Induction of qualitative models using discrete Morse theory

Jure Žabkar and **Gregor Jerše** and **Neža Mramor** and **Ivan Bratko**
Laboratory of Artificial Intelligence, Faculty of Computer Science, SI-1000 Ljubljana, Slovenia

Abstract

Qualitative models are often a useful abstraction of the physical world. Learning qualitative models from numerical data is a possible way to obtain such an abstraction. We present a new approach to induction of qualitative models from numerical data which is based on discrete Morse theory (DMT). Our algorithm QING (Qualitative INduction Generalized) has a firm theoretical background in computational topology. This makes it possible to extend the capabilities of state-of-the-art algorithms for qualitative modelling substantially. The output of QING is a labeled graph, which enables a visualisation of the qualitative model. Induced qualitative models can also be used for numerical regression by applying the Q^2 method. To illustrate the power of QING we present its application on an artificial function, add noise, and finally show how it performs on a dynamic domain such as inverted pendulum.

Introduction

Every day more and more data from real-life processes, such as measurements of weather variables, measured data from simulations, technological processes, chemical reactions etc is being recorded. Only a small subset of this data is later analyzed in hope to obtain the models that would imitate the processes from which the data was gathered. Such models enable the experts to run simulations and make predictions about something before it really happens. Numerical prediction and quantitative modelling, both suited for such a purpose, are common tasks in machine learning. Their quality is usually judged on the numerical accuracy they achieve on yet unseen data. It often happens that numerically accurate model fails to explain the underlying processes hidden in the data or the explanation is too complex. Recently, quantitative machine learning has been combined with qualitative learning in the method called Q^2 learning (Šuc, Vladušič, & Bratko 2004) which turned out to be very successful. In Q^2, a qualitative model is induced first and is later used to force the numerical model to be consistent with the induced qualitative constraints. This usually contributes to better accuracy of numerical predictions while qualitative models themselves are useful as comprehensible models that intuitively explain how the system works.

Qualitative models have been neglected for several reasons. Not only the induction of a qualitative model is a complex task but it is also not possible to estimate the true value of the induced model. How good is it? How does it compare to the model induced by another algorithm? There are several estimates for numerical models but none for qualitative models. More or less it is the matter of one's taste and habit when one decides which algorithm to use.

In this paper we present an algorithm QING (Qualitative INduction Generalized) which is based on discrete Morse theory (DMT) (Forman 2001) from the field of computational topology. We consider this powerful theoretical background in mathematics an advantage. Given a learning set of examples with numerical attributes and a numerical class variable, the goal of QING is to perform qualitative analysis of class variable w.r.t. attributes. The output of QING is a qualitative field (qfield), a set of critical points and a labeled qualitative graph (qgraph), which is a visualisation of the qualitative model. Detailed definitions of these terms are given in section 'Algorithm QING'. Induced qualitative models can also be used for numerical regression by applying the Q^2 method. The main difference between QING and other algorithms for induction of qualitative models is in attribute space partitioning. Unlike algorithms that split on attribute values (e.g. trees, rules), QING triangulates the space (domain) and constructs the qualitative field which for every learning example tells the directions of increasing/decreasing class. Doing so it finds all maxima, minima and saddles, so called *critical points*. One of the main features of QING is *canceling*, a direct way to handle noisy data. Another important advantage over state-of-the-art algortihms is that monotonic qualitative constraints are generalized so that most of the qualitative ambiguity is removed. This paper is mainly focused on the theoretical background of our approach that greatly contributes to many features of QING. However, we also present some experiments to show how QING works in practice, how it handles noise and how it compares to state-of-the-art algorithms for induction of qualitative models.

The most relevant of related work is algorithm QUIN which we briefly summarize in 'Related work'. Algorithm QING is described and accompanied with a simple example in section 'Algorithm QING'. In section 'QING with inverted pendulum' we aplly QING to the dynamic system of inverted pendulum. For mathematically oriented readers we summarize discrete Morse theory in section 'Discrete Morse theory'.

Related work

The problem of automatic induction of qualitative models has been addressed several times (Bratko & Šuc 2003; Kuipers 1994). In one way or another, most of the approaches use mainly background knowledge and not learning examples. The first algorithm for induction of qualitative trees from numerical data was QUIN.

