

Point-feature lettering of high cartographic quality: A multi-criteria model with practical implementation

Andreas Reimer*

Maxim Rylov†

Abstract

There have been numerous and varied research efforts to automate point-feature label placement (PFLP). It seems that many well-established precepts for point-feature annotation used by human cartographers have been neglected so far. As a consequence, the currently implemented, fully automated solutions are limiting computer generated maps in their expressive power. In this paper we present a comprehensive multi-criteria model that complies with almost all well defined cartographic placement principles and requirements for PFLP. This allows for a significant increase in toponym density without effecting legibility. The proposed model expressed as a quality evaluation function can be employed by any mathematical optimization algorithm for solving the automated label placement problem. An application of the proposed model was tested on Volunteered Geographic Information data and sample parameter settings were devised. The results illustrate that a high level of cartographic quality for PFLP can be achieved through the integrated approach. The resulting quality is comparable to the lettering produced by an expert cartographer.

1 Introduction

Over the last four decades many attempts have been made to automate the process of map labelling; see bibliography of papers on this topic maintained by [1]. The main reason is to reduce the cost of manual label placement. The cartographic lettering problem comprises (e.g. [3]):

- toponym considerations (e.g. exonyms/ endonyms, multilingual labelling and gazetteers);
- toponym selection in respect to task;
- typeface (e.g. font choice, font form (spacing, colouring, italics etc.), font size, font color);
- geometric placement of labels.

Automated label placement research has concentrated on the geometric placement and assumes the preceding problems have been solved, i.e. are input parameters. A general function to measure the geometric quality of label placement was proposed by [2], but

*Department of Geography, University of Heidelberg, andreas.reimer@uni-heidelberg.de

†Department of Geography, University of Heidelberg, maxim.rylov@geog.uni-heidelberg.de

Research Article	PFLP Requirements and Constraints Applied in Research Article											
	R1	R2	R3	R4	R5	R6	R7	R8	R9	C1	C2	C3
Yoeli (1972)	X	X	X							X	X	X
Hirsch (1982)	X	X	X						X	X	X	X
Zoraster (1986)	X	X	X							X	X	X
Doerschler and Freeman (1992)	X	X	X						X	X	X	X
Edmondson and others (1997)	X	X	X						X	X	X	X
Strijk and van Kreveld (1999)	X	X							X	X	X	X
Huffman and Cromley (2002)	X	X	X							X	X	X
Ebner, Klau, and Weiskircher (2003)	X	X				X	X					
van Dijk, Thierens, and de Berg (2004)	X	X								X	X	X
Stadler, Steiner, and Beiglbock (2006)	X	X				X				X	X	X
Mote (2007)	X	X	X							X	X	X
Luboschik, Schumann, and Cords (2008)	X	X	X						X	X	X	X
Bae and others (2011)	X	X										
Gomes, Ribeiro, and Lorena (2013)	X	X				X	X					

Figure 1: The PFLP requirements and constraints in previous research works.

neither implemented nor experimentally evaluated to the crucial level of the partial functions. We introduce a comprehensive multi criteria model that covers more cartographic placement rules than the related work and show the practical applicability by implementation and experiment.

2 Multi-criteria model

The extant guidelines [4, 5, 6] that refer to labelling of point features can be divided into two categories: rules and constraints. The list of rules adapted to the requirements of our purposes is as follows:

- R1 Type arrangement should reflect the classification, importance and hierarchy of objects.
- R2 Labels should be placed horizontally.
- R3 The lettering to the right and slightly above the symbol is prioritized.
- R4 Names of coastal settlements should be written in water.
- R5 Labels should be placed completely on the land or completely on the water surface.
- R6 Names should not be too close to each other.
- R7 Labels should not be excessively clustered nor evenly spread out.
- R8 Each label should be easily identified with its point feature. Ambiguous relationships between symbols and their names must be avoided.
- R9 Labels should not overlap other significant features of the cartographic background or do this as little as possible.
- C1 Names must not overlap point feature symbols.
- C2 Two names must not overlap each other.
- C3 Two point features must not have any overlap.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	C1	C2	C3
	X	X	X	X	X	X	X	X	X	X	X	X

Figure 2: The PFLP requirements and constraints in the presented model.

It is important to note, that R6 and especially R8 are in fact the most important of all rules according to the literature, as the whole usability of the map hinges upon them. Previous work often concentrates on avoiding overlapping (C1-C3) labels with few regards to disambiguation (Fig. 1 and 2).

