The Use of a Genetic Algorithm to Model Vasculature of a Dicotyledon Leaf

Daniel Zimarev

Abstract

Thick tissue vascularisation is one of the major obstacles in tissue engineering. This paper looks at a novel application of genetic algorithms – to solve tissue vascularisation problems. As a first step towards modelling vasculature in human tissue, simple vasculature of a dicotyledon (Populus sect. Aigeiros) leaf is investigated. The work consists of two parts. One presents the model and the algorithm to describe and solve 2-d vascularisation problems using genetic algorithms. It provides a description of chromosome encoding, tissue shape encoding and, most importantly, a multi-objective cost function based on two criteria: path length and number of cells vascularised. The second part of the paper compares solutions produced by different configurations of the algorithm (using 2-point and uniform crossover) to the ones found in nature. Three different phenotypes are observed, with one having strong resemblance to the natural pattern. The latter can be seen as giving insight into evolutionary basis of venation patterns. The second part also examines the general performance of the algorithm and reveals the advantages and limitations of applying genetic algorithms to the problem domain, as well as suggesting future areas of improvement.

Keywords: Tissue Engineering; Vascularisation; Genetic Algorithms; Optimisation; Binary Tree; Dicotyledon Leaf

1. Introduction

Current tissue engineering techniques can produce bone cartilage and skin tissue, but not thick tissue. One of the main barriers to achieving this is inability to create sophisticated vascularisation, necessary to provide all the cells with nutrients [1]. There are many dimensions to the problem, including material science and manufacturing aspects. However, this paper will focus solely on the development of an algorithm, which, for a given tissue shape, generates the necessary vascular geometry for the survival of the tissue. Such an algorithm has the potential to be incorporated into CAD software for automated tissue printing systems [2].

* Corresponding author. Tel.: +44 7929 914254;
E-mail address: dz260@cam.ac.uk.
Presumably, vasculature in organisms is optimal, having evolved over millions of years. Therefore, the problem can be approached not from a biochemical point of view (which is often the case [3-5]), but from an optimisation one. Among various optimisation methods, GAs stand out as a method most closely resembling the evolutionary processes found in nature. They have the capacity to produce vasculature for a given shape over a number of generations. Moreover, GAs have successfully been applied to many network optimisation problems: circuit and flow channel design, path routing and scheduling [6, 7]. A similar problem was investigated Aragón et al who used multiobjective GAs to find microvasculature for self-healing bio-mimetic materials [7]. Because vasculature is also a type of a network (representable by a graph) it is natural to apply GAs to this problem domain.

To simplify the modelling process a 3-d problem was reduced to 2-d. Given a 2-d tissue shape, the algorithm had to find a suitable vascular network. To validate the accuracy of the algorithm a simple 2-d segment of the human vascular system had to be chosen for visual comparison. Unfortunately, vasculature in the human body is mostly 3-dimensional. Any 2-d section cuts (using CT or MRI) present only a fragmented view. A solution was found by modelling the planar venation found in dicotyledon (one of two types of flowering plants) leaves. This was possible due to dicotyledon leaf venation and that of human tissue exhibiting morphological similarities at micro (net like structures) and macro (branched structures) levels (Fig.1).

Fig.1 Similarities between human (A) and plant (B) macro/micro vascularisation morphologies. A–after information contained in a micrograph (LM X 125) of a human capillary bed [8], B – after information contained in a micrograph of a Populus sect. Aigeiros (dicotyledon) tree leaf [9].

2. The Model and the Algorithm

The leaf shape [9] is subdivided into discrete Cartesian coordinate points (these points can be viewed as cells or clusters of cells) forming the tissue matrix $\mathbf{\tau}$ (1). Coordinate points are often simply referred to as ‘cells’ in this paper.

$$\mathbf{\tau} = \begin{bmatrix} x_1 & \cdots & x_n \\ y_1 & \cdots & y_n \end{bmatrix}$$ (1)

The greater number of points for a shape, the greater the accuracy of the simulation (model used had 325 coordinate points). However this also increases the solution space and, subsequently, simulation time. It also makes it more difficult for the algorithm to find a solution due to the number of combinations it has to process.