QUIN (QUalitative INduction) looks for qualitative dependencies in numerical data and induces qualitative trees to express such dependencies. The induction process is similar to the induction of decision trees (Breiman *et al.* 1984; Quinlan 1992). In a qualitative tree the leaves are labeled with MQCs (monotonic qualitative constraints), a kind of monotonicity constraints that are widely used in the field of qualitative reasoning (Kuipers 1994).

An MQC is best described by an example, let's say $y = M^{+}(x)$. This says that y monotonically increases whenever x increases. In general, MQCs can have more than one argument, e.g. $z = M^{+,-}(x, y)$ says that z monotonically increases whenever x increases and z monotonically decreases when y increases. Each qualitative constraint in an MQC requires a strict increasing/decreasing dependency in its variable while keeping the other variables constant. Therefore, an MQC may be qualitatively ambiguous. Qualitative ambiguity occurs when the qualitative value of the constraint cannot be predicted (e.g. the qualitative change in $z = M^{+,-}(x, y)$ cannot be determined in the case of x and y both changing). The degree of fit between the data and an MQC is evaluated by two measures: *qualitative consistency* and *qualitative ambiguity*. Qualitative consistency of an MQC is the percentage of the learning examples that are qualitatively consistent with the MQC. Qualitative ambiguity is the percentage of examples for which the MQC allows ambiguous predictions.

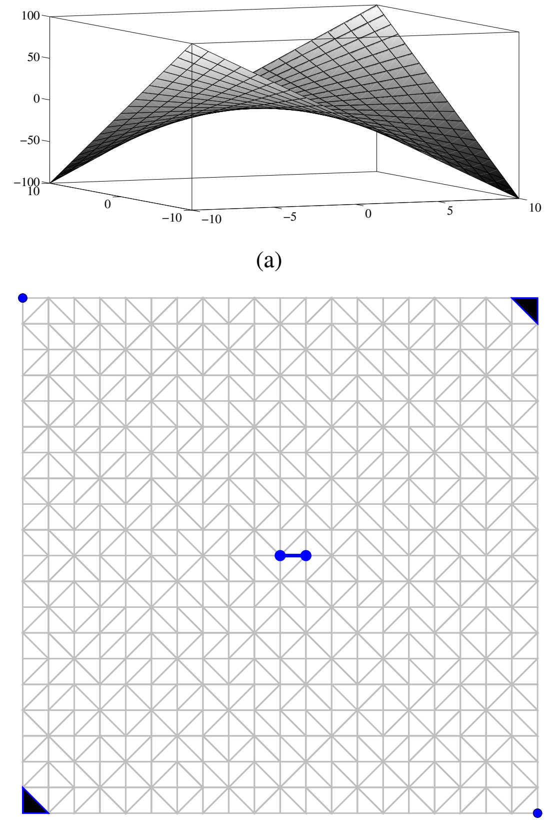
The QUIN algorithm has quite a high complexity. Empirical results (Bratko & Šuc 2003; Šuc, Vladušič, & Bratko 2004) show that QUIN can handle noisy data and, at least in simple domains, produces qualitative trees that correspond to human intuition.

Algorithm QING

QING's task is to perform qualitative analysis of continuous class variable f w.r.t. given attributes $(x_1, \ldots, x_n)$, where n is the dimension of the attribute space. For simplicity we will in this paper restrict ourselves to two attributes. Theoretically, QING works for any dimension n but is practical for $n \leq 5$ due to the complexity of triangulation. The input to QING is a set of learning examples with continuous attributes. Its output is:

- a qualitative field, (*qfield*)
- a set of *critical points* – minima, maxima and saddles of f, where in the case $n > 2$ the saddle are of different types,
- a qualitative graph, (*qgraph*)

Definition A qfield is a qualitative model represented as a set of pairs (p_i, p_j) which determine vectors pointing in the direction of incresing f. The points p_i in attribute space can



(a)

(b)

Figure 1: Function $f(x, y) = xy$ and the triangulation of its domain with two minima (circles), two maxima (triangles) and a saddle in the middle.

be either data points or midpoints between the data points, i.e. centers of mass of the segments and triangles forming the triangultaion. An example of qfield is shown in Fig. 2.

The qfield determines the critical points of f. They are simply the points which do not appear in any one of the pairs (p_i, p_j).

