A Relationship Between CNF and DNF Systems Derivable from Examples

EVANGELOS TRIANTAPHYLLOU / Department of Industrial and Manufacturing Systems Engineering, Louisiana State University, 3134C CEBA Building, Baton Rouge, LA 70803-6409; Email: etrian@lsuvm.snc.lsu.edu

ALLEN L. SOYSTER / Department of Industrial and Management Systems Engineering, Penn State University, 207 Hammond Building, University Park, PA 16802; Email: alsie@engr.psu.edu

(Received: November 1991; final revision received: May 1992; accepted: January 1994)

In learning from examples, the main goal is to use a collection of positive examples and a collection of negative examples to derive a Boolean expression which satisfies the requirements imposed by the examples. In order to represent such a Boolean expression, the conjunctive normal form (CNF) and the disjunctive normal form (DNF) have been proposed. This paper makes two contributions. First it shows how to use any DNF algorithm to derive a CNF formula (or vice-versa). Furthermore, it demonstrates how to make efficient use of DNF algorithms which cannot handle a large number of positive (or negative) examples by using them as negative (or positive) examples and deriving CNF (or vice-versa). Therefore, the findings of this paper can be used to solve efficiently large scale learning problems. Two learning algorithms are used to illustrate the above issues.

Learning from examples has attracted the interest of many researchers in recent years. This is mainly due to the belief that any intelligent system should be able to improve its performance over time. In the typical learning problem of this type, both positive and negative examples are available and the main goal is to determine a Boolean expression that accepts all the positive examples, while it rejects all the negative examples. This kind of learning has been examined intensively in the last years (see, for instance, [2, 6, 8–13]). Typically the Boolean expression found by the system is either in the conjunctive normal form (CNF) or in the disjunctive normal form (DNF) (see, for instance, [1, 3–6, 11, 12]).

The findings of this paper reveal a useful relationship between the CNF and DNF systems derivable from the same data. This relationship can benefit approaches which attempt to solve large problems and use either the CNF or the DNF form in representing a Boolean expression. Suppose that a learning algorithm depends heavily on the number of positive (or negative) examples. If the number of positive (or negative) examples is larger than the number of negative (or positive) examples, then it is more efficient to solve a slightly different problem. That is, apply the same algorithm as before, but instead of using the original examples now use their complements. However, now the complements of the original positive (or negative) examples should be treated as the new negative (or positive) examples.

If the original algorithm derives a CNF (or DNF) expression, then in this way it will derive a DNF (or CNF) expression which will satisfy the constraints of the original data. The previous situation is similar to the strategy of solving the dual of an LP problem. Recall that if the primal LP problem has many constraints and few variables, then the Simplex approach is faster for the dual problem (which will have few constraints and many variables).

Let \( v \) be an example (either positive or negative). Then, \( \bar{v} \) is defined as the complement of example \( v \). For instance, if \( v = (1, 0, 0) \), then \( \bar{v} = (0, 1, 1) \). The following definition introduces the concept of the complement of a set of examples. Let \( E \) be a collection of examples (either positive or negative). Then, \( \bar{E} \) is defined as the complement of the collection \( E \).

1. Generating Systems in CNF and DNF Form

Define the general form of a CNF and DNF system as (1) and (2), respectively,

\[
\bigwedge_{j=1}^{k} \left( \bigvee_{i \in p_j} a_i \right) \tag{1}
\]

and

\[
\bigvee_{j=1}^{k} \left( \bigwedge_{i \in p_j} a_i \right) \tag{2}
\]

where \( a_i \) is either \( A_i \) or \( \bar{A}_i \). Then the following theorem states an important property which exists when CNF and DNF systems are inferred from collections of positive and negative examples. The proof of this theorem is based on the observation that an example \( v \) is accepted (rejected) by a conjunction \( \bigwedge_{i \in p} \bar{a}_i \) if and only if the example \( \bar{v} \) is rejected (accepted) by the disjunction \( \bigvee_{i \in p} a_i \).

**Theorem 1.** Let \( E^+ \) and \( E^- \) be the sets of positive and negative examples, respectively. A CNF system given as (1) satisfies the constraints of the \( E^- \) and \( E^+ \) sets if and only if the DNF system given as (2) satisfies the constraints of \( \bar{E}^- \) (considered as the positive examples) and \( \bar{E}^+ \) (considered as the negative examples).

This theorem will be applied to two algorithms which use learning from examples. In [11] an algorithm which
infers CNF systems from positive and negative examples is developed. In that approach, CNF clauses are generated in a way which attempts to minimize the number of CNF clauses that constitute the recommended CNF system. The strategy followed there is called the One Clause At a Time (OCAT) approach. A new and enhanced version of the OCAT approach with some extensions is given in [12]. The OCAT approach is sequential. In the first iteration it determines a single clause in CNF form which accepts all the positive examples in the $E^+$ set while it rejects as many negative examples in $E^-$ as possible. In the second iteration it performs the same task using the original $E^+$ set but the revised $E^+$ set has only those negative examples which have not been rejected by any clause so far. The iterations continue until a set of clauses is constructed which reject all the negative examples in the original $E^-$ set.

The second algorithm formulates the clause inference problem as a clause satisfiability (SAT) problem. In turn, this satisfiability problem is solved by an interior point method proposed by Karmakar, et al. [7] Let $M_1$ and $M_2$ be the numbers of examples in the $E^+$ and $E^-$ sets, respectively, and $N$ the number of atoms. In [6] it is shown that given two collections of positive and negative examples, then a DNF system can be inferred to satisfy the requirements of these examples. This approach pre-assumes the value of $k$; the number of conjunctions in the DNF system. In general, this SAT problem has $k(N(M_1 + 1) + M_2) + M_1$ clauses, and $k(2N(1 + M_1) + N M_2 + M_1)$ Boolean variables. Besides the fact that the first algorithm infers CNF systems, while the second infers DNF systems, the two approaches have another major difference. The first approach attempts to minimize the number of disjunctions in the CNF system, while the second approach assumes a given number, say $k$, of conjunctions in the DNF system and solves a SAT problem. If this SAT problem is infeasible, the conclusion is that there is no DNF system which has $k$ or less conjunctions and satisfies the requirements imposed by the examples.

