum of record weights is the same total as in the two original databases. Furthermore, since each original TAM and TGI person is present in the fusion database – sometimes in more than one record – with the same relative weight, this method satisfies the requirements to preserve sample sizes, media currencies and product usage incidences.

APPENDIX 2
SOME COMMENTS ABOUT PREDICTIVE ISOTONIC FUSION

Once we have created a predictive model and applied it to the TAM database, the predictive modeler would have declared that there is a target group in its own right based upon applying some threshold (e.g. the top 20% of the predicted scores). This target group is clearly identifiable and its properties can be documented (e.g. the top 20% of the predicted scores covered 80% of the product users). The target group would have a label such as 'those who are in the top quintile of predicted scores.' We are sympathetic to this viewpoint, but we continue through with the statistical matching because of the other extensions (such as multimedia planning/buying/optimization).

Predictive models can be constructed by any number of techniques, such as multiple linear/nonlinear regression, discriminant analysis, logistic regression, probit regression, tobit regression, proportionate hazard regression, neural networks, support vector machines, kernel methods, nearest neighbor matching, AID, CHAID, CART, MARS, and so on. These methods differ in their technical assumptions about functional forms, homogeneity/heterogeneity of variance, error distributions, distance functions, etc. As a practical matter, for reasons that we will explain, we are indifferent to the choice of the method as long as the execution time is reasonable and the full information is being utilized.

In the present context, we are not seeking precise numerical estimates. All we are looking for is a way of ranking people by the predicted scores. The ranking is invariant under nonlinear, order-preserving transformations, which means that most of these methods will yield approximately the same ranking. This being the case, we would use the computationally simplest method (such as multiple linear regression) instead of the more computationally complex method (such as neural networks).

We do warn against the use of classification tree methodologies such as AID, CHAID, CART and MARS because they may not be able to leverage the full information. Consider the example of a TAM panel of 10,000 persons. It seems reasonable that we would require 36 = 2 x 6 x 3 critical strata formed by gender (male/female), age (18-24, 25-34, 35-44, 45-54, 55-64, 65+) and TV viewing (heavy/medium/light) as the starting points of the classification trees. Thus, the average stratum contains about 10,000 / 36 ≈ 280 persons. There may be 20 more common variables, but the classification tree methodology can sub-divide a stratum at most once or twice more before declaring that no further 'statistically significant' splitting is available. But it is not that the remaining variables are really insignificant in a substantive sense; it is just that the cross-tabulation-based system cannot accommodate them within this sample size. By contrast, for example, a logistic regression can use all of these variables as predictors and find them to be 'statistically significant.'

Some predictive modeling methods result in estimates which have the meaning of being probabilities of target group membership. Once applied to the TAM database, this opens up a couple of strategies other than statistical matching.

For one thing, the TAM respondents can be assigned to target group membership (or not) by referencing the estimated probabilities to a random number generator. This was the approach adopted by Baron (2001). For another thing, the probabilities can be applied directly to the case weight of the TAM respondents in order to obtain a projected target group universe.

In either case, there is the risk that the target group incidence may be different than the original incidence. More significantly, this approaches precludes other extensions such as multimedia planning/buying/optimization.

So far, we have set up a structure with the same TGI-based predictive model being applied to both the TAM and TGI databases. It is in fact not required to deploy the same model to the two databases. The sole purpose is to obtain the best ranking of the respondents for target group membership, by whatever means possible.

In fact, to push it this point further, the predictive model does not even have to be derived from these databases, as there may be an auxiliary database that contains better information for fusion purposes.

Consider this hypothetical situation. Our goal is to fuse a local market television diary sample (such as NSI) with a local radio diary sample (such as Arbitron). Indeed, the only common variables are age, sex and geography, which would probably make for a weak fusion. Suppose there exists a local market multimedia study (such as Scarborough), which contains demographics, product usage, radio listening and television viewing. From the multimedia study, we construct a predictive model of target group membership from age, sex, geography and television viewing to be applied to the television diary sample. From the multimedia study, we construct a predictive model of target group membership from age, sex, geography and radio listening to be applied to the radio diary sample. The television and radio samples can now be fused by isotonic matching. In theory at least, this should be a more powerful fusion than one based upon age, sex and geography only.