Definition A qgraph is a labeled graph describing the qualitative behaviour of f. The vertices are in the critical points and two critical points are connected if a path along which the function values monotonically increase. It is an abstraction of qfield, ment as a visualization of the qualitative model. An example of qgraph is shown in Fig. 3.

To be more illustrative, the description of the algorithm is accompanied with an example $f(x, y) = xy$ defined on an orthogonal mesh (see Fig. 1(a)) on the domain $[-10, 10] \times [10, 10]$.

Before we continue, let us slightly extend the notation of an MQC: $f = M^{c}_{(x)}$ means that f stays constant with increasing x. We also note here that the specific qualitative ambiguity described in section 'Related work' is removed in QING – the values of all the variables may change simultaneously.

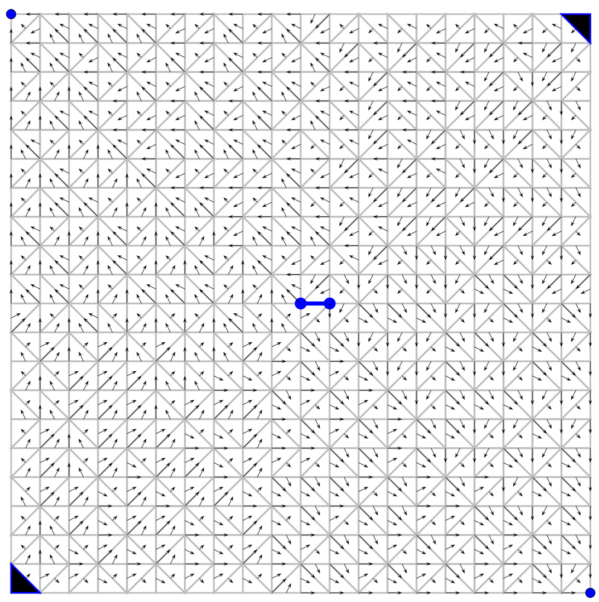


Figure 2: Qualitative field for $f(x, y) = xy$. The arrows point in the direction of function decrease.

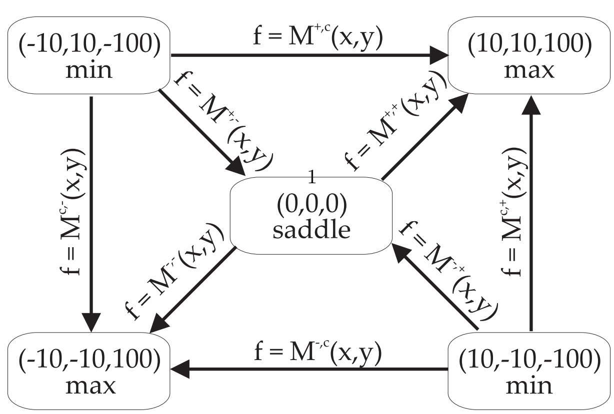


Figure 3: Qualitative graph for $f(x, y) = xy$.

The outline of the QING algorithm is as follows. Learning examples are represented as points in the attribute space, each point having assigned a value of its class variable. The domain is triangulated in order to be analysed with discrete Morse theory. Critical points are reconstructed using the algorithm of (King, Knudson, & Mramor Kosta 2005). Canceling is performed to remove the noise.

In the following paragraphs we explain each main step of the algorithm in more detail followed by examples. We finish this section with the analysis of QING's complexity.

Preprocessing

In the topological setting, learning examples are represented as points in $\mathbb{R}^n$, where n is the number of attributes, and the class variable f represents the values of a smooth Morse function in these points. In the case $n = 2$, a set of points $\{(x_i, y_i, z_i), i = 1, \dots, k\}$ which represent sampled values of a function $z = f(x, y)$ over some domain $D \subset \mathbb{R}^2$ is given, and our goal is to analyse the function f using DMT to obtain a qualitative behaviour of f. To do so we first triangulate our domain D. The class values at these points are extended to a discrete Morse function defined on the triangles. In QING we use Delaunay triangulation implemented in a free software library Qhull (Barber, Dobkin, & Huhdanpaa 1996) which is very robust and works in arbitrary dimension. Since triangulation is the basis for further analysis it is worth using it carefully. Delaunay triangulation triangulates the convex hull of the given points causing some undesired effects on the edge, namely, triangles connecting distant points appear. To avoid this we embed our points in an artificial polygon, triangulate and remove the triangles that connect to the points on the polygon.