### 2. An Example of Deriving CNF and DNF Systems

Suppose that the following are two collections of positive and negative examples:

$$E^+ = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad E^- = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}. $$

When the SAT approach is used on the previous data, the resulting satisfiability problem has 31 clauses and 66 Boolean variables (it is assumed that $k = 2$). Since there are more positive examples than negative ones, the complemented problem is smaller. It has 58 variables and 26 clauses. The complemented sets are as follows:

$$\bar{E}^+ = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \bar{E}^- = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}. $$

When the SAT approach is applied on the sets $\bar{E}^+$ and $\bar{E}^-$ (treating the first set as the positive examples and the second set as the negative examples), then the following DNF system is derived: $(A_1 \land A_2) \lor (\bar{A}_1 \land \bar{A}_2)$. Therefore, according to Theorem 1, the following CNF system satisfies the requirements of the original $E^+$ and $E^-$ data: $(A_1 \lor A_2) \land (\bar{A}_1 \lor \bar{A}_2)$.

Similarly, when the OCAT approach is applied on the previous $\bar{E}^+$ and $\bar{E}^-$ data, then the following CNF system is derived: $(A_1 \lor A_2) \land (\bar{A}_1 \lor A_2)$. According to Theorem 1, the following DNF system satisfies the requirements of the original $E^+$ and $E^-$ data: $(A_1 \land A_2) \lor (\bar{A}_1 \lor \bar{A}_2)$.

Table I presents some computational results when the OCAT approach is used on random test problems with $N = 30$ atoms. In this table, $|E^0|$ indicates the total number of examples and $|E^+|$ represents the number of positive examples. To understand the table, consider the third line of the first column. This represents a problem with 5 positive examples and 95 negative examples which required 2 CPU seconds to derive a set of CNF clauses (which accepts all 5 positive examples and rejects the 95

| $|E^0|$ | $|E^+|$ | Time | $|E^0|$ | $|E^+|$ | Time | $|E^0|$ | $|E^+|$ | Time | $|E^0|$ | $|E^+|$ | Time |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 1 | 1 | 200 | 10 | 6 | 400 | 5 | 3 | 500 | 35 | 194 |
| 100 | 3 | 1 | 200 | 11 | 38 | 400 | 7 | 10 | 600 | 8 | 16 |
| 100 | 5 | 2 | 200 | 18 | 18 | 400 | 10 | 10 | 600 | 11 | 15 |
| 100 | 5 | 3 | 200 | 19 | 4 | 400 | 15 | 7 | 600 | 23 | 184 |
| 100 | 7 | 2 | 200 | 51 | 212 | 400 | 16 | 8 | 600 | 44 | 41 |
| 100 | 7 | 6 | 300 | 2 | 1 | 400 | 17 | 23 | 600 | 49 | 315 |
| 100 | 7 | 3 | 300 | 3 | 2 | 400 | 36 | 282 | 600 | 83 | 300 |
| 100 | 8 | 2 | 300 | 14 | 29 | 400 | 47 | 97 | 700 | 13 | 26 |
| 100 | 9 | 2 | 300 | 14 | 25 | 400 | 49 | 700 | 18 | 30 |
| 100 | 15 | 7 | 300 | 17 | 107 | 500 | 6 | 13 | 700 | 19 | 60 |
| 200 | 1 | 1 | 300 | 22 | 102 | 500 | 7 | 15 | 700 | 19 | 15 |
| 200 | 2 | 1 | 300 | 22 | 70 | 500 | 13 | 8 | 700 | 56 | 467 |
| 200 | 4 | 2 | 300 | 24 | 12 | 500 | 16 | 38 | 800 | 64 | 739 |
| 200 | 5 | 2 | 300 | 36 | 243 | 500 | 20 | 19 | 900 | 72 | 836 |
| 200 | 6 | 2 | 300 | 71 | 524 | 500 | 34 | 73 | 1000 | 47 | 80 |

$^a$Time is in CPU seconds on an IBM ES/3090-600S machine.
negative examples). For a given problem size, e.g. $|E^0| = 100$, note how the CPU time increases as the number of positive examples, $|E^+|$, increases.

The savings in applying theorem 1 for the SAT approach can be easily determined. Compute the difference in the number of clauses and variables when $M_1$ and $M_2$ are interchanged. This difference is

$$(M_1 - M_2)(k(N - 1) + 1)$$

(clauses)

and the difference in the number of variables is

$$(M_1 - M_2)k(N + 1).$$

(variables)

Hence, when $M_1 > M_2$ (positive examples exceed negative examples) the number of clauses and the number of variables can be reduced by a factor proportional to $(M_1 - M_2)$ and a term of the order of $k \times N$ when the positive and negative examples are interchanged. For instance, if $M_1 - M_2 = 50, k = 10$ and $N = 20$, the reduction in the number of clauses and variables needed for the SAT approach would be 9,550 and 10,500, respectively.

3. Concluding Remarks

In this paper we examined an interesting relationship that exists between CNF and DNF systems derivable from collections of positive and negative examples. The findings of this paper can benefit any algorithm which derives CNF or DNF clauses from positive and negative examples. This relationship was demonstrated on two learning algorithms. Furthermore, the findings of this research can lead to increased efficiency for solving large learning problems of the type described in this paper.

References