How many cuttings to cut from a mother plant?
Combining Linear Programming and Data Mining

DÜMMEN ORANGE

Han Hoogeveen, Utrecht University

Joint work with Jakub Tomczyk and Tom van der Zanden

Dümmen Orange

- “New” company: merger of 67 companies.
- Largest producer of cuttings (stekken) in the world.
- Over 1.4 billion cuttings sold each year; market share 8%.
- Production sites all over the world.
- 4500 varieties; 57 crops.
- Presented a problem at ‘Mathematics with Industry’.

Problem description

- Mother plants are planted some months before harvesting; cuttings are cut from these mother plants weekly.
- The number of mother plants to plant is based on forecasts; you must plant enough to fulfil the forecasts for the demand.
- For each variety, the majority of sales takes place in the 'peak weeks', which is a period of approximately 10 weeks. We use T to denote the number of weeks here.
- For each week you have to decide how many cuttings to cut from the mother plants.
- The number of cuttings that can be cut from a mother plant depends (in some way) on the number of cuttings that have been cut in previous weeks.

Problem description

- Example: Can obtain (for instance) 2 cuttings per week per plant.
- Problem: can not consistently obtain 2 cuttings, goes down after a few weeks.
- Currently: plant $0.5\times$ peak demand $+$ 10% "buffer".
- Forecast: 80.000 plants/week in peak $\rightarrow$ plant 44.000 mother plants.
- Pressure from sales to offer more (from the buffer) for sale $\rightarrow$ may not have enough later.

Research questions

1. Model how the number of cuttings produced goes down as more cuttings are taken.
2. Determine how many mother plants should be planted to meet predicted demand.
3. Determine how many cuttings to offer for sale in each week (and thus how many to cut).

- Currently, we only look at one variety of plant in isolation and use a deterministic model (we ignore random disturbances).
- Guidance by domain expert; some data available
- Available data: 160 numbers (10 years, 16 weeks, average harvest per mother plant)

LP approach (How many mother plants at least)

- Since the current decisions depend on earlier decisions, we work with feasible *cutting patterns*.
- A cutting pattern tells you how many cuttings to cut on average from a mother plant in each week; this number can be fractional. For example $(2.0; 1.8; 1.9; 1.7; 2.0; \ldots)$.
- If the domain expert has a list with all N feasible cutting patterns that we can use, then ...

... we can solve the problem of finding the minimum number of mother plants by formulating it as a **Linear Programming** problem. Remark that we have to meet the forecasts.

LP formulation

- Define x_j as the number of mother plants that are cut according to cutting pattern j, for $j = 1, \ldots, N$.
- Define a_{jt} as the number of cuttings that are obtained in week t $(t = 1, \ldots, T)$, when one mother plant is cut according to cutting pattern j.
- Define b_t as the forecasted demand in week t.

LP formulation (2)

- LP formulation

$$\min M = x_1 + \ldots + x_N$$

subject to

$$\sum_{j=1}^{N} a_{jt} x_j \geq b_t \quad \forall t$$
$$x_j \geq 0 \quad \forall j$$

- The solution of this LP program gives you a lower bound to the number of mother plants that have to be planted. The company can decide to add more (for example to build in some safety margin).
- Unfortunately, the domain expert did not have a list of feasible cutting patterns available.

Next attempt

- This linear programming formulation looks a lot like the one to solve the **Cutting Stock** problem (find the minimum number of bars to cut a number of pieces with given length).
- In Cutting Stock we work with cutting patterns as well; they indicate how many pieces of each length to cut from a bar.
- Cutting Stock is solved using **column generation**: let's try that here.
- Major difference: for Cutting Stock it is clear when a cutting pattern is feasible. For the flower cutting problem the constraints to decide whether a cutting pattern is feasible are unknown (this was one of the research questions).

Find the unknown constraints

- Use the data that we have to find these constraints (and have these confirmed by the experts).
- We have data specifying the average number of cuttings taken in week t ($t = 1, \ldots, T$) per year; we assume that these are related to what is actually possible.
- Use **data mining** (or data search) to search for constraints that are satisfied by all feasible cutting patterns; to specify a cutting pattern we must specify the numbers $(a_1, \ldots, a_T)$.
- For example: infer constraints $a_t \leq C_1 \;\; \forall t$ by determining C_1 as the maximum over all a_t found in the data.
- Similarly, $a_t + a_{t+1} \leq C_2$: find C_2 as the maximum value of two consecutive weeks, etc.