Obtaining qualitative model

To calculate the critical points of a function on a discrete set of points we use discrete Morse theory of Forman (Forman 2001). Critical points are reconstructed from the qualitative field which is obtained using the algorithm of (King, Knudson, & Mramor Kosta 2005). Possible pairs of critical points with function values differing by less than a given margin (parameter persistance) are cancelled. This becomes useful in noisy domains to set the threshold for noise reduction, where persistance is set to the value of the measuring tolerances at data acquisition.

Critical points together with the qfield represent a qualitative model of our function, the class variable. So described, the qualitative model could be used in Q^2 learning but it still lacks a comprehensive explanatory power. Especially in higher dimensions, it is too complex for a human to comprehend. Therefore we abstract the qualitative field to a qualitative graph which serves as a visualization tool.

Algorithm complexity

The algorithm consists of three major steps: constructing a triangulation and a discrete vector field on it, and constructng the qgraph connecting the critical points. In the first step, an additional feature is the possibility of cancelling neighboring pairs of critical points where the values differ by less than a given margin, which is an efficient method for dealing with noise. The complexity of this first step is $O(h)$ without canceling, and $O(h^{2\times\lceil\frac{d}{2}\rceil})$ with cancelling, where h is the number of points and d is a dimension of the attribute space. The second step requires for each critical point a search through the paths leading through this critical point. The complexity of this step is $O(N)$, where N is the number of triangles. The last step requires a linear search through the points and therefore has the complexity of $O(h)$ where h is a number of learning examples (i.e. points).

How QING handles noise

Noise is disturbing but inevitable in real data. Therefore it is very important that the algorithm is able to deal with it and still induce a usefull model. QING has a straightforward solution to this problem. Its only parameter, *persistence*, cancels the pairs of critical points that differ in function values for less than the persistence.

To demonstrate canceling in practise we added 10% noise to our artificial domain $f(x, y) = xy$. Fig. 4 shows how different values of the parameter persistence influence the qualitative field. Starting with persistence 0, which corresponds to assuming that there is no noise, we encounter many critical points in the qfield. Increasing persistence we finally come to the point where the qgraph very much resembles the one on Fig. 3 with no noise. Both qgraphs are isomorphic, i.e. qualitatively equal. Inspite of noise we managed to discover the correct qualitative model. In practice, domain experts can usually asses the persistence value (e.g. the measuring tolerances) very well.

QING with inverted pendulum

The inverted pendulum (also known as 'pole and cart') is a well known dynamic domain that is, due to its simplicity, often used in experimenting with new algorithms. The system is shematically shown in Fig. 5. Equations 1 and 2 give its physical model. To build a qualitative model of the inverted pendulum we would have to model both equations. Since the procedure is the same, we choose to present only the more complex half of the qualitative model, $\ddot{x}$, and omit $\ddot{\varphi}$.

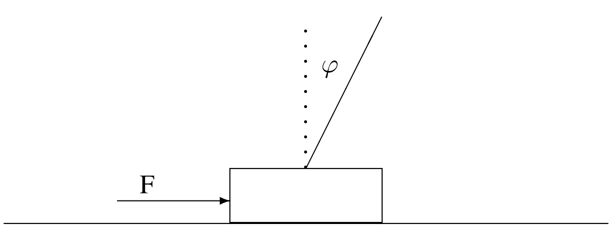


Figure 5: Inverted pendulum, also known as pole and cart.

$$\ddot{x} = \frac{4F + 2lm\dot{\varphi}^2 \sin\varphi - 1.5mg\sin 2\varphi}{4M + 4m - 3m\cos^2\varphi} \quad (1)$$

$$\ddot{\varphi} = \frac{(M+m)g\sin\varphi - F\cos\varphi - \frac{1}{2}ml\dot{\varphi}^2\sin\varphi\cos\varphi}{\frac{1}{6}(4M + 4m - 3m\cos^2\varphi)l} \quad (2)$$

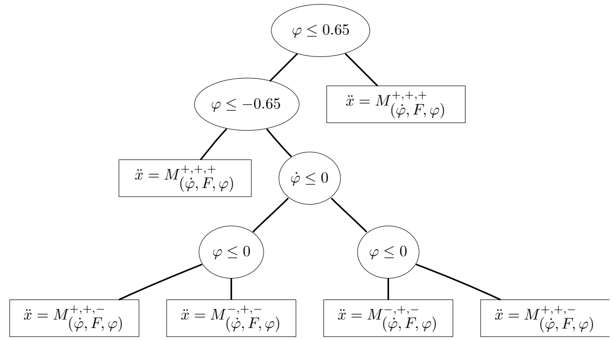


Figure 6: Qualitative tree for $\ddot{x} = \ddot{x}(\varphi, \dot{\varphi})$ built analytically from Eq. 1.

