Journal of Manufacturing Science and Engineering

Technical Briefs

Data Mining for Subassembly Selection

Bruno Agard
Assistant Professor, Départment de Mathématiques et de Génie Industriel, École Polytechnique de Montréal,
C.P. 6079, succ. Centre-ville, Montréal
(Québec), H3C3A7, Canada
e-mail: bruno.agard@polymtl.ca

Andrew Kusiak
Professor, Intelligent Systems Laboratory, 2139 Seamans Center, Department of Mechanical and Industrial Engineering, The University of Iowa, IA 52242-1527
e-mail: andrew-kusiak@uiowa.edu

The paper presents a model and an algorithm for selection of subassemblies based on the analysis of prior orders received from the customers. The parameters of this model are generated using association rules extracted by a data mining algorithm. The extracted knowledge is applied to construct a model for selection of subassemblies for timely delivery from the suppliers to the contractor. The proposed knowledge discovery and optimization framework integrates the concepts from product design and manufacturing efficiency. The ideas introduced in the paper are illustrated with an example and an automotive case study.
[DOI: 10.1115/1.1763182]

1 Introduction

Customers expect to obtain products of perceived functionality at the right time and cost. To meet these customer needs, manufacturers often produce a large number of different products, which may lead to the excessive product diversity—the economy of scale dilemma (see [1] and [2]). It is natural that the price of a specialized product is higher than that of a simple one. To decrease production costs while ensuring the diversification needed, manufacturers standardize products to benefit from the economy of scale [1,3,4]. Most products are neither fully standardized nor specialized, rather a mix of the two [5].

To manufacture widely diversified products, two approaches have been advocated in the literature: (1) modular design aiming at the use of identical components and subassemblies for customized products, and (2) delaying the differentiation of products or processes (known as the delayed differentiation strategy).

1.1 Modular Design.

Product flexibility and the use of common components within various products are of importance in modular design [6]. The flexibility of a module (the number of its uses) depends on its surplus functionality and the required standard interfaces.

Prior research on modular design has emphasized consistency of the design process and manufacturing. For example, the taboo search algorithm presented in [7] aims at the design of an assembly system for modular products. The development of modular products with the minimum test cost is discussed in [8]. A modular design methodology intended to produce a large variety of products at a low cost is discussed in [9]. Other examples of modular concepts are presented in [6].

Modular design leads to a large number of different products using a limited number of modular components. One aspect of product modularity, the design of product families, has been discussed in Newcomb et al. [10], Simpson et al. [11], Kusiak and Huang [12], Martin [13], Dahmus et al. [14] Erens and Verhulst [15], Gonzalez-Zugasti et al. [16], and Jiao and Tseng [17].

1.2 Delayed Product Differentiation.

The delayed product differentiation concept implies delaying the point of differentiation of the product or the process (in which a product acquires its identity) [18]. The goal of the delayed product differentiation is to maximize the use of standard elements and to push back, to the latest time possible, the point when each product differs from another. Some authors, e.g., [18,19], used the term postponement as a synonym of delayed differentiation.

2 Problem Statement

The research reported in this paper has been motivated by an industrial application. A supplier realizes a component (wire harnesses) for a contractor that produces cars. The contractor assembles different components received from the suppliers on a synchronous production line. The contractor requires that the right subassemblies be delivered in the right sequence and at the right time. As the product assembly sequence cannot be modified, delaying assembly of a one product implies delaying assembly of other products, which is not acceptable.

The supplier faces the problem of realizing a large variety of subassemblies that have to be delivered in the right sequence and at the right time.

2.1 Wire Harness.

A wire harness is a collection of electric cables that are used to connect different elements in electromechanical or electronic systems. The functions of a wire harness are to provide electric power and electronic signals to the different peripheral units. An example wire harness is shown in Fig. 1.

A wire harness includes different kinds of elements, such as:

Contributed by the Manufacturing Engineering Division for publication in the JOURNAL OF MANUFACTURING SCIENCE AND ENGINEERING. Manuscript received December 2002; revised December 2003. Associate Editor: S. Raman.