The vascular path is modelled using a binary tree. To decrease the solution space, two-level branching (7 nodes in total) is used. The chromosome hence consists of 14 integer parameters $n_{par}$: 7 ‘x’ and 7 ‘y’ coordinates (2). To prevent the vascular path from being generated outside the boundaries of the tissue shape, chromosome values are limited to the ones available in the tissue matrix (1), or $\mathbf{\rho} \in \mathbf{\tau}$.

$$\mathbf{\rho} = \begin{bmatrix} x_1 & \cdots & x_{n_{par}/2} & y_{n_{par}/2+1} & \cdots & y_{n_{par}} \end{bmatrix}$$ (2)

A binary tree based vascular path is formed by the particular way parameters are decoded and encoded within the chromosome (see Fig.2). The left ‘child’ of a node is found by $2n$ and the right ‘child’ by $2n+1$, where $n$ is the
location of the parental node. This is applied to both sides (2) of the chromosome, as each node consists of x and y Cartesian values.

\[ d = \sum_{n=1}^{N} \sqrt{(x_{2n} - x_n)^2 + \left( y_{2n \frac{n_{par}}{2}} - y_{n \frac{n_{par}}{2}} \right)^2} + \sqrt{(x_{2n+1} - x_n)^2 + \left( y_{2n \frac{n_{par}}{2} + 1} - y_{n \frac{n_{par}}{2}} \right)^2} \]  

The population \((\text{popsize})\) is formed of 20 chromosomes. Mutation \((\text{mutrate})\) and selection rates are set at 0.2. These parameters have been found experimentally to produce the fastest convergence.

The whole population of chromosomes is evaluated by the cost function, which can be viewed as a crude approximation of natural selection. There are two components to the cost function (or two optimisation criteria): (i) path length and (ii) vascularised area estimation.

The area vascularised is obtained by using two loops. Within the first loop the shortest distance is estimated from the point (or cell) to each of the segments of the path. If the distance is bellow a threshold value, the cell is vascularised and vice versa. The second loop repeats the procedure for all of the cells within the tissue matrix. Finally the total number of cells non-vascularised \(n\) is obtained.

The method of shortest point-to-line (PTL) distance calculation deserves a separate mention. The algorithm first finds an intersection point of a normal from the point to the path and the path itself (Fig.3). If the intersection point \((i_1)\) is within the boundaries of the line \((L_1-L_2)\), the normal will be the shortest PTL distance. If the point is outside the bounds \((i_2)\), the shortest distance will be to one of the ends of the segment line. The algorithm measures the distance to each of the two ends and selects the shortest one \((P_2-L_2)\) [10].
Fig. 3 A diagram illustrating the location of the shortest point-to-line distances depending on the positions of the points (P₁ and P₂) relative to the line segment L₁-L₂.

Initially, all the distances (path length and nutrient supply threshold) are calculated based on the discrete coordinate (cell) locations. However, subsequently these values are converted to anatomical dimensions using the Anatomical Length Coefficient (ALC). This is done to ‘ground’ the model to natural dimensions. \[ ALC = \frac{ALH}{CLH} \]

where \( ALH \) - Anatomical Lamina Height (distance from the stem to the very tip of the leaf), equal to 39200 \( \mu \)m according to Raven et al. [9], and \( CLH \) - Coordinate Lamina Height, equal to 18 (unitless) in the model used.

The shortest PTL distance is compared against the vascularisation threshold \( VT \). \( VT \) represents the maximum distance from the vessel at which cells are still able to receive nutrients.

In line with Multiple-Objective Optimisation (MOO) described by Haupt & Haupt [6] and De Jong [11], the two parameters, (i) path length and (ii) the number of cells non-vascularised, are multiplied by weighting coefficients \( w_l \) for length and \( w_v \) for vascularisation and added up to obtain the total cost (fitness) of the chromosome (4). The lower the path length (lower resistance to flow) and the fewer cells are not receiving nutrients, the lower the overall cost and hence better the solution.

\[ f = w_l d + w_v n \]  

(4)

The sum of the weighting coefficients \( w_l \) and \( w_v \) must be equal to one. In this particular research \( w_l=0.000001 \) and \( w_v=0.999999 \) were used. It is easy to vascularise all of the cells by having a long chaotic, intersecting path. Conversely, when the path becomes shorter and more structured, it becomes more difficult to vascularise all of the cells. For this reason \( w_l \) was chosen big enough to prevent vascular path lines from intersecting, but small enough as to not shorten the vascular path at the expense of not vascularising all of the cells. Note that the relative size of the coefficients is not comparable as length and cell numbers are not dimensionally uniform. Therefore, it cannot be said with absolute certainty that one criterion is more important than the other. However, it does seem sensible to assume that vascularising all of the cells is more important than having the shortest path.