Since the equations are known, a straightforward way to obtain the qualitative model, would be to calculate the derivatives $\frac{\partial\ddot{x}}{\partial F}$, $\frac{\partial\ddot{x}}{\partial\varphi}$ and $\frac{\partial\ddot{x}}{\partial\dot{\varphi}}$ and look for the areas where they are positive/negative. By hand, with some approximations, we can get the qualitative tree shown in Fig. 6. Approximations are necessary because the area in $\mathbb{R}^2$ where $\frac{\partial\ddot{x}}{\partial\varphi}$ is close to 0 is an ellipse and using a qualitative tree, we can only approximate it with a rectangle.

Analytical solutions are nice to play with but in practise we often have only data, obtained by a sampling some process. For the sake of experiment, we use Eq.1 to obtain a data sample. Without loss, we neglect F. Our domain is therefore a plane spanned by φ and $\dot{\varphi}$, specifically, a rectangle $[-\pi/2, \pi/2] \times [-10, 10]$. To keep things simple we again have an orthogonal mesh and no noise.

On this data we use QUIN to construct a qualitative tree of depth 6 with 27 nodes, of which 14 are leaves – 7 $M^-(\dot{\varphi})$ and 7 $M^+(\dot{\varphi})$. The root splits on $\dot{\varphi} \leq -0.5$. All internal splits are made on different values of φ. As QUIN says, the coverage is perfect and there is no qualitative ambiguity in this tree. We can of course tell QUIN to build a smaller tree. The one of depth 3 has 8 leaves and its splits are the same as those to the third level in the larger tree.

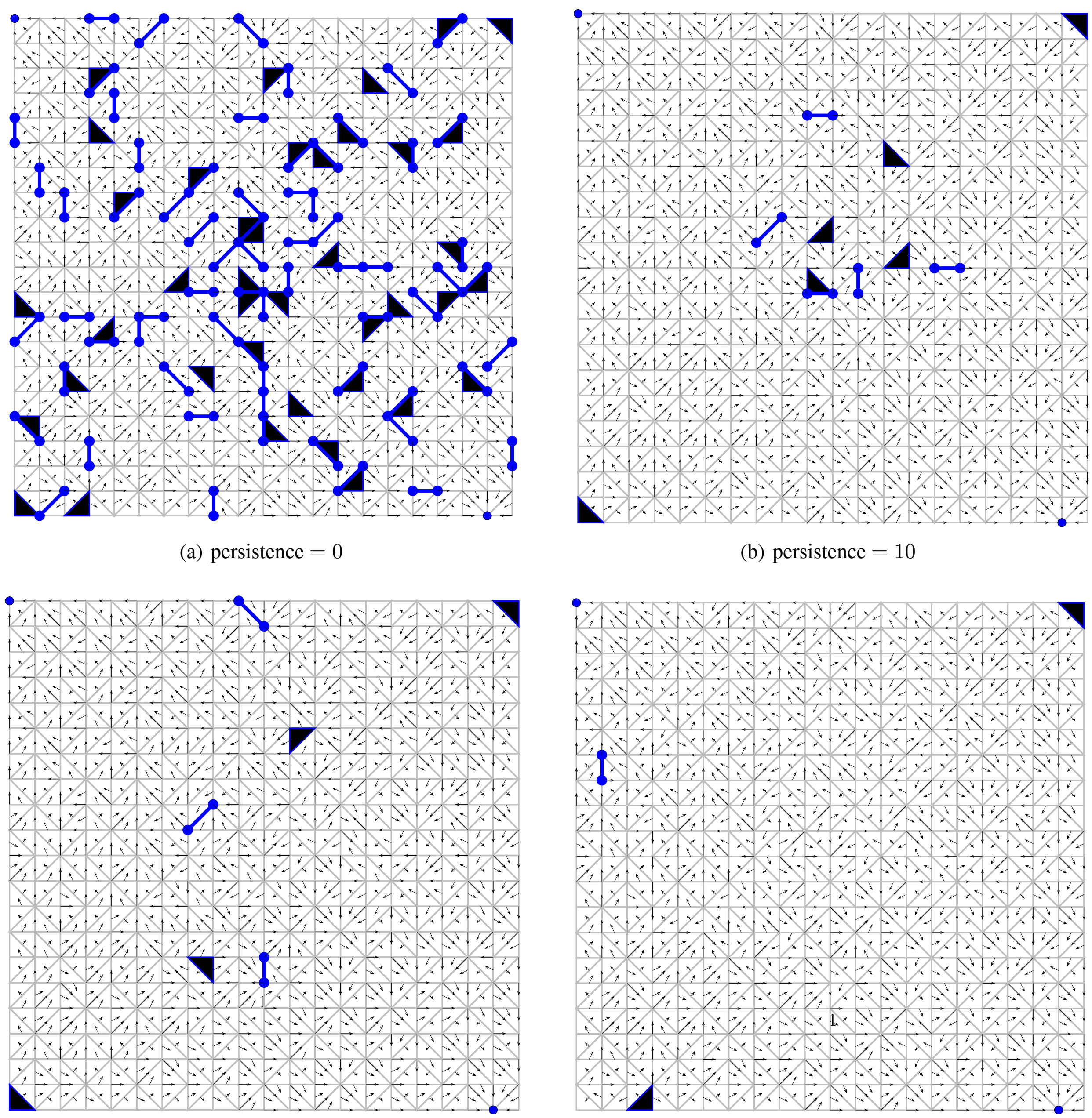
At the end, we use QING on the same data. The induced qualitative graph, Fig. 7, has 8 nodes (critical points) and 15 segments (MQCs) between them.

