OPTIMIZATION OF BOOLEAN CLASSIFICATION PROBLEMS

E. Triantaphyllou et al.

For the reference see the last page.

objective optimization models therefore tend to be inadequate. A realistic model has multiple objective functions, ~~and usually these functions have~~ different weights for the decision makers. In fact, multi-objective optimization has two subfields: the identification of the nondominated solutions, and the selection of a nondominated solution where the objective functions are felt to be in a proper balance. The first-named subfield can be studied in the splendid isolation of mathematical research. The second subfield, however, straddles the boundary between mathematics and other disciplines because human subjectivity is an integral part of the selection process. One cannot just leave it to an optimization expert to formulate a model and to calculate an acceptable nondominated solution. At various stages the expert has to interrupt the computational process in order to fathom the preferences of the decision makers. Certain parameters (weights, targets, desired levels) are adjusted on the basis of new preference information, whereafter the computations proceed in a somewhat modified direction. It is not always clear, however, how the parameters should be adjusted in order to guarantee a rapid convergence towards an acceptable compromise. This is crucial because decision makers cannot spend much time on a particular decision problem. It is an illusion to think that many interruptions would be possible to elicit preference information. One or two sessions of the decision-making body, communication via the mail and the telephone in the time intervals between interruptions, and that's all.

In the 1980s, experts could work with questionnaires to fathom the preferential feelings in a group. The decision makers leisurely answered the questions in their office or at home, and returned the responses via the mail. In the next session of the group the calculated results were available for discussion [3]. Today, however, information technology provides sophisticated facilities for *group decision making*. Group Decision Rooms with networked PCs and a public screen for electronic brainstorming and weighted voting are commercially available. The sessions in a GDR have more impact than the questionnaires. First, the technology of the GDR eliminates the advantages of certain discussion techniques. The group members with strong verbal skills who usually dominate a meeting lose their grip on the silent majority ~~as soon as the buttons~~ are to be pressed. Second, the anonymous brainstorming and voting procedures promote an egalitarian attitude in the group (this may be a stumbling block in authoritarian cultures where decisions are deferred to the boss). GDR sessions create a certain commitment among the participants, possibly due to the intense communication, so that the decisions cannot easily be reverted thereafter. In summary, one may observe a power shift in a GDR which strongly affects the choice of a strategy [4].

References

[1] Bosch, P.P.J. van den, and Lootsma, F.A.: 'Scheduling of power generation via large-scale nonlinear optimization', *J. Optim. Th. Appl.* **55** (1987), 313–326.

[2] Energy Study Centre: 'National energy outlook', *Report* **ESC-42** (1987).

[3] Kok, M.: 'Conflict analysis via multiple objective programming', *Doctoral Diss. Fac. Math. Inform. Delft Univ. Techn.* (1986), Mekelweg 4, 2628 CD Delft, Netherlands.

[4] Laplante, A.: 'Nineties style brainstorming', *Te Suppl. Forbes Magazine* **25** (Oct. 1993), 44–61.

[5] Lootsma, F.A., Boonekamp, P.G.M., Coo R.M., and Oostvoorn, F. van: 'Choice of a lo term strategy for the national electricity supply scenario analysis and multi-criteria analysis', *Europ Oper. Res.* **48** (1990), 189–203.

Freerk A. Loo

Fac. Math. Informatics Delft Univ. Tech

Meke

2628 CD Delft, The Nether

E-mail address: F.A.Lootsma@twi.tudel

MSC2000: 90B50

Key words and phrases: linear programming, mixed-integer programming, integer programming, nonlinear programming, multi-objective optimization, cutting-stock load dispatching, energy modeling, group decision

Optimization in Boolean Classification Problems

There are many situations in which it is or desirable to classify objects into two mutually exclusive sets or classes. *Medical sis*, for example, has been the focus of research efforts over the past several years set of attributes, or relevant characteris describe a patient, the problem then is ification of various biological and/or ributes in order to determine the correct or classification.

ate the magnitude of the problem, conse of *breast cancer diagnosis*. Based on medical history and on the results of *phy screening* (the most effective diag- available to health care professionals), tempt to classify *breast tumors* as becious for malignancy or benign. Unfortu- all breast tumors which are suspected lignant, over 70% are later found to be rough an expensive and emotionally try- gical procedure called a biopsy [5]. In ad- almost 50% of those patients who actually east cancer are classified as benign by their ans, so that many malignancies go unrecog- [27].