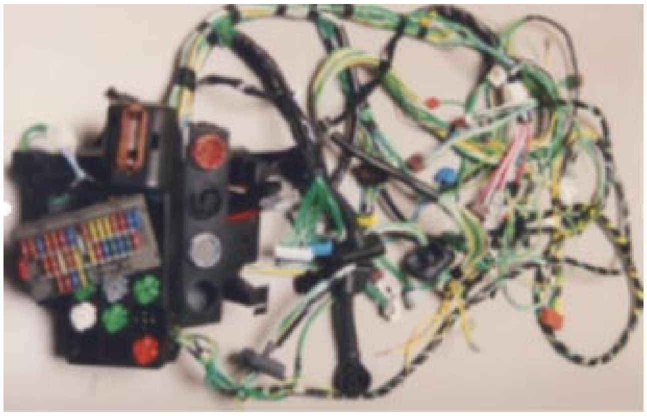


Fig. 1 A wire harness

- — Cables that are used to transmit information and energy;
- — Connectors linking the wire harness with the other elements;
- — Branches are places on the wire harness where cables change directions.

The functions of a wire harness are illustrated in Fig. 2.

2.2 Subassemblies. The supplier has to consider both, the wide variety of products and the delivery time, which is shorter than the total time to manufacture a wire harness. A viable solution is in the standardization of wire harnesses, which simplifies the delivery. The latter is due to diminished impact of the order and the delivery time as the wire harnesses can be produced to stock.

The cost of a unique standardized wire harness is relatively high because some customers may not select some options or alternatives and yet they may receive the same wire harnesses as those who have ordered them.

The supplier has attempted to use a small number of standardized wire harnesses. Each type of a wire harness was manufactured to stock. The synchronous delivery is a complicating factor when a supplier has to deliver a wire harness that meets the minimum customer requirements. Moreover for each standardized wire harness there are some options that are not used at a cost that no one wants to bear.

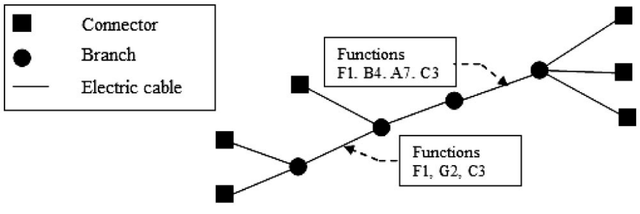


Fig. 2 Functional representation of a wire harness

In the above described scenario, there are many options and alternatives and standardization does not provide a totally satisfactory answer. The supplier has decided to produce some modular subassemblies to be realized in advance and stocked. Each subassembly encompasses a set of functions to be stored and used as modules. Those modules are called Industrial Modules (IMs). The work that is done in advance for each IM includes only the elements of the IM that will not be modified in the future assembly.

The supplier waits for an order and when it arrives a specific wire harness is produced from the subassemblies at the supplier's disposal. Then the supplier delivers the wire harness that matches best the consumer requirements. Besides, the supplier has to produce all the subassemblies necessary to manufacture a specific wire harness subject to the delivery time constraints.

In this paper, the time to produce all subassemblies in order to manufacture a specific wire harness is called the time of final assembly and it is discussed in the next section.

2.3 Time of Final Assembly. The time of final assembly (*TFA*) is the time to manufacture a specific wire harness to meet customer's requirements.

This *TFA* depends on the number of subassemblies that are used to produce a wire harness and the way the different subassemblies are combined.

The time to manufacture an industrial module is called the Reduced Final Assembly Time (*RFAT*) and it is computed as follows [20]:

$$RFAT_k = \sum_{i=0}^{No\_of\_branches} time\_to\_realize\_branch(i) + \sum_{j=0}^{No\_of\_nodes} time\_to\_realize\_node(j) \tag{1}$$

$$Time\_to\_realize\_branch(i) = \begin{cases} \bullet\ Type\_of\_assembly(i) \times length\_of\_branch(i), \\ \text{if all functions of branch (i) belong to the module} \\ \bullet\ 0,\ \text{otherwise} \end{cases} \tag{2}$$

$$Time\_to\_realize\_node(j) = \begin{cases} \bullet\ Type\_of\_node(j),\ \text{if all branches of j belong} \\ \text{to the module (time to assemble a connector or the} \\ \text{time to manufacture all branches)} \\ \bullet\ 0,\ \text{otherwise} \end{cases} \tag{3}$$

The final assembly time is then computed from (4).

$$TFA = TFAm - \sum_{k=0}^{No\_of\_modules} RFAT_k \tag{4}$$

3 Data Mining Approach

In this research product (wire harness) options and alternatives are considered. Options are the non-standard product functions at a customer disposal, e.g., a sun roof is an option that a customer may request. Alternatives are characteristics necessary for the product that contain different attributes. For example, the color of a car is alternative as it is possible to have cars of different colors, but each car has to have a certain color. For each wire harness a set of options and alternatives is defined.

