Project

Alexey Maykov

March 12, 2007

1 Quadratic Assignment Problem

In this write-up, I talk about an application of a Quadratic Assignment Problem to a multi-class classification problem. First, I describe QAP and prove it's NP-Completeness. Then, I review an application of error correcting codes to a multi-class recognition problem. Then, I show how solution to QAP can be applied to a multi-class classification problem.

2 Quadratic Assignment Problem

2.1 Description

The Quadratic Assignment Problem (QAP) is a fundamental combinatorial optimization problem. The classical definition of the problem is the following:

There is a need to build n facilities. There are n possible locations. There is a distance between each pair of locations. There is also a transport flow between any pair of facilities. We want to place facilities in locations in a most optimal way. Here is a formal definition of a problem:

Given a set of potential facilities P and locations L, flow between facilities $w(p_1, p_2),\ p_1 \in P, p_2 \in P$ and a distance between locations $d(l_1, l_2),\ l_1 \in L, l_2 \in L$, find such a bijection function $f : P \mapsto L$, which minimizes the following:

$$\sum_{p_1, p_2 \in P} d(p_1, p_2) \cdot w(f(p_1), f(p_2))$$

2.2 NP-Completeness Proof

I'm going to use a decision formulation of a QAP to prove it's NP-completeness:

For a given k, is there such $f : P \mapsto L$, that

$$\sum_{p_1, p_2 \in P} d(p_1, p_2) \cdot w(f(p_1), f(p_2)) \leq k \qquad (1)$$

First, let's prove that the problem is NP. Indeed, for a given f and k, it will take only a polynomial time to calculate the value of the function above and check if it's greater or less then k.

To prove that the problem is NP-complete, I'll show that the TSP reduces to QAP. Here is a decision version of TSP:

For a given set V with a distance metric $d(a, b), a, b \in V$, and a number k, is there such an ordering of elements of V : $(v_1, v_2, ..., v_n, v_1)$ that the following is true:

$$\sum_{1 \leq i < n} d(v_i, v_{i+1}) + d(v_n, v_1) \leq k \quad (2)$$

To solve this TSP, let' formulate the QAP. We'll have a set P of n facilities ordered in a linear sequence: $P = (p_1, p_2, ...p_n)$. The flow between two sequential facilities is 1 and between any other facilities is 0: $w(p_i, p_j) = 1,\ if\ \|i - j\|\ mod\ n = 1,\ 0\ otherwise$. And a set L is equal to the set V with the same distance metric. The interpretation of P are stops of a salesman along his way. The QAP will place these stops into cities the most optimal way. Indeed, the formula 1 turns into 2 in such configuration.

3 ECC for classification

The single class classification problem is formulated as following: given a set X, and a value of a function $F : X \mapsto (0, 1)$ on some subset of X, find such a function $f : X \mapsto (0, 1)$ which minimizes the following $\sum_{x \in X} E(f(x), F(x))$ ([1]). In other words, the set X can contain items which belong to a class or not. $F(x)$ is equal to 1 if x belongs to a class and equals to 0 if it is not. Unfortunately, we know the value of F only for some subset of the entire set X (training set). Our task is to find such an approximation of F with minimal error. E is an error function can be equal to $\|F(x) - f(x)\|$. For example, X can be a set of all people. Function $F(x)$ is 1 if a person x is able to graduate from a department. The task of the enrollment committee is to determine whether or not a person is capable to graduate. The commission has information about a small subset of X which is a set of student who already graduated. In practice, X is usually a high dimensional linear space. I.e., each x can be represented as a set of real numbers $(x_1, x_2, ..., x_n)$, where each component is called a feature and usually normalized between 0 and 1. Usually, the output of the function f is a number between 0 and 1 which represents the probability of x belonging to the class. This number is then rounded to come up with the final answer. Very often, we search for f in a form of $f'(x, w)$, where w is another parameter based on a training set, and f is a smooth function. In a process called training, we look for such w which minimizes $\sum_{x \in T} E(f(x), F(x))$ where $T \subset X$.

In the multiclass classification problem, we have a set of classes $C_1, C_2, ...C_q$. The task is to determine for any $x \in X$, what class it belongs to. Classes are mutually exclusive. I.e., x can belong to one and only one C_i. An example of such a problem may be handwriting recognition. In this case, classes are letters of a target alphabet.

One of the usual approaches to for the multiclass recognition problem is to create a single class recognizer for each class: $f_1(x), f_2(x), ...f_q(x)$. Then, find the $argmax_{i \in (1,2,...q)} f_i(x)$ to identify the class. This approach seems wasteful. We can represent 2^q combinations using q bits while we use q bits to represent just q values. Thomas G. Dietterich in [2] describes the usage of Error Correction Codes (ECC) for the problem of multiclass classification. In there, each class is represented by a codeword, i.e. by a bit string of length k. Generally, k is not equal to q. k recognition functions are trained to output a bit vector: $(f_1(x), f_2(x), ..., f_k(x))$. The value of this function is compared to codeword for each class. The one with the shortest distance is declared as a class for x. The distance between two bit vectors is defined as a hamming distance (a number of matching bits). The article [2] describes how these codewords can be chosen. (The maximum Hamming distance between codewords and between values for the f_i for the same i but different x). The article also shows that ECC provide much better accuracy comparative to traditional methods. The interpretation is that the recognition can be viewed as a process of information transfer and ECC helps to correct errors arose during a transfer (recognition).

4 Using QAP for ECC classification

It is still unclear how to assign error correction codes to classes. While [2] provides a way to come up with a set of robust error correction codes. One of the ideas which makes sense is to put confusable classes as far apart as possible to avoid confusions. Let's say that we have a set of classes C. Then, we have a confusion metric $d : C \cdot C \mapsto Z$. We use a set of codes F. For every pair of such codes, we can calculate a hamming distance: $h : F \cdot F \mapsto Z$. To find the most optimal assignment of codes to classes, we need to find such $e : C \mapsto F$ which minimizes the following function:

$$\sum_{c_1, c_2 \in C} d(c_1, c_2) \cdot h(e(c_1), e(c_2))$$

This is QAP which we reviewed above.

In Tablet PC handwriting recognition team, this task appeared for a task of recognition of Japanese characters. The number of classes in such task is equal to a number of characters which is very high (about 4000). Solving QAP in this case wasn't practical, so this method was never applied even though it was formulated.

The QAP may also be applied to a variety of tasks such as the original problem with facilities and their locations. QAP can be also applied for placing components on electronic circuit boards or inside chips. The paper [3] provides a polynomial time approximation algorithm for QAP.

References

[1] Christopher M.Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[2] Thomas G. Dietterich and Ghulum Bakiri. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.

[3] Gregory Gutin and Anders Yeo. Polynomial approximation algorithms for the tsp and the qap with a factorial domination number. *Discrete Appl. Math.*, 119(1-2):107–116, 2002.