decision maker, in this example the med- doctor, must infer from existing information characteristics or combinations of characteris- which are indicative of a benign or malignant mor in order to correctly classify new cases. In ir most basic form, the characteristics used to cribe each patient are represented by one or re binary attributes. That is, each object (pa- nt) may be represented by a Boolean vector in hich an attribute value is either 1 (true) or 0 alse). Often, the problem is compounded by the ct that complete information is not available. Continuing with the breast cancer example, sup- ose that the information related to all pertinent characteristics is not available due to the patient's ability to undergo certain tests because of exces- ive cost, the possibility of indeterminate test re- ults, lack of knowledge related to family histories, etc. Thus, in addition to the binary data indicating the presence or absence of a given characteristic, the attribute value or level of some characteristics may be unknown. The doctor is then faced with the problem of assessing a limited set of characteristics to determine whether a biopsy is warranted. He/She must decide if the available characteristics and/or combinations of characteristics provide sufficient information for an accurate classification of the tumor.

This problem, referred to as the *inductive inference problem* or *Boolean classification problem*, is illustrative of a vast number of similar situations throughout business, industry and medicine. Technological advances have created a 'data explosion', providing decision makers with ever increasing amounts of information. Unfortunately, this information is usually not exploited in an optimal way, and at times, not at all. Clearly, the classification problem becomes more complex as the amount of information related to the object increases. Individuals, or groups of individuals find themselves incapable of consistently and reliably handling, manipulating and analyzing the available information. As a result, the creation of computer systems capable of learning the concepts underlying the data and subsequently classifying new examples accurately and efficiently has become a practical necessity.

Background Information. As informally presented above, solving the Boolean classification problem generally involves the development of a system that learns from feature-based examples. That is, each example is described by a set of Boolean attributes. The binary vector [0 1 1 1], for instance, describes an example in which the first attribute (or characteristic) is false, and the remaining attributes are true. Each example also carries a classification: positive or negative. The goal of a *learning algorithm* is to infer from these examples a Boolean function (logical system) that is capable of accurately predicting the class of new examples. Generally, the inferred system is expressed as a Boolean function in *conjunctive normal form* (CNF) or *disjunctive normal form* (DNF).

The general form of a CNF and DNF Boolean function is defined as (1) and (2), respectively. That is:

$$\bigwedge_{j=1}^{k}\left(\bigvee_{i\in\rho_j} a_i\right) \qquad (1)$$

and

$$\bigvee_{j=1}^{k}\left(\bigwedge_{i\in\rho_j} a_i\right), \qquad (2)$$

where a_i is either A_i or $\overline{A}_i$. That is, a CNF expression is a conjunction of disjunctions, while a DNF expression is a disjunction of conjunctions. Any Boolean function can be transformed into CNF or DNF format [15].

To clarify the concepts presented thus far, suppose that the following sets of positive and negative examples are somehow known:

$$E^+ = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

and

$$E^- = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

The goal of a learning algorithm is to infer a Boolean function which correctly classifies all the examples. One such function (in CNF) is as follows:

$$(A_2 \vee A_4) \wedge (\overline{A}_2 \vee \overline{A}_3) \wedge (A_1 \vee A_3 \vee \overline{A}_4).$$

These three clauses, when are taken together, accept the previous four positive examples and reject the six negative examples.

Traditionally, there are two main methods with the goal of creating these intelligent systems: *decision trees* and *neural networks*. These tools, which have evolved over a forty-year period, represent a large portion of the literature on learning algorithms which propose to solve the Boolean classification problem. When applied to classification problems in which the goal of the system is to learn from feature-based examples, decision trees have been one of the most popular methodologies for the extraction of knowledge. Because of the natural interpretation of knowledge, symbolic decision trees can be easily translated into a set of rules suitable for use in *rule-based systems*. The size and form of the decision tree is significantly affected by the ordering of the attributes, and often the resulting tree is nonoptimal or it may be overspecialized. A complex tree is not only more difficult to validate, but as J.R. Quinlan [16] demonstrated, a simpler tree is more likely to capture structures inherent in the data.

Neural networks comprise the other extreme in artificial intelligence approaches. These systems consist of a set of programs based on the structure of biological neural systems. Knowledge is represented in the form of a series of interconnected neurons, the structure of their interconnections, and the strength of their interconnections. To the user the process is a 'black box'. Though these systems have demonstrated the ability to provide accurate classifications in many applications, the examples are classified without explanations or justifications. In an attempt to overcome this deficiency, hybrid systems which combine neural networks and rule-based systems have been developed [4]. While these hybrid systems are efficient and effective in terms of both time and storage requirements, unfortunately, an exponential number of rules may be derived [17]. This renders attempts at justification of the process virtually useless due to the complexity of the explanation. The problem is that the logical rules are not derived within a complete logical framework.