The wire harness options and alternatives considered in this study are as follows:

Table 1 Data set with eight orders

Order No.	ABS	ASC	Single AC	Dual AC	Single CD	Multiple CD	Powered seats	Sun roof	Cruise control
12335	1	0	0	1	0	1	1	0	0
15645	1	1	1	0	1	0	0	0	1
16544	0	0	0	1	1	0	1	0	0
67165	0	0	1	0	1	0	0	1	0
16871	1	1	1	0	0	1	0	1	0
32116	0	0	0	0	0	0	1	0	1
11544	1	1	0	1	0	1	0	0	1
15423	0	0	1	0	0	0	1	1	0

- ABS (Anti Blockage System)-an electronic system that prevents the wheels form blocking when the car is breaking, it's an option that can be required or not;
- ASC (Anti Skid Control)-an electronic system that prevents the car from sliding when it's running, it is an option;
- AC (Air Conditioner)-an option with two alternatives: "Single AC" or "Dual AC;"
- CD (Compact Disc)-an option with two alternatives "Single CD" or "Multiple CD;"
- Powered seats, Sun roof and Cruise control are other three more options available to the customer.

To determine the best set of subassemblies for manufacturing (the best set of options), a data mining approach was used. The database containing the options and alternatives selected by the customers was used as a training data set.

To realize the subassemblies, the relationships between the options are needed. For example:

- The percentage of orders where "ABS" and "ASC" appear together;
- The percentage of orders where "Dual AC" and "Powered seats" appear together;
- The percentage of orders where "ABS," "ASC," "Dual AC" and "Powered seats" appear together.

An association rule learning algorithm derives relationships among different attributes characterizing sub-assemblies [21]. Each association rule is characterized by two metrics, support and confidence.

For example, for the association rule A⇒B extracted from the database:

- Support represents the quantity of A in the data set;
- Confidence represents the ratio (A & B)/A.

3.1 Illustrative Data Set. The following example mimics the database used by the supplier in our case study. The data in Table 1 includes nine options and alternatives (ABS, ASC, single AC, dual AC, single CD, multiple CD, powered seats, sun roof and cruise control).

Each row of Table 1 represents an order of a customer that is to be realized. Each entry (besides the Order No. column) specifies whether an option has been selected (1) or not (0) by a customer.

For example, in the order No. 12335 the customer has selected the following options: ABS, Dual AC, Multiple CD, and Powered seats.

Based on the data in Table 1 additional metrics can be derived, e.g.:

- ABS option has been required by 50% (4/8) of the customers;
- ASC option has been required by 37.5% (3/8) of the customers.

3.2 Association Rules. For the data in Table 1 rules have been extracted with the association rule algorithm [21]. Three sample rules are presented next.

Rule 1. (ABS=0)⇒(Multiple CD=0) [4, 100%]

Rule 7. (Multiple_CD=0) & (Powered_seats
(Powered_seats=0)⇒(Single_CD=1) [2, 100%]

Rule 17. (ASC=1)⇒(ABS=1) [3, 100%]

These sample rules are interpreted as follows:

Rule 1. In the data set, 4 orders call for (ABS=0), and in 100% of those cases the orders ask for (Multiple CD=0).

Rule 7. 2 orders with (Multiple_CD=0) and (Powered_seats=0) call for (Single_CD=1) and this orders make up 100% of such cases.

Rule 17. 3 customers require (ASC=1) and in 100% of those cases they also ask for (ABS=1).

3.3 Rule Selection. Further analysis of the three sample rules indicates that:

Rule 1. If a customer did not request "ABS," then the "Multiple CD" option was not requested, which occurred 4 times out of 8 orders (50%). This rule implies that "ABS" and "Multiple CD" do not frequently appear together in the same product. This rule will not be used for subassembly as the corresponding functions have not been requested.