Technically speaking all three models are graphs so we can compare them simply by looking at their complexity.

Discrete Morse theory

In this section we review the basics of Forman's discrete version of Morse Theory (Forman 2001).

In the classical, smooth version of Morse Theory, a Morse function is a function defined on a smooth manifold M of dimension n, which has only nondegenerate critical points. In our case, M will be a domain in Euclidean space $\mathbb{R}^n$, and in this case a critical point p of a function $f : M \to \mathbb{R}$, is a point where $\mathrm{grad}\, f = 0$, i.e. the linear term in the Taylor expansion of f around p is 0. A critical point is nondegenerate if the second degree term in the Taylor expansion is nonzero. In the neighbourhood of a nondegenerate critical point, the



(a) persistence = 0

(b) persistence = 10

(c) persistence = 15

(d) persistence = 23

Figure 4: Domain $f(x, y) = xy$ with added 10% noise. Different values of parameter persistence are used to show how noise is removed through canceling of critical points.

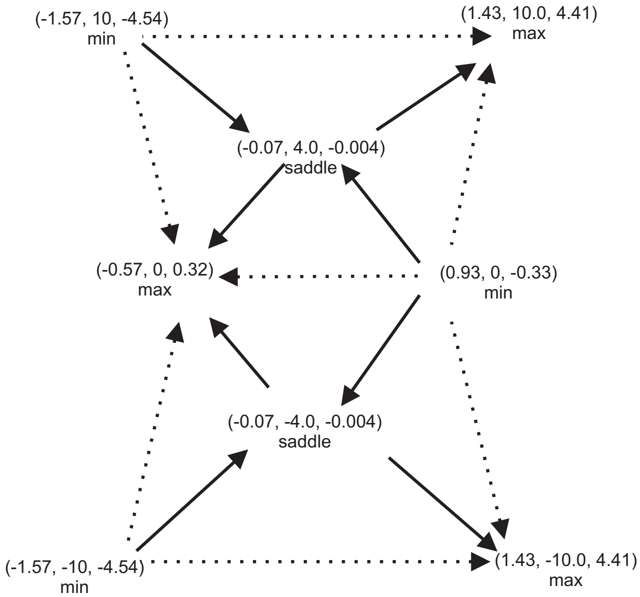


Figure 7: Qualitative graph for $\ddot{x} = \ddot{x}(\varphi, \dot{\varphi})$ built by QING.

function f can be expressed as $-\sum_{i=1}^{k} u_i^2 + \sum_{i=k+1}^{n} u_i^2$, where the number of negative terms k is called the *index* of p. In the case $n = 2$, a critical point of index 0 corresponds to a minimum, a critical point of index 1 to a saddle, and a critical point of index 2 to a maximum. In higher dimensions, saddles of different types exist. A nondegenerate Morse function determines a flow on the manifold M which corresponds to the vector field $\mathrm{grad}\, f$. A good introduction to Morse theory is (Milnor 1963).