Optimization Approaches. Recognizing the need to minimize the resulting system, the problem of inferring a Boolean system from positive and negative examples was formulated as a satisfiability problem (SAT) and a method for inferring a minimal DNF system was proposed [8]. The SAT problem is next translated into an integer programming (IP) problem that is then solved by using an *interior point method* developed in [9]. The method makes use of a parameter, say k, which preassumes the number of disjunctions in the DNF system to be derived. The IP problem, if feasible, is solved and k is successively lowered until infeasibility is encountered and then it is concluded that there exists no system of size k or smaller which accepts all positive examples and rejects all negative examples. Many solution methods exist for solving the SAT problem (see, for instance, [8], [7] and [26]). Unfortunately, trying to determine a minimum size Boolean function may be computationally very difficult since it is much harder to prove that a given SAT problem is infeasible than to prove it is feasible. Thus, while the proach can be used with success on small data and a *minimal number of DNF clauses* thus be derived, when dealing with real world data, minimizing the size of the system may be neither sible nor desirable due to the vast amounts of time and storage required by such algorithms.

In [25] a logical (Boolean) function approach the classification problem has been introduced with the *one clause at a time* (OCAT) approach. Like the SAT approach, the OCAT approach formulates the *Boolean classification problem* as a series of integer programming problems. The OCAT algorithm is sequential and greedy in nature. The first iteration takes as input the E^+ and E^- sets, and generates a single clause which accepts all positive examples and rejects as many negative examples as possible. This is the greedy aspect of the method. In the next iteration, it performs the same task using the original E^+ set and a revised E^- set which includes only those negative examples not rejected by the preceding CNF clause. The iterations continue until a set of clauses is constructed which rejects all the negative examples and, of course, each clause accepts all the positive examples. This algorithm is as follows:

```
        $i = 0$; $C = \emptyset$;
DO      WHILE ($E^- \neq \emptyset$)
1       $i \leftarrow i + 1$;
2       Find a clause $c_i$ which accepts all members of $E^+$ while it rejects as many members of $E^-$ as possible;
3       Let $E^-(c_i)$ be the set of members of $E^-$ which are rejected by $c_i$;
4       Let $C \leftarrow C \cup c_i$;
5       Let $E^- \leftarrow E^- - E^-(c_i)$;
REPEAT;
```

The one clause at a time (OCAT) approach (the CNF case).

The core of the method lies in step 2, the application of a branch and bound algorithm. Through the development of new search strategies and fathoming tests, E. Triantaphyllou [19] improved the performance of the branch and bound step. Still, like all branch and bound algorithms, it suffers from exponential time complexity. However, computational experiments indicate that the OCAT approach, when combined with the branch and bound algorithm, is a very efficient method for inferring logical clauses from sets of positive and negative examples. In fact, in over half of the test cases, this approach generated a minimum number of clauses. In addition, when compared to the SAT approach of [8], OCAT and the branch and bound was found to be considerably faster while performing at the same level of predictive accuracy [19]. Thus, while the OCAT approach may not always derive an absolute minimal system, computationally it is much less expensive that the SAT approaches and therefore more applicable to real world applications.

Continued research in this area resulted in the development of two *randomized heuristics* [2]. It should be stated here that this approach is similar, in principal, to the GRASP (greedy random adaptive search procedure) approach presented in [3]. The first heuristic (RA1) was developed to overcome the exponential time complexity of the OCAT's branch and bound algorithm. That is, RA1 derives a Boolean system from positive and negative examples in polynomial (quadratic) time. The primary difference between the branch and bound algorithm and the RA1 heuristic is that in each iteration, the branch and bound attempts to reject as many negative examples as possible; while RA1 attempts only to reject many negative examples. Again, the increased speed resulted in generally larger systems. When comparing the two algorithms, A.S. Deshpande and Triantaphyllou [2] found that the branch and bound used in the original OCAT approach produced in general, fewer conjunctions and required higher CPU times than the RA1 heuristic. Additionally, it was concluded that a conjunction of the RA1 heuristic and the branch and bound method performs much better in terms of both computational time/memory requirements of the process and the size of the derived system than either approach used alone.