Quality evaluation function

We have written out the requirements for good label placement for our use case. These requirements are also metrics for the quality evaluation function. The requirement that a point feature can be labelled only once or even can be left unlabelled can be written as:

$$\sum_{i=1}^{P_n} x_{i,j} \leq 1, \forall j = 1 \dots N \quad (1)$$

where $x_{i,j} \in 0, 1$ is the decision variable that defines whether the j th point feature is labelled in the i th position and N is the number of map features to be labelled and P_n is the number of possible label positions for one feature. In our implementation $P_n = 8$, being a common model in PFLP. An important constraint of map lettering is that no two labels may in any way overlap each other or point features. This is expressed with the inequality: $x_{i,j} + x_{k,m} \leq 1$, which is valid for all intersections between the i th position of the j th point feature and the k th position of the m th point feature.

The full form of the quality evaluation function of cartographic preferences is:

$$Q(x) = \sum_{i=1}^{P_n} \sum_{j=1}^N (\beta_1 F_{i,j}^{prior} + \beta_2 F_{i,j}^{pos} + \beta_3 F_{i,j}^{over} + \beta_4 F_{i,j}^{disamb} + \beta_5 F_{i,j}^{clut} + \beta_6 F_{i,j}^{coast}) x_{i,j} / N_v(x) \quad (2)$$

where $N_v(x)$ is the number of visible labels from input N , $\beta_1 \dots \beta_6$ are the weights for the corresponding metric $F_{i,j}^*$ that are explained below.

Priority of the point feature: The difference in presentation of two cities helps a reader to see the difference in population, importance, or administrative status of a place [7]. Such differentiation can be done by assigning a priority to a place, often this is done by population. After normalization over the minimum and maximum of the extant population values, the metric is in range $[0, 1]$, fulfilling requirement R1.

Positioning of the name around its point feature in terms of cartographic desirability. The metric has the maximum value $[1]$ when the label position is

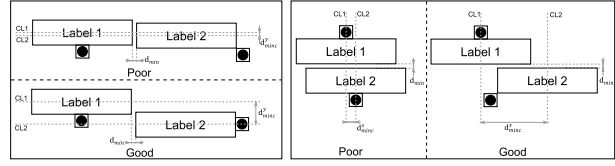


Figure 3: Presentation of ambiguity between two neighbouring labels.

somewhat above and to the right of its symbol, other positions are ranked lower after an ordered list such as found in [4]. Fulfilling requirement R2 and R3.

Overlap of symbol and its label with other significant map features. We use a raster-based method. We prefer this method as it can be used on both small and large scales. We define a metric that can measure homogeneity of the cartographic background under a label. As an input for this metric we require a raster image I in which the nontextual objects are already rendered. Each element $p \in I$ is a pixel. Assume that we applied some image segmentation algorithm φ which transforms pixels into clusters $\varphi(I) = c^1, c^2, \dots, c^M$, where $c^m, m = 1$, are the clusters. This metric is designed to yield a value of 1.0 for the case when all elements within the bounding rectangle $R(i, j)$ of a label belong to one cluster, i.e. the region of the map background under a label is homogeneous. R9.

Disambiguation: The magnitude of ambiguity between neighbouring point features and their names; measured as normalized distances between labels themselves and between their respective centrelines. The distances are compared to certain thresholds beyond which no ambiguity can occur anymore at the given resolution. Illustrated in Fig 3; R6 and R8.

Clutter magnitude of ambiguity between and density of neighbouring point features and their names. Calculated with a modified version of the force-based model of [11]. R6 and R7.

Coastal places are measured by the percentage of water under the label for the point features that describe coastal places. The actual procedure uses an approach similar to the abovementioned raster-based technique with a separate colour mask generated from ocean polygons. R4 and R5.

3 Experimental results

In our implementation we used a couple of techniques that allow an increase in performance of label placement algorithms. Firstly, we made use of a quad tree data structure to store labels and examine whether any label overlaps other characters or symbols on the map. Secondly, in a pre-processing step, we construct a conflict graph, whose nodes are all labels of the map, and whose edges indicate potential overlap with other labels (nodes). The graph-based approach is

one of the most common approaches in the field of interactive and dynamic labelling [8, 9]. In order to compare effectiveness and accuracy of our model we used three well-known heuristic search algorithms for solving PFLP as a mathematical programming problem: greedy, discrete gradient descent and simulated annealing. In our experiments these algorithms are used to find a feasible solution for label placement by treating the proposed model as the objective function to be optimized. In our implementation the greedy algorithm is restricted to the selection of the first candidate position for a point feature that can be placed on a map, i.e. the improvement of a final solution within the candidate positions of a point feature is not allowed. As an annealing schedule for the simulated annealing algorithm we chose a polynomial-time cooling schedule that was proposed by [10].

We performed our experiments on a dataset that represents geospatial data granted by the OpenStreetMap project that is one of the most promising crowd sourced projects. The test dataset represented the northern part of Denmark. From the dataset we extracted all settlements and divided them into 4 groups. To each group we assigned a different font size and image for point features which reflect the population and administrative status of a place.

