Individualized Hiking Time Estimation

Arthur Pitman and Markus Zanker
Alpen-Adria-Universität Klagenfurt
9020 Klagenfurt, Austria
{arthur.pitman, markus.zanker}@aau.at

Johann Gamper and Periklis Andritsos
Freie Universität Bolzano/Bozen
39100 Bozen-Bolzano, Italy
{johann.gamper, periklis.andritsos}@unibz.it

***Abstract*—Route planning algorithms attempt to find the optimum path between two nodes in a graph, where a cost function specifies a weight for each edge. In many situations cost is related to distance or time. While time estimates are rather straightforward for automotive applications, real-world route planning for hiking and other outdoor activities requires careful consideration of a variety of different factors in order to produce reliable time estimates. Static estimates such as Naismith's Rule for estimating hike times do not consider individual factors such as a hiker's fitness or current progress along the trail. In this paper we address these aspects and develop a model for individual weight estimation that can be exploited in route planning applications. An evaluation conducted on GPS traces from hikes in South Tyrol, Italy indicates that the model can outperform Naismith's estimate by up to 23%.**

***Index Terms*—route planning; time estimation; hiking; tourism**

I. Introduction

Calculating time requirements has been an important part of journey planning since at least Roman times. Vegetius, for instance, wrote in his *De Re Militari* that the Roman Army should march "with the common military step twenty miles in five summer-hours" [1]. In this paper, we focus on the more peaceful pursuit of recreational hiking, where such estimates form an important planning element, allowing walkers to determine what is feasible. Furthermore, in the case of unexpected bad weather or emergencies, realistic time estimates may help walker reach suitable shelter or aid search and rescue operations.

In the tourism domain, mobile travel aids are becoming an increasingly valuable tool for decision making 'on the go'. Here, accurate time estimates are of particular importance, allowing hiking routes to be synchronized with the schedules of public transportation facilities or the opening hours of visitor centers. In the following section, we will examine related work in the area of itinerary and route planning. In Section III we describe the modeling approach and the baseline estimate based on Naismith's Rule. Finally, Section IV describes our evaluation methodology and presents the results of applying the empirical model to historical hiking data.

II. Related work

Itinerary or route planning has attracted the attention of the research community, especially since the development and wide-spread use of GPS devices. Many well-known software companies offer web-based and mobile solutions that assist users in planning and following a route. Most of these tools, however, are developed around the assumptions of *constancy* and *universality*, two notions dictating that route planning is independent of environmental parameters and user preferences [2].

Vansteenwegen and Souffriau [3] present a deep overview of personalized electronic tourist guide systems. They model the problem as an Orienteering Problem [4] and extend it to include time-dependent information related to public transportation systems. Hence, every itinerary produced is different depending on the transportation system used and the departure time. Apart from the aforementioned, a number of objectives/criteria can be combined in order to improve the planning process for users. The study presented in [5] offers a decision-theoretic process based on spatial ontologies, which models quantitative and qualitative data in a hierarchical way and aims to provide a user-centric route plan.

Letchner et al. [2] offer a solution that is based both on time-variant parameters, such as traffic speeds and patterns, as well as previous GPS traces logged by the system. To arrive at a personalized route, their TRIP system employs Hidden Markov Models (HMM) that, given prior observations and the current position, scores different locations and chooses the one most likely to satisfy all constraints. Using travel-related preferences, Tumas and Ricci [6] employ a knowledge-based recommendation technology and a scoring function that takes user input and point of interest characteristics and determines the most relevant points to be included in a user's itinerary.

With the advent of social media, it is increasingly easy to reach users and, at the same time, share information and produce content. For example, media sharing sites such as Flickr[1] act as a medium for storing pictures that are associated with a timestamp, geographic information and user-provided descriptions and tags, making such photo albums a rich source of tourist information. Work on automatic mining and information extraction from such repositories aims to develop recommendation algorithms to enhance a tourist's experience. McGinty and Smyth [7] propose a route planning algorithm using Case-Based Reasoning (CBR) that produces new individualized routes based on prior experience of users with similar preferences and environmental characteristics. In the most recent work in the area of personalized route

[1] http://www.flickr.com

1529-4188/12 $26.00 © 2012 IEEE
DOI 10.1109/DEXA.2012.51

planning, Luccese et al. [8] offer a graph-based algorithm for the discovery of individual routes. As input, the algorithm takes a weighted graph of points-of-interest that are connected if the probability (based on past experience) that they belong to the same itinerary is high. They then employ a Random Walk with Restarts algorithm [9] to rank the candidate points-of-interest and come up with the best route for a user.