Faced with real-world problems, in which there is often *incomplete information* related to both the attribute values and the classifications, the goal to optimize the system becomes more desirable and necessary. Each of the methods discussed thus far have considered only positive and negative examples with complete data. That is, there is no *missing information* in the data set. Often, the complete examples represent only a portion of the available data, since in general data bases are plagued by missing information. In [1] a logical method for deriving a Boolean function from positive and negative examples was introduced in which some of the attribute values may be unknown. Since the missing information did not inhibit the classification process, these attributes are treated as 'don't care' values by the algorithm. Using a network flow algorithm, the method has been shown to efficiently derive a Boolean function with a very high predictive accuracy. The fact that the algorithm is capable of effectively handling missing information makes it more applicable to real data bases. Note, however, that the method makes no attempts at minimizing the size of the derived function.

Deshpande and Triantaphyllou [2] extended the RA1 heuristic for complete positive and negative examples, to include the use of incomplete data through the development of a second randomized heuristic, termed RA2. This method, allows not only for the inclusions of missing information in the attribute values, but it also makes use of examples which are *unclassifiable* due to the presence of missing information. That is, for some examples, the correct classification cannot be determined due to the lack of sufficient information. The objective of the second heuristic, similar to the first one, is to interactively derive a small-sized Boolean function from these three mutually exclusive sets: positive, negative, and unclassifiable examples. The algorithm consists of two phases. In each iteration of the first phase, the objective of the algorithm is to reject many negative examples while accepting all positive examples, and rejecting no unclassifiable examples. Once all negative examples have been rejected by the current set of clauses, phase II then assures that none of the unclassifiable examples are accepted by the system. When compared to the RA1 the accuracy obtained with the inclusion of unclassifiable data was always higher than the corresponding accuracy obtained without the inclusion of the unclassifiable data. This method has satisfactorily addressed the issues of efficiency and system size. Furthermore, it demonstrated that the process of extracting knowledge from examples can be expedited by exploiting the patterns contained in missing information and unclassifiable examples.

A related issue in this area is the development of approaches for partitioning large scale problems in optimal or semi-optimal ways [24]. That was done by using a graph-theoretic approach. The same approach also allows to establish lower limits on the minimum number of clauses derivable from two given collections of input examples. Also, an approach for *guided learning* of a target Boolean function is proposed in [22]. In [10], [21], and [18] the above problem was studied when the property of *monotonicity* can be established in the input data. Finally, in [13], [12] and [11] some methods were presented for dealing with *fuzziness* and uncertainty.

Concluding Remarks. Clearly, minimizing the size of the inferred system is an attractive goal. A complex system is difficult to validate, difficult to apply, and difficult to understand. On the other hand, a method which seeks to minimize the size of the system creates an inefficient process which is both computationally difficult and limits the method's applicability due to the vast time and storage requirements. In light of the success of the RA2 heuristic, the authors of this article are currently conducting research aimed at the development of an 'optimal' logical method which has the ability to handle missing information not only in the attribute values, but in the classification of the examples as well.

The new method works in conjunction with the OCAT approach. Through the application of modified B&B algorithm, CNF clauses are interactively generated such that the set of clauses, wh taken together, accepts all positive examples, jects all negative examples and neither accepts rejects any unclassifiable example. We consider this effort three optimization goals: efficiency the process, accuracy of the derived function the number of clauses which comprise the deri function. Thus 'optimal' in this sense implies derivation of a small (hopefully minimum) and curate Boolean system through the efficient ploitation of information contained in uncla able examples and, of course, the positive and ative examples.

In our current research efforts, optimizatio comes even more vital. By allowing missing

ion and unclassifiable examples, the amount of
ilable data increases and necessitates the use
learning algorithm which does not require ex-
sive amounts of time and/or memory require-
nts. In addition, the goal of a logical approach
to derive a system capable of accurately classi-
ng new examples and providing justification for
e decision. A logical system derived from incom-
ete data may encounter an example which can-
t be classified due to insufficient information.
his system must be capable of explaining why
e example is unclassifiable. That is, it has the
dditional responsibility of assisting the decision
aker in identifying the minimal amount of addi-
onal information required for classification of the
xample. In a minimal system, this information is
ore readily accessible.

See also: **Boolean and fuzzy relations;**
Checklist paradigm semantics for fuzzy log-
cs; Alternative set theory; Finite complete
ystems of many-valued logic algebras; Op-
imization in classifying text documents; In-
ference of monotone Boolean functions; Lin-
ear programming models for classification;
Statistical classification: Optimization ap-
proaches; Mixed integer classification prob-
lems.