In Figures 5 and 6 we present the results of Test 1 for the greedy and gradient descent algorithms respectively. In this test we used only two metrics common in the literature such as positioning around a point-feature and feature hierarchy. From these figures it is clear that the resulting map is far from a high level of functionality of the lettering, as the names partially hide some important and relevant geographic features such as roads (see Fig. 5, ref. 2) and bays (see Fig. 6, ref. 1). Moreover, we can see some distinct ambiguities between the names and the features they label (see ref. 1 in Fig. 5, 6).

A demonstrative example is a group of villages that consist of Møldrup, Roum, Bjerregrav, Skals. Figure 8 shows the changing label configuration with simulated annealing using the weights for Test Nr. 2 [3 metrics] (a) Nr. 3 [4 metrics] (b), Nr. 4 [3 metrics] (c) and Nr. 5 [6 metrics] (d).

It is evident from the tabular data (see Figure 4) that the simulated annealing algorithm predominated over other algorithms in the quality of the label assignment, while the greedy heuristic outperforms

№	N	β_1	β_2	β_3	β_4	β_5	β_6	Greedy				Gradient Descent				Simulated Annealing			
								$N_c(x)$	%	$Q(x)$	$t_{pm}[s]$	$N_c(x)$	%	$Q(x)$	$t_{pm}[s]$	$N_c(x)$	%	$Q(x)$	$t_{pm}[s]$
1	82	0.6	0.4	0	0	0	0	59	71.95	0.617	0.01908	66	80.49	0.657	0.02553	73	89.02	0.706	1.02975
2	82	0.3	0.2	0.5	0	0	0	59	71.95	0.673	0.08279	66	80.49	0.717	0.09524	73	89.02	0.784	1.29695
3	82	0.2	0.1	0.3	0	0	0.4	59	71.95	0.712	0.39249	66	80.49	0.761	0.42134	74	90.24	0.839	1.48764
4	82	0.2	0.1	0	0	0	0.7	59	71.95	0.729	0.34698	66	80.49	0.779	0.35373	74	90.24	0.858	1.57519
5	82	0.2	0.1	0.3	0.1	0.05	0.25	59	71.95	0.710	0.50949	66	80.49	0.760	0.51101	74	90.24	0.834	3.56920

Figure 4: The results of the PFLP algorithms with different parameter weights.



Figure 5: The results of the greedy test run No 1.

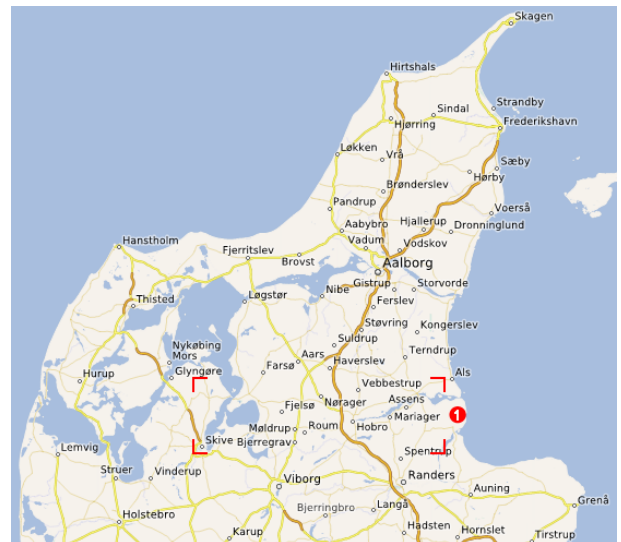


Figure 6: The results of the gradient descent test run No 1.

other algorithms by returning always the worst solution of PFLP. It is also clear that computational resources required for each PFLP algorithm and set of metrics vary greatly. In order to determine how strong the impact of different metrics is upon performance, we calculated a score as the running time of Test 5 divided by the running time of Test 1. For the tested PFLP algorithms this score is 26.70, 20.01 and 3.46 respectively. Comparing the scores, we can conclude that a substantial amount of computation time is spent on pre-processing for different metrics.

4 Conclusion

Our model of a quality evaluation function for the PFLP problem satisfies almost all cartographic re-

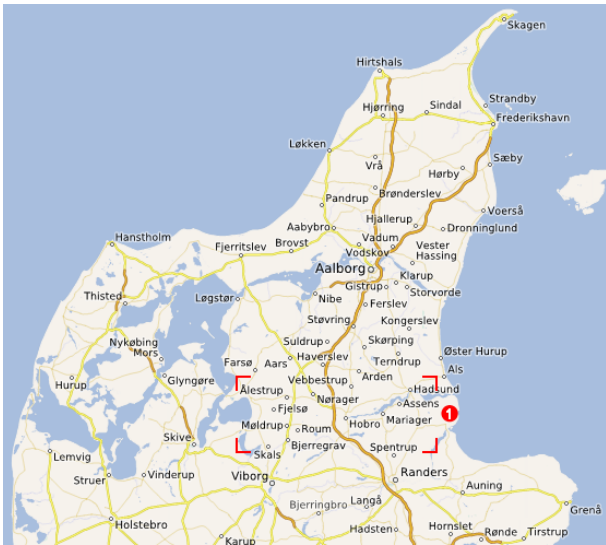


Figure 7: The results of the simulated annealing test run No 5.