After each iteration the chromosomes are arranged in an ascending order according to their cost. Based on the selection rate a fraction of the best chromosomes is kept. Their genetic material is used to recreate the population during crossover. Two different types of crossover are used in this research.

a) A variation of a two-point crossover, where the first locus point (location within the chromosome at which gene exchange occurs) is selected randomly. However, it must be within the bounds of \( x \) coordinate values (first half of the chromosome) in (2). The second locus point is pre-determined by the first one and exchanges the corresponding \( y \) values between the two parents. This is to avoid accidental creation of path points, which are outside the boundaries of the tissue matrix.

b) Unlike in two-point crossover, the intensity of genetic material exchange in uniform crossover is far greater. When creating a new individual the program stops at the location of each gene (or variable) and decides whether to take the value from the paternal or the maternal chromosome. The selection probability for any of the parent genetic material is set at 50%. Like the two-point crossover, this crossover is unique in that in addition to exchange of the \( x \)
value, the exchange of the corresponding $y$ value also has to occur. The corresponding $x$ and $y$ values have to be from the same parent, and hence from the tissue matrix.

In both cases parents are selected randomly from the surviving chromosomes. Only one offspring is produced from each pair. The number of crossover iterations corresponds to the number of chromosomes that need to be re-created in the population.

After crossover, the mutation operator introduces a number of random changes to the chromosomes. At each instance, like in crossover, an $x$ value and a corresponding $y$ value are replaced by random values from the tissue matrix. The number of mutation iterations is the positive infinity of the product of $npar$, $popsize$ and $mutrate$. All but the two best chromosomes are susceptible to mutation. This is to avoid accidentally destroying the very best of genetic material and experiencing a sudden jump in the lowest cost of a subsequent generation.

### 3. Results and Discussion

Overall, it has been observed that the morphology of venation was dependent on 3 factors: (i) weight coefficients assigned within the cost function, (ii) nutrient delivery (vascularisation) threshold ($VT$) and (iii) shape, or the coordinate matrix $\tau$ (1), of the tissue being vascularised. The higher the number of cells the tissue contained and the more parameters were present within the chromosome, the bigger the solution space the algorithm had to examine, making it more difficult to find an optimal solution.

It has also been observed that the two optimisation components within the algorithm’s cost function performed different roles. It was relatively easy to vascularise all of the cells (in any chaotic way possible). But it was more difficult to vascularise all of the cells while using a short path. Hence, it was the length function that was the primary optimisation vehicle, gradually and continuously generating improved (lower cost) solutions.

Three scenarios have been examined in greater detail: (a) a two-point crossover based algorithm running for 500 iterations, (b) a two-point crossover based algorithm running for 1000 iterations and (c) a uniform crossover based algorithm running for 1000 iterations (see Fig. 4). During each scenario the algorithm ran 20 times using a vascularisation threshold $VT=20,000 \, \mu m$. In all three instances three phenotypes (vascular patterns) emerged: ‘U’, ‘Y’ and ‘T’-like.

a) In the first session (Fig.5, Fig.6, & Fig.7) the mean cost after 500 generations was $50.45E-3$ with a population standard deviation of $5.75E-3$. 

---

**Fig. 4** Graphs showing the mean, minimum and maximum costs at each iteration for trials a (A), b (B) and c (C).
The ‘Y’ like pattern (Fig.6) was expected, because the model was based on binary branching. What was unexpected is that a binary tree based encoding managed to produce linear ‘U’ like features (Fig.5) and even ternary ‘T’-like branching (Fig.7). This phenomenon demonstrates one of the advantages of the application of GAs to vascular modelling. They require very little knowledge about the actual system and will adapt accordingly to what is truly optimal with the means available.

Fig.8 illustrates how the transformation occurs. A ‘U’ pattern (A) is obtained by shortening of the branches to the extent that their start and end coordinates merge. Likewise, a ternary pattern (B) is obtained by shortening of the branches (a), change of their direction (c), and, most importantly, an overlap occurring at the stem (b). The overlap, however, results in an artificial increase in cost (due to the length component), which might help to explain, why ternary morphology (Fig.7) had the highest cost and the rarest occurrence in all three scenarios tested.
Another interesting observation was that the cost of two completely different solutions could be very similar (see Fig.5: B/C & Fig.6: A/B/C). This suggested that there could be not one, but several equally optimal solutions. The observation was in line with findings by Roth et al, suggesting that leaf venation, which is dependent on multiple functions, could have evolved not one, but several equally optimal patterns [12]. The findings were also characteristic of MOO, where several solutions can lie on the Pareto optimal front.

b) Surprisingly, during the second trial (Fig. 8) the mean cost after 1000 generations was slightly higher at 51.39E-3 (0.93E-3 increase), having a population standard deviation of 4.49E-3. It is very likely that (statistically) for a larger sample size the costs would have been identical. 1000 iterations not producing a lower cost was indicative of the cost reaching steady state and 500 iterations being sufficient for a 2-point crossover. In fact, cost v generation graphs revealed that the most significant change was occurring within ≈ 215 generations.