Example constraints from data mining

$$a_t \leq 2$$
$$a_t + a_{t+1} \leq 3,9$$
$$a_t + a_{t+2} \leq 3,85$$
$$a_t + a_{t+3} \leq 3,9$$
$$a_t + a_{t+1} + a_{t+2} \leq 5,75$$
$$a_t + a_{t+1} + a_{t+4} \leq 5,71$$
$$a_t + a_{t+3} + a_{t+4} + a_{t+5} \leq 7,4$$
$$a_t + a_{t+3} + a_{t+5} + a_{t+6} \leq 7,41$$
$$a_t + a_{t+4} + a_{t+5} + a_{t+6} \leq 7,41$$
$$a_t + a_{t+5} + a_{t+6} + a_{t+7} \leq 7$$
$$a_t + a_{t+1} + a_{t+3} + a_{t+4} + a_{t+5} + a_{t+6} \leq 10,9$$
$$a_t + a_{t+6} + a_{t+7} + a_{t+8} + a_{t+9} + a_{t+10} \leq 10,5$$

Back to column generation

- We can formulate the pricing problem then as: find the best cuttern pattern $(a_1, \ldots, a_T)$ that satisfies the constraints determined using data mining.
- Since the constraints are linear, this is an LP problem (again).
- We can use Constraint Satisfaction to solve the pricing problem in case of non-linear complicated constraints.
- **Useful Observation**
  In case the feasibility of a cutting pattern is described using linear constraints only, then we do not need column generation to solve the problems!

Convex combination of patterns

- Suppose that the cutting patterns $\mathbf{a_1}, \ldots, \mathbf{a_N}$ satisfy the linear constraints, where $\mathbf{a_j} = (\mathbf{a_{j1}}, \ldots, \mathbf{a_{jT}})$.
- Take any set of non-negative real values $\lambda_1, \ldots, \lambda_N$ with $\sum_{j=1}^{N} \lambda_j = 1$.
- Define cutting pattern $\mathbf{C}$ as the convex combination of the cutting pattterns $\mathbf{a_1}, \ldots, \mathbf{a_N}$:

$$\mathbf{C} = \sum_{\mathbf{j=1}}^{\mathbf{N}} \lambda_{\mathbf{j}} \mathbf{a_j}$$

- Then $\mathbf{C}$ satisfies the linear constraints and is a feasible cutting pattern.

Computing number of mother plants (1)

- Define x_j as the number of mother plants that are cut according to cutting pattern j, for $j = 1, \ldots, N$.
- Define b_t as the forecasted demand in week t.
- LP formulation

$$\min M = x_1 + \ldots + x_N$$

subject to

$$\sum_{j=1}^{N} a_{jt} x_j \geq b_t \quad \forall t$$
$$x_j \geq 0 \quad \forall j$$

Theorem

Let $x^* = (x_1^*, \ldots, x_n^*)$ denote an optimal solution. Then there exists an equivalent solution in which we use only 1 cutting pattern $C = (C_1, \ldots, C_T)$

- Here: put $\lambda_j = x_j^*/M \implies MC_t = \sum_{j=1}^{N} a_{jt} x_j^*$.

Computing number of mother plants (2)

- Observation: If you want to produce b_t cuttings in week t using one cutting pattern only, then you should cut $a_t = b_t/M$ cuttings on average.
- Determine M such that $(a_1, a_2, \ldots, a_T)$, where $a_t = b_t/M$, corresponds to a feasible cutting pattern.
- Satisfying for example constraint $a_t + a_{t+1} \leq 3,9$ implies that $\frac{b_t}{M} + \frac{b_{t+1}}{M} \leq 3,9$.
- Working things out yields the constraint

$$M \geq \frac{b_t + b_{t+1}}{3,9}.$$

- Each constraint from data mining yields a lower bound on M; put M equal to the maximum of these bounds.

Optimize sales income

- The number of mother plants M is a decision variable; cost per mother plant is c_M.
- Again, x_j denotes the number of mother plants cut according to pattern j ($j = 1, \ldots, N$).
- We assume that for each week we know how many cuttings we can sell additionally; call this AD_t.
- Define decision variables y_t as additional sales realized in week t.
- Define p_t as profit of selling additional cuttings in week t (refinements are possible).
- The problem can be formulated as an LP again.

LP formulation

$$\max \sum_t p_t y_t - Mc_M$$

subject to

$$\sum_{j=1}^{N} a_{jt} x_j - y_t \geq b_t \quad \forall t$$
$$\sum_{j=1}^{N} x_j \leq M$$
$$0 \leq y_t \leq AD_t \quad \forall t$$
$$x_j \geq 0 \quad \forall j$$

- Again, in the optimal solution we need only one cutting pattern.

Reformulation of the LP

- Define z_t as the number of cuttings produced in week t.
- Since we only need one cutting pattern, $a_t = z_t/M$. This yields the LP:

$$\max \sum_t p_t(z_t - b_t) - Mc_M$$

$$\text{subject to}$$

$$b_t \leq z_t \leq b_t + AD_t \quad \forall t$$

$$\text{`the variables } a_t = z_t/M \text{ form a feasible cutting pattern'}$$

- z_t/M is not linear; rewrite the constraints.
- For example: $a_t + a_{t+1} \leq 3,9$; multiply with M.
- This yields the constraint $z_t + z_{t+1} \leq 3,9M \implies$ LP-again.

Extensions

- More varieties of plants: just extend the LP.
- Uncertainty in growth of mother plants: work with scenarios (use information from historical data).

Conclusions

- Having just a few data may make a huge difference.
- Using data mining to describe an unknown part of the model is a neat trick (which we have not seen being used before).
- If linear constraints suffice, then sometimes you can simplify the model a lot.