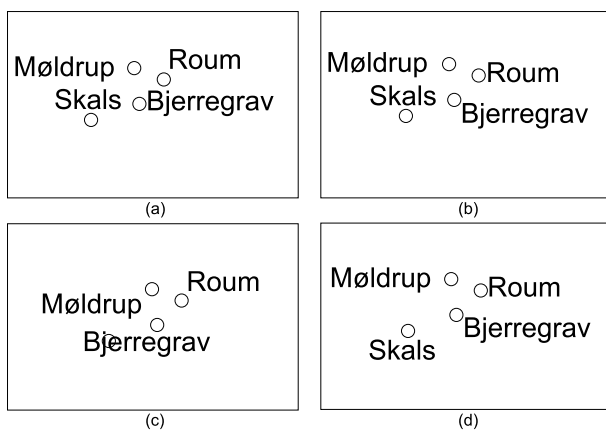


Figure 8: A sketch representation of ambiguity for a group of villages.

quirements for point features more than any other previous model given in the literature. Producing unambiguously labelled maps has been traditionally recognized as being the most important aim of the whole labelling process [4, 5, 6]. To our knowledge this is the first attempt to address, among other things, the quantification of ambiguous label-feature relations. The proposed model is highly adjustable and provides the human cartographer a handy tool to make an appropriate label placement. It also conceptually opens the possibility to automate the preceding stages of label placement (see Introduction), which have previously been neglected in research. The experiments argue that the model together with an appropriate mathematical optimization algorithm for PFLP, which is able to find an approximation to the global optimum, produces visually plausible lettering with high cartographic quality and is capable of

considerably enhancing the functionality of the map. The presented model can also be used for labelling other feature types (lines, areas). We implemented the model on top of the platform MapSurfer.NET for publishing spatial data to the web. We prepared and published a web map that is based on the global dataset provided by the OpenStreetMap project. It is available online on the OpenMapSurfer web page (OSM Roads layer) [12]. The name placement on small scales (zoom levels 2 – 12) is done using the multi-criteria model devised in this paper.

Acknowledgments

Maxim Rylov gratefully acknowledges DAAD (Deutscher Akademischer Austausch Dienst, German Academic Exchange Service) for awarding the DAAD-Scholarship to pursue his doctoral degree.

References

- [1] Wolff, A., and T. Strijk *A map labeling bibliography*. <http://i11www.iti.uni-karlsruhe.de/~awolff/map-labeling/bibliography/> 2009.
- [2] van Dijk, S., M. van Kreveld, T. Strijk, and A. Wolff *Towards an Evaluation of Quality for Names Placement Methods*. *International Journal of Geographical Information Science* 16/7: 641-661.
- [3] Imhof, E. *Thematische Kartographie*. Berlin, Walter de Gruyter 1972
- [4] Imhof, E. *Positioning Names on Maps* *The American Cartographer* 2/2: 128-144. 1975
- [5] Wood, C.H. *Descriptive and Illustrated Guide for Type Placement on Small Scale Maps*. *The Cartographic Journal* 37/1: 5-18. 2000
- [6] Brewer, C.A. *Designing Better Maps: A Guide for GIS Users*. ESRI Press, Redlands CA. 2005
- [7] Butzler, S.J., C.A. Brewer, and W.J. Stroh. *Establishing Classification and Hierarchy in Populated Place Labeling for Multiscale Mapping for The National Map*. *Cartography and Geographic Information Science* 38/2: 100-109. 2011
- [8] Been, K., E. Daiches, and C. Yap. *Dynamic Map Labeling*. *IEEE Transactions on Visualization and Computer Graphics* 12/5, 2006
- [9] Mote, K. *Fast Point-Feature Label Placement for Dynamic Visualizations*. *Information Visualization* 6/4: 249-260, 2007
- [10] Aarts, E.H.L., and P.J.M. van Laarhoven. *A New Polynomial Time Cooling Schedule*. *Proc. IEEE International Conf. on Computer Aided Design*, Santa Clara, 206-228, 1985
- [11] Ebner, D., G.W. Klau, and R. Weiskircher. *Force-Based Label Number Maximization*. Technical Report TR-186-1-03-02, Vienna U of Technology, 2003.
- [12] OpenMapSurfer *A web mapping service provided by University of Heidelberg*. Available at <http://openmapsurfer.uni-hd.de>, 2014.