c) During the third trial 2-point crossover was replaced with uniform crossover. The results produced (Fig. 9) had a mean cost of 49.09E-3, with a population standard deviation of 11E-3. The standard deviation was significantly higher than that for a 2-point crossover, possibly indicating that uniform crossover was more effective at exploring the solution space. Although the mean cost was lower than that of 2-point crossover, a larger sample size would have been needed to confirm this with certainty.
Fig. 10. The phenotypes emerged after 1000 generations using uniform crossover. A) ‘U’-like appeared in 45% of all cases. The mean cost was $44.41E-3$, with a population standard deviation of $4.40E-3$. The lowest cost achieved within the phenotype was $38.453E-3$. B) ‘Y’-like phenotype emerged in 50% of cases. The mean cost was $50.06E-3$, with a population standard deviation of $5.12E-3$. Lowest cost = $44.3E-3$. C) ‘T’-like phenotype emerged in 5% of cases, having a cost of $52.37E-3$.

The phenotype having the lowest minimum cost (in all three trials) had a linear ‘U’-like morphology. The second best solution exhibited binary ‘Y’ branching and the third best solution had ternary ‘T’ structures. This raised a question of why ‘nature-like’ patterns did not possess the lowest cost. This might have been due to: (i) particular weight coefficient values selected, (ii) vascularisation threshold value selected, (iii) optimisation criteria omitted (e.g. pressure), (iv) nature-like solutions not being the most optimal or (v) the encoding mechanism favouring a linear or binary patterns.

‘U’-like pattern showed a decreased median cost, when switching from 500 iterations to 1000 and a further decrease when using uniform crossover instead of a 2-point one at 1000 iterations. ‘Y’ and ‘T’ phenotypes did not exhibit improved median cost over all of the subsequent trials, but did see a decrease in the median cost after a switch has been made from the use of a 2-point crossover to uniform crossover. In general, uniform crossover appeared to outperform the 2-point crossover.

After a 500-generation optimisation 70% of solutions had a ‘U’-like pattern, whereas after 1000 generations only 45%, indicating that some of the solutions were transitioning from phenotype ‘U’ to phenotype ‘Y’.

The algorithm managed to produce nature-like patterns using only 2 relatively simple parameters. This stood in contrast to the complexity of modelling all of the biochemical variables.

A feature that often emerged in many of the ‘T’ and ‘Y’ like solutions was a straight middle core. In addition to being ‘nature-like’, it was also an unusual feature from a probabilistic perspective. The algorithm could randomly pick coordinates points at any location on the tissue matrix. It selecting for the ones forming a straight line was unusual and indicated a beneficial trait which was propagated over the generations.

A ‘ternary’ phenotype was the rarest one to occur. This might have been the result of the model being based on binary branching. In order for a ternary pattern to appear, a small overlap in the path had to occur at the stem of the leaf (Fig 7, B, b), which not only increased the cost, but also had a low probability of occurrence. The abundance of ‘U’-like solutions indicated that the algorithm had no difficulty in transitioning to a lower magnitude (binary-to-unary) branching. This suggested that ternary based branching was better suited for modelling vascular networks.

4. Conclusions and Future Directions

The paper presents a novel application of genetic algorithms to a biological vascularisation problem. The main contribution of the paper lies in the method presented for solving vascularisation problems, including: chromosome encoding, tissue shape encoding and, most importantly, the cost function. Although the model is 2-dimensional, it can be scaled to 3-d by adjusting the tissue matrix and the cost function.

During experimental trials, the algorithm managed to produce nature-like patterns, which suggests that the optimisation parameters selected may be present in the natural processes as well.

The research also evaluates the feasibility of applying GAs to vascularisation problems. On the one hand, GAs are well suited for the problem domain. They are very adaptive (can produce unary paths from a binary based system) and require very little knowledge about the system. This is well suited for the complex vascular systems. On the other hand, the algorithm suffers from a common GA shortcoming – premature convergence. The algorithm tends
to get stuck a local minimum, without examining the whole of the solution space. Although uniform crossover seems to be more successful at doing this compared to 2-point crossover, it remains to be examined whether there are crossover, selection and any other operators that could be effectively implemented to prevent premature convergence in this application domain. Currently, highly fit individuals can survive indefinitely. Perhaps their lifespan could be limited to one generation only. The variation could also be preserved by giving unfit individuals a small chance of mating, instead of none. Another limitation, a large solution space, could be reduced by introducing a ternary-branching based model. The solutions space could be further decreased by adding additional optimisation criteria, e.g. pressure.

Acknowledgements

The author would like to thank Prof Mike Hughes for useful advice and suggestions.

References