Rule 7. If a customer did not request "Multiple CD" and "Powered seats," then the "Single CD" option was requested, which occurred 2 times (25%). This rule is supported by a few examples and therefore it will not be used for subassembly.

Rule 17. When a customer requested "ASC," the "ABS" option was requested in 3 out of 8 times (37.5%). It means that "ASC" and "ABS" are associated in 37.5% of the requests. This rule has been selected for future study because it has a relatively high support and it associates the existing options and/or alternatives ("ASC" and "ABS").

The three above sample cases indicate that the useful rules need to be selected. A three-step procedure for the rule selection is presented in Fig. 3.

For the previously presented example, rules 1 and 7 have not been selected for building subassemblies.

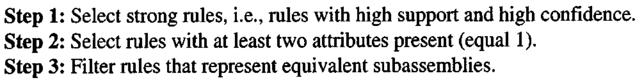
The association between "ASC" and "ABS" of rule 17 suggests the realization of a subassembly with the two attributes. In this case, cost analysis determines its implementation.

From the rules extracted based on the data in Table 1 only rules 3, 16 and 17 have been selected.

Rule 3. (Sun_roof=1)⇒(Single_AC=1); [3, 100%]

Rule 16. (Multiple_CD=1)⇒(ABS=1); [3, 100%]

Rule 17. (ASC=1)⇒(ABS=1); [3, 100%]



Step 1: Select strong rules, i.e., rules with high support and high confidence.
Step 2: Select rules with at least two attributes present (equal 1).
Step 3: Filter rules that represent equivalent subassemblies.

Fig. 3 Selection of rules for subassembly.

Table 2 Candidate modules with the corresponding $RFAT_j$, m_j, and c_j

	(1) ABS	(2) ASC	(3) Single AC	(4) Dual AC	(5) Single CD	(6) Multiple CD	(7) Powered seats	(8) Sun roof	(9) Cruise control	(10) Rule 3	(11) Rule 16	(12) Rule 17
ABS	1										1	1
ASC		1										1
Single AC			1							1		
Dual AC				1								
Single CD					1							
Multiple CD						1					1	
Powered seats							1					
Sun roof								1		1		
Cruise control									1			
$RFAT_j$	0.5	0.5	0.4	0.8	0.4	0.5	0.9	0.5	0.6	0.7	0.8	0.6
m_j	5	6	8	10	3	7	11	5	10	12	9	8
c_j	10	12	20	12.5	7.5	14	12.2	10	16.7	17.1	11.3	13.3

4 Subassembly Selection

The model for assembly selection presented next assumes that association rules have been extracted for a database.

4.1 The Model and Algorithm. The model presented here is known in the optimization literature as the set covering problem [22].

Denote:

I = set of functions e_i to be realized

J = set of modules V_j to be manufactured

$$a_{ij}=\begin{cases}1 & \text{if elementary function } e_i \text{ is realized in module } V_j\\ 0 & \text{otherwise}\end{cases}$$

$RFAT_j$ = reduced final assembly time, if e_i is realized in module V_j

m_j = manufacturing cost to realize module V_j

c_j represents the manufacturing cost per unit of reduced time and it is expressed as

$$c_j=\frac{m_j}{RFAT_j}$$

The decision variable is defines as

$$x_j=\begin{cases}1, & \text{if module } V_j \text{ is selected, } j\in J\\ 0, & \text{otherwise}\end{cases}$$

Based on the above notation, the following model is formulated:

$$\mathrm{Min}\sum_{j\in J} c_j x_j$$

subject to:

$$\sum_{j\in J} a_{ij}x_j \geqslant 1 \quad \text{for all } i\in I$$

$$x_j=0,1 \quad \text{for all } j\in J$$

Chvatal's [23] proposed a heuristic algorithm for solving this model.

Define $|V_j|$ = cardinality of vector V_j (number of nonzero elements in vector V_j)

Step 0. Set the solution set $J^*=\varnothing$.

Step 1. If $V_j=\varnothing$ for all j, stop; J^* is a solution. Otherwise, find a subscript k maximizing the ratio $|V_j|/c_j$ and proceed to Step 2.

Step 2. Add k to J^*, replace with V_j with V_j-V_k, and go to Step 1.

Using the data in Table 1, Table 2 has been created, where the columns represent options and modules with the corresponding values of $RFAT_j$, m_j, and c_j. For example, module 12 (represented by rule 17) includes the options "ABS" and "ASC."