In the discrete version of Morse theory, a triangulation of the domain M is given. A discrete Morse function f associates a value to each simplex in the triangulation, and satisfies the following conditions. For each simplex α there is at most one simplex $\beta^{(k+1)}$ which contains α as a face such that $f(\beta) \leq f(\alpha)$, and there is also at most one faces $\gamma^{(k-1)}$ of α such that $f(\gamma) \geq f(\alpha)$.

As we can see from these two conditions, the values of a Morse function generally increase with dimension, with one possible exception. It is easy to see that the two conditions above are exclusive, and so each simplex appears in at most one pair $(\alpha^{(k)}, \beta^{(k+1)})$, where α is a face of β and $f(\beta) < f(\alpha)$. A simplex $\alpha^{(k)}$ is a *critical simplex of index* k, if it does not appear in any such pair, i.e. if the function values on all its faces are lower, and the function values on all simplexes which contain it as a face are higher.

The collection of pairs $F = \{(\alpha^{(k)}, \beta^{(k+1)})\}$ with α_k face of β_{k+1} and $f(\beta) \leq f(\alpha)$ is the discrete analogue of the gradient vector field of a smooth function f. The discrete analogue of a trajectory of the gradient vector field is a V-path which is a sequence of simplices

$$\alpha_0^{(k)}, \beta_0^{(k+1)}, \alpha_1^{(k)}, \beta_1^{(k+1)}, ..., \beta_r^{(k+1)}, \alpha_{r+1}^{(k)}$$

such that pair $(\alpha_i, \beta_i) \in F$, for each $i = 0, 1, .., r$, $\alpha_i \neq \alpha_{i+1}$ and α_{i+1} (as well as α_i) is a face of β_i. Then $f(\beta_i) < f(\alpha_i)$ because (α_i, β_i) belongs to F and $f(\alpha_{i+1}) < f(\beta_i)$ because α_{i+1} is a face of β_i (but (α_{i+1}, β_i) does not belong to F). A V-path corresponds to a path through the simplices in M along which f decreases.

A discrete gradient vector which has no nontrivial closed paths, i.e. no V-paths such that $r \geq 0$ and $\alpha_0 = \alpha_{r+1}$ corresponds to a discrete Morse function (Forman 2001). So if we want to extend a function given on set of vertices to a discrete Morse function on the entire triangulation, we only have to find a discrete vector field that has no nontrivial closed paths (King, Knudson, & Mramor Kosta 2005).

Conclusions and further work

We applied the discrete Morse theory, which is a 'hot issue' in the field of computational topology, to qualitative machine learning. We used it to induce a qualitative model from numerical data. Qualitative rules are used to describe the qualitative constraints of class variable using given attributes. We focused mainly on the theoretical issues yet showing how QING performs in practise. We are aware of the fact that QING's true power should be tested on real domains but still believe that all the theoretical background should be carefully considered first.

References

Barber, C. B.; Dobkin, D. P.; and Huhdanpaa, H. 1996. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software* 22(4):469–483.

Bratko, I., and Šuc, D. 2003. Learning qualitative models. *AI Magazine* 24(4):107–119.

Breiman, L.; Friedman, J.; Olshen, R.; and Stone, C. 1984. *Classification and Regression Trees*. CRC Press.

Forman, R. 2001. A user's guide to discrete morse theory. In *Proceedings of the 2001 International Conference on Formal Power Series and Algebraic Combinatorics, A special volume of Advances in Applied Mathematics.*

King, H. C.; Knudson, K.; and Mramor Kosta, N. 2005. Generating discrete morse functions from point data. *Exp. math.* 14(4):435–444.

Kuipers, B. 1994. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, Massachusetts.

Milnor, J. 1963. *Morse Theory*. Princeton University Press.

Quinlan, J. 1992. Learning with continuous classes. In *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, 343–348.

Šuc, D.; Vladušič, D.; and Bratko, I. 2004. Qualitatively faithful quantitative prediction. *Artificial Intelligence* 158(2):189–214.