III. Models for time estimation

In 1892, Scottish mountaineer William W. Naismith proposed a basic rule for calculating hike times: "men in fair condition should allow for easy expeditions, namely, an hour for every three miles on the map with an additional hour for every 2000ft of ascent" [10]. While this rule forms the basis for most estimates, many adjustments have been made over the years that have followed. For instance, in 1965 Philip Tranter introduced further corrections to compensate for hiker fitness, as determined by the time required to climb 300 meters over a distance of 800 meters. Other common corrections include subtracting 30 minutes for each kilometer of gentle descent ($5° - 12°$) and adding 30 minutes for each kilometer of steep descent ($> 12°$) [11]. Additional variations may be used to account for weather, terrain and pack load. In addition to these factors, we suppose that the length of hike, motivation and fatigue may influence hiker speed. For example, longer hikes may require more gear and therefore be slower. Alternatively, hikers may speed up due to increased enthusiasm as they approach their goal.

Several works have investigated the validity of Naismith's Rule. Scarf examined its applicability to both fells and treadmill running, finding an 8:1 ratio between flat distance and elevation for fells and an 1:3 ratio for treadmills [12], [13]. Norman confirmed these results, adding an additional ratio of 6:1 for roads based on the results of some very specific events and concluding that while the rule was generally applicable, its parameters were dependent on the terrain [14].

Despite this, Naismith's rule is ambiguous with respect to the treatment of downhill sections: [15] explains that the correction for ascent may be considered to account for descent as well or that it implies that descent is possible at the same speed as that on level ground. [16] provides a survey of many alternative models and advocates polynomial models as they can be easily fitted to both ascent and descent data.

A significant shortcoming of all of these models is that they are unable to react to the particular conditions of a hike or to the current abilities of the hikers. Despite their simplicity, the models are designed to be applied before a hike to create a static estimate, often based on traditional mapping materials such as topographic maps. The wide availability of GPS devices, integrated for example into mobile phones, means that, in addition to standard mapping data, detailed information is available about a hiker's current position (p) and speed/velocity (v). We propose to exploit this data in two steps:

1) Modeling: To create a model for the speed of hikers under various conditions.
2) Personalization: To exploit current hike data to adapt estimates to the real-time performance of the hiker.

While active, a GPS receiver periodically produces data about its current position. Each measurement typically includes information about the longitude, latitude and altitude, as well as a timestamp. Over time, these may be accumulated as a sequence of positions $P = \langle p_0, p_1, ..., p_n \rangle$ and a sequence of timestamps $T = \langle t_0, t_1, ..., t_n \rangle$, collectively known as a *trace*. A speed v_i may be calculated for each segment spanning from $i-1$ to i:

$$v_i = \frac{dist(p_{i-1}, p_i)}{t_i - t_{i-1}} \tag{1}$$

Furthermore, a feature vector $\vec{z_i}$ of length m may be extracted from a given P at point i by a function g, i.e.:

$$\vec{z_i} = g(P, i) \tag{2}$$

The feature vector could, for example, include information about the segment, such as its gradient, contextual information about the segment's position in the trace, such as the proportion of elapsed distance, and information about the trace as a whole, such as its length. Hiker speed may be modeled by a function of $\vec{z_i}$ for each segment:

$$\hat{v_i} = f(\vec{z_i}) = \sum_{j=1}^{m} \alpha_j z_{ij} \tag{3}$$

where $A = (\alpha_1, \alpha_2, ..., \alpha_m)$ is a vector of coefficients. The modeling task is thus reduced to linear optimization: given a set of n speeds V and feature vectors Z, find $\hat{A}$ such that the cost function C is minimized:

$$C_{min} = \min_A \sum_{i=0}^{m} (v_i - \hat{v_i})^2 = \min_A \sum_{i=0}^{m} \left(v_i - \sum_{j=1}^{m} \alpha_j z_{ij} \right)^2 \tag{4}$$

i.e.:

$$\hat{A} = \arg\min_A \sum_{i=0}^{m} \left(v_i - \sum_{j=1}^{m} \alpha_j z_{ij} \right)^2 \tag{5}$$