References

[1] Boros, E., Hammer, P.L., and Hooker, J.: 'Predicting cause-effect relationships from incomplete discrete observations', *RUTCOR Report Rutgers Univ.*, no. RRR 9-93 (1993).

[2] Deshpande, A.S., and Triantaphyllou, E.: 'A greedy randomized adaptive search procedure (GRASP) for inferring logical clauses from examples in polynomial time and some extensions', *Math. Comput. Modelling* **27**, no. 1 (1998), 75–99.

[3] Feo, T.A., and Resende, M.G.C.: 'Greedy randomized adaptive search procedures', *J. Global Optim.* **6** (1995), 109–133.

[4] Fu, L.M.: 'Knowledge-based connectionism for revising domain theories', *IEEE Trans. Syst., Man Cybern.* **23**, no. 1 (1993), 173–182.

[5] Hall, F.M., Storella, J.M., Silverstone, D.Z., and Wyshak, G.: 'Non palpable breast lesions: Recommendations for biopsy based on suspicion of carcinoma at mammography', *Radiology* **157** (1988), 353–358.

[6] Hooker, J.N.: 'Generalized resolution and cutting planes', *Ann. Oper. Res.* **12**, no. 1-4 (1988), 217–239, R.G. Jeroslow (ed.).

[7] Hooker, J.N.: 'A quantitative approach to logical inference', *Decision Support Systems* **4** (1988), 45–69.

[8] Kamath, A.P., Karmarkar, N.K., Ramakrishnan, K., and Resende, M.: 'An interior point approach to Boolean vector function synthesis': *IKE Proc. 3rd Midwest Symp. Circuits and Systems*, Vol. 1, 1994, pp. 185–189.

[9] Karmarkar, N.K., Resende, M., and Ramakrishnan, K.: 'An interior point algorithm to solve computationally difficult set covering problems', *Math. Program.* **52**, no. 3 (1992), 597–618.

[10] Kovalerchuk, B., Triantaphyllou, E., Deshpande, A.S., and Vityaev, E.: 'Interactive learning of monotone Boolean functions', *Inform. Sci.* **94**, no. 1-4 (1996), 87–118.

[11] Kovalerchuk, B., Triantaphyllou, E., and Ruiz, J.F.: 'Monotonicity and logical analysis of data: A mechanism for evaluation of mammographic and clinical data': *Proc. 13-th Symp. Computer Applications in Radiology (SCAR)*, June 6-9 1996, pp. 191–196.

[12] Kovalerchuk, B., Triantaphyllou, E., Ruiz, J.F., and Clayton, J.: 'Fuzzy logic in digital mammography: Analysis of lobulation': *Proc. FUZZ-IEEE '96 Internat. Confer. (New Orleans, LA, September 8-11, 1996)*, Vol. 3, 1996, pp. 1726–1731.

[13] Kovalerchuk, B., Triantaphyllou, E., Ruiz, J.F., and Clayton, J.: 'Fuzzy logic in computer-aided breast cancer diagnosis: Analysis of lobulation', *Artif. Intell. in Medicine*, no. 11 (1997), 75–85.

[14] Kovalerchuk, B., Triantaphyllou, E., Ruiz, J.F., Torvik, V.I., and Vityaev, E.: 'The reliability issue of computer-aided breast cancer diagnosis', *Computers and Biomedical Res.* (to appear).

[15] Peysakh, J.: 'A fast algorithm to convert Boolean expressions into CNF', *Techn. Report IBM Computer Sci. RC 12913*, no. 57971 (1987).

[16] Quinlan, J.R.: 'Learning efficient classification procedures and their application to chess endgames', in R.S. Michalski (ed.): *Machine Learning: An Artificial Intelligence Approach*, Tioga Publ., 1983.

[17] Shavlik, J.W.: 'Combining symbolic and neural learning', *Machine Learning* **14** (1994), 321–331.

[18] Torvik, I.V., and Triantaphyllou, E.: 'Inferring a monotone Boolean function by asking a small number of membership questions', *submitted for publication* (2000).

[19] Triantaphyllou, E.: 'Inference of a minimum size Boolean function from examples by using a new efficient branch-and-bound approach', *J. Global Optim.* **5**, no. 1 (1994), 69–94.

[20] Triantaphyllou, E., Kovalerchuk, B., and Deshpande, A.S.: 'Some recent developments in logical analysis', in R. Barr, R. Helgason, and J. Kennington (eds.): *Interfaces in Computer Sci. and Operations Research*, Kluwer Acad. Publ., 1996, pp. 215–236.