The values of $RFAT_j$ are computed from (1). The manufacturing cost m_j depends on the components used to realize the modules and on the process.

The modules 10, 11 or 12 indicate that the time to manufacture two options together is lower than manufacturing each of them individually.

In iteration 1 and Step 1 of the algorithm compute: $\mathrm{Max}|V_j|/c_j = \max\{1/10,1/12,1/20,1/12.5,1/7.5,1/14,1/12.2,1/10,1/16.7,2/17.1,2/11.3,2/13.3\}=2/11.3=0.18$ with the corresponding $k=11$ is computed.

In Step 2, the set $J^*=\{11\}$ is updated and each V_j is replaced with V_j-V_{11}, which correspond to the removal of columns 1 and 6 from Table 2.

In iteration 2 and Step 1, the value

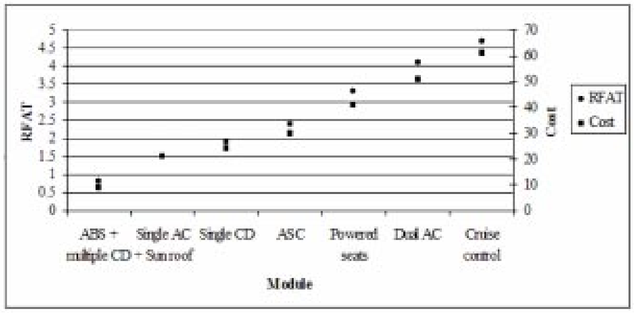


Fig. 4 Relationship between cumulative *RFAT* and cost for selected modules

$\text{Max}|V_j|/c_j = \max\{1/12,1/20,1/12.5,1/7.5,1/12.2,1/10,1/16.7,2/17.1,1/13.3\} = 2/17.1 = 0.12$ with the corresponding $k=10$ is computed.

In Step 2, the set $J^*=\{11,10\}$ is updated and each V_j is replaced with $V_j - V_{10}$, which correspond to the removal of columns 3 and 8 from Table 2.

The final solution is $J^*=\{11,10,5,2,7,4,9\}$.

4.2 Analysis of Subassemblies. The final solution $J^*=\{11,10,5,2,7,4,9\}$ generated in the previous section implies that the following modules should be manufactured {rule 16, rule 3, Single CD, ASC, Powered seats, Dual AC, Cruise control}.

Figure 4 illustrates the relationship between the cumulative *RFAT* and manufacturing cost for the selected modules.

If the first module (ABS + multiple CD) is incorporated in a subassembly, then the manufacturing cost is 9 and the *RFAT* is 0.8. Incorporating the second module (Single AC + Sun roof) results in the manufacturing cost of 21.

To meet the minimum *RFAT* requirement of 2.5, the following modules have to be realized: ABS + multiple CD, Single AC + Sun roof, Single CD, ASC, and Powered seats. The manufacturing cost to realize these modules is 41.

The minimum *RFAT* requirement of 2 can be met by manufacturing Powered seats at the cost of 30.

5 Industrial Case Study

The framework discussed in this paper has been applied to industrial data. In the industrial case study 35,989 orders were considered, each with 61 attributes. The attributes defined each customer, the car type, the model, the color of the car and all the options and alternatives necessary to fully describe each car.

5.1 Extraction and Selection of Decision Rules. It has been decided that in rules with the minimum support of 10% minimum and confidence grater than 50% will be considered for optimization (Step 1 of Fig. 3).

All rules with less than two attributes with the value of 1 have been deleted (Step 2 of Fig. 3). Finally, the domain experts were consulted to analyze the rules and delete ones that led to equivalent subassemblies.

5.2 Selection of Modules. The rules selected in Section 5.1 have been incorporated into the incidence matrix, analogous to the one shown in Table 2. The values of $RFAT_j$ were computed from (1) and the company has provided other data needed.

The results generated in this case study are intended for the negotiation of the cost structure between the suppliers and the contractor. Any delay of subassembly delivery by the supplier translates into reduced cost incurred by the contractor.

6 Conclusion

In this paper, the selection of subassemblies based on customers' requirements in the presence of the time and cost constraints was discussed. A data mining algorithm was used to determine a candidate set of modules and options to be considered for building subassemblies. The subassembly selection problem has been formulated as an integer programming model. A heuristic algorithm was applied to select the best set of subassembly structures.

This paper demonstrated that a data mining approach provides useful knowledge that could be applied to fairly price products as well impact future product designs. The proposed approach can be generalized to applications other than the wire harness case study discussed in the paper.

References

[1] Child, P., Diederichs, R., Sanders, F. H., and Wisniowski, S., 1991, "The Management of Complexity," Sloan Management Review, Fall, pp. 73–80.

[2] Kekre, S., and Srinivasan, K., 1990, "Broader Product Line: A Necessity to Achieve Success," Manage. Sci., **36**(10), pp. 1216–1231.

[3] Fisher, M. L., and Ittner, C. D., 1999, "The Impact of Product Variety on Automobile Assembly Operations: Empirical Evidence and Simulation Analysis," Manage. Sci., **45**(6), pp. 771–786.

[4] Hyer, N., and Wemmerlöv, U., 1984, "Group Technology and Productivity," Harvard Bus. Rev., July-August, pp. 140–149.

[5] Zinn, W., 1990, "Should We Assemble Products Before an Order is Receive?" Business Horizons, March-April, pp. 70–73.

[6] Kusiak, A., 1999, *Engineering Design: Products, Processes, and Systems*, Academic Press, San Diego, CA.

[7] He, D. W., and Kusiak, A., 1997, "Design of Assembly Systems for Modular Products," IEEE Trans. Rob. Autom., **13**(5), pp. 646–655.

[8] Huang, C. C., and Kusiak, A., 1999, "Synthesis of Modular Mechatronic Products: A Testability Perspective," IEEE/ASME Transactions on Mechatronics, **4**(2), pp. 119–132.

[9] Huang, C. C., and Kusiak, A., 1998, "Modularity in Design of Products and Systems," IEEE Trans. Syst. Man Cybern., **28**(1), pp. 66–77.

[10] Newcomb, P. J., Bras, B., and Rosen, D. W., 1998, "Implications of Modularity on Product Design for the Life Cycle," ASME J. Mech. Des., **120**, pp. 483–490.

[11] Simpson, T. W., Maier, J.-R. A., and Mistree, F., 2000, "Product Platform Design: Method and Application," Res. Eng. Des., **13**, pp. 2–22.

[12] Kusiak, A., and Huang, C. C., 1996, "Development of Modular Products," IEEE Trans. Compon., Packag. Manuf. Technol., Part A, **19**(4), pp. 523–538.

[13] Martin, M. V., 1999, "Design for Variety: A Methodology for Developing Product Platform Architectures," PhD Thesis, Stanford University, Stanford, CA.

[14] Dahmus, J. B., Gonzalez-Zugasti, J. P., and Otto, K., 2001, "Modular Product Architecture," Des. Stud., **22**, pp. 409–424.

[15] Erens, F., and Verhulst, K., 1997, "Architectures for Product Families," Computers in Industry, **33**, pp. 165–178.

[16] Gonzalez-Zugasti, J., Otto, K., and Baker, J., 2000, "A Method for Architecting Products Platforms," Res. Eng. Des., **12**, pp. 61–72.

[17] Jiao, J., and Tseng, M., 1999, "A Methodology of Developing Product Family Architecture for Mass Customization," Journal of Intelligent Manufacturing, **10**, pp. 3–20.

[18] Lee, H. L., and Tang, C. S., 1997, "Modeling the Costs and Benefits of Delayed Product Differentiation," Manage. Sci., **43**(1), pp. 40–53.

[19] He, D. W., Kusiak, A., and Tseng, T. L., 1998, "Delayed Product Differentiation: A Design and Manufacturing Perspective," Comput.-Aided Des., **30**(2), pp. 105–113.

[20] Agard, B., and Tollenaere, M., 2002, "Design of Wire Harnesses for Mass Customization," *Proceedings of IDMME 2002*, Clermont-Ferrand, France.

[21] Fayyad, U. M., Smyth, P., and Uthurusamy, R., 1996, *Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press, Cambridge, MA.

[22] Kusiak, A., 2000, *Computational Intelligence in Design and Manufacturing*, John Wiley, New York.

[23] Chvatal, V., 1979, "A Greedy Heuristic for the Set-Covering Problem," Math. Op. Res., **4**(3), pp. 233–235.