[21] Triantaphyllou, E., and Lu, J.: 'The knowledge acquisition problem in monotone Boolean systems', in A. Kent and J.G. Williams (eds.): *Encycl. Computer Sci. and Techn.*, M. Dekker, 1998, pp. 89–106.

[22] Triantaphyllou, E., and Soyster, A.L.: 'An approach to guided learning of Boolean functions', *Math. Comput. Modelling* **23**, no. 3 (1995), 69–86.

[23] Triantaphyllou, E., and Soyster, A.L.: 'A relationship Between CNF and DNF systems derivable from examples', *ORSA J. Comput.* **7**, no. 3 (1995), 283-285.

[24] Triantaphyllou, E., and Soyster, A.L.: 'On the minimum number of logical clauses which can be inferred from examples', *Computers Oper. Res.* **23**, no. 8 (1996), 783–799.

[25] Triantaphyllou, E., Soyster, A.L., and Kumara, S.: 'Generating logical expressions from positive and negative examples via a branch-and-bound approach', *Computers Oper. Res.* **21**, no. 2 (1994), 185–197.

[26] Trick, M., and Johnson, D.: *Second DIMACS challenge on cliques, coloring, and satisfiability*, DIMACS. Amer. Math. Soc., 1995.

[27] Van Dijck, J., Verbeek, A., Hendricks, J., and Holland, R.: 'The current detectability of breast cancer in a mammographic screening program', *Cancer* **72**, no. 6 (1993), 1933-1938.

Evangelos Triantaphyllou
Dept. Industrial and Manufacturing Systems Engin.
3128 CEBA Building
Louisiana State Univ.
Baton Rouge, LA 70803-6409, USA
E-mail address: trianta@lsu.edu
Web address: www.imse.lsu.edu/vangelis

Jennifer Austin-Rodriguez
Dept. Industrial and Manufacturing Systems Engin.
3128 CEBA Building
Louisiana State Univ.
Baton Rouge, LA 70803-6409, USA

MSC2000: 90C09, 90C10

Key words and phrases: medical diagnosis, inductive inference problem, Boolean classification problem, learning algorithm, conjunctive normal form, CNF, disjunctive normal form, DNF, satisfiability problem, SAT, minimum number of clauses, one clause at a time approach, OCAT, GRASP approach, randomized heuristics, missing information, unclassifiable examples.

Optimization in Classifying Text Documents

From the 1950s onwards, the search for computerized tools and mathematical models that can speed up the *classification of large collections of documents* has been the focus of many research efforts. These efforts have been centered in developing tools that can speed up the classification of documents according to some underlying context. A current example of this situation is the *Internet*. In this worldwide conglomerate of databases, one can easily see the speed at which documents on the topic, say, 'basketball' are retrieved from among the millions of documents produced daily on the Internet. Document classification is also of paramount importance in many information retrieval applications, such as news routing [7], classification/declassification of official documents [15] e-mail filtering [27], and context derivation of electronic meetings [3].

From the 1950s onwards, various fields of the human knowledge have produced several solutions for the document classification problem (see, for example, [23], [21], and [2]). Some examples of these fields are mathematical optimization, computational linguistics, expert systems, neural networks, and genetic algorithms. These methodologies have been severely limited to some degree by the huge amounts of information, both textual and graphical, generated by today's information driven society. On the other hand, this 'technological' limitation has been the boost for the development of more efficient and effective classification procedures [15].

The purpose of this article is to exhibit some contributions of discrete optimization during the process of *automatic document classification*. This paper illustrates these contributions by presenting three cases (application areas) in which optimization is used in the classification process. The first case deals with a generic procedure for the selection of a set of *indexing terms* (*keywords* or *context descriptors*). The second case deals with the selection of an optimal set of indexing terms to minimize the overlapping of keywords used in different documents. The last case deals with the classification of text documents from mutually exclusive classes. These three cases are only a tiny sample of a vast collection of related instances in the field of information retrieval systems; see [11], [1[illegible] [25] for additional literature.

This article is organized as follows. The [illegible] section presents an overview of the document classification process. The subsequent section

Triantaphyllou, E. and J. Austin-Rodriguez, (2001), "**Optimization in Boolean Classification Problems,**" *Encyclopedia of Optimization,* (P.M. Pardalos and C. Floudas, Eds.), Kluwer Academic Publishers, Boston, MA, U.S.A., Vol. 4, pp. 176-182.