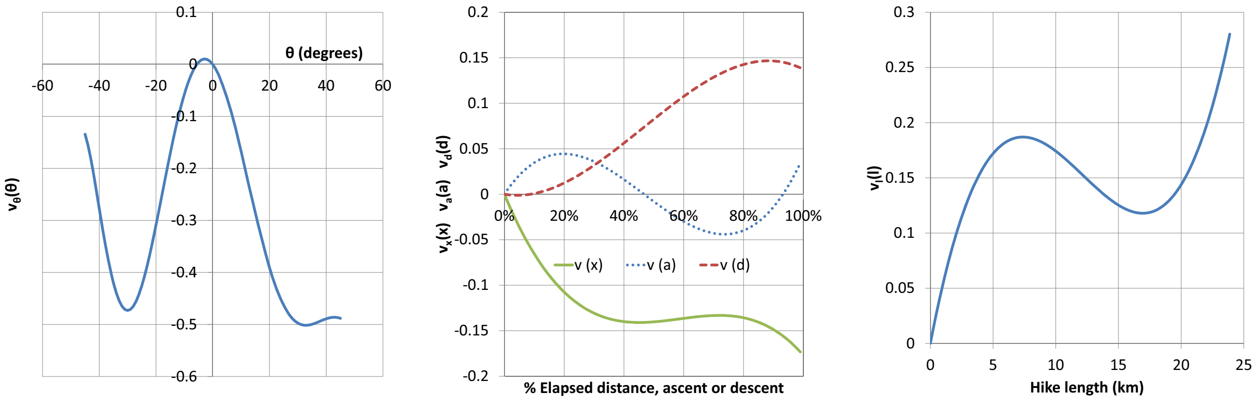
where $v_i \in V$, $z_i \in Z$. The model may then be applied to estimate the time $\hat{r}$ required for an unknown hiker to walk a known hike specified by P:

$$\hat{r} = \sum_{i=1}^{n} \frac{dist(p_{i-1}, p_i)}{\hat{v_i}} = \sum_{i=1}^{n} \frac{dist(p_{i-1}, p_i)}{A \cdot g(P, i)} \tag{6}$$

IV. Evaluation

A dataset of 360 hikes was constructed from GPS traces logged in South Tyrol, Italy. The GPS data itself was obtained from a leading publicly available repository using a filter restricting the traces to hiking (*activity type = "hiking"*) logged between March 2011 and March 2012. In general, each trace originates from a different route and hiker.

Fig. 1. Plots of components $v_\theta, v_x, v_a, v_d, v_l$



(a) v_θ (b) v_x, v_a, v_d (c) v_l

Furthermore, the data was preprocessed and cleaned to mitigate any bias introduced by the range of logging devices. In particular, as logging frequencies can vary substantially, from a second to several minutes, the longitude and latitude data was resampled to produce points at five minute intervals using piecewise cubic splines. This has the additional benefits of minimizing variations in measurement accuracy and reducing the dataset to a manageable size. Furthermore, as loggers may or may not contain specific hardware for measuring elevation, e.g. using barometric pressure, the recorded altitude values were replaced with data from a single elevation web service.

Following pre-processing, each trace was checked for consistency ensuring that it did not contain any velocities in excess of 4 ms^{-1} which, for example, might indicate that a logger was left active during car travel. Furthermore, any sections of a hike with an average speed of less than 0.2 ms^{-1} were explicitly marked as pauses and removed from the traces to prevent them from having any bearing on the time estimates.

For the analysis, the 360 hikes were randomly split into two equally sized sets, namely a learning set and a testing set. An overview of the dataset is given in Table I. The anonymized dataset is available for download from http://www.isbi.at/media/OSTAR-SmoothedTracesForTimeEstimation.zip.

$$\vec{z} = b\,(1, \theta, \theta^2, \theta^3, \theta^4, \theta^5, \theta^6, \theta^7, x, x^2, x^3, a, a^2, a^3, d, d^2, d^3, l, l^2, l^3) \tag{7}$$

For the evaluation, the function g was specified to create feature vectors containing polynomials of each segment's gradient angle θ (in radians), proportion of elapsed distance x, ascent a and descent d (each standardized to $[0, 1]$), as well as track length l (kilometers). With the exception of θ, which was included up to degree seven, cubic polynomials were used as higher degree polynomials produced no improvement. Furthermore, as some walkers are faster the others due to fitness or the difficulty of the hike, we introduced a scaling factor b determined by the average speed of the associated hiker on flat ground (calculated using sections with an angle of gradient between $-5°$ and $5°$). Thus each feature vector had the form:

TABLE I
DATASET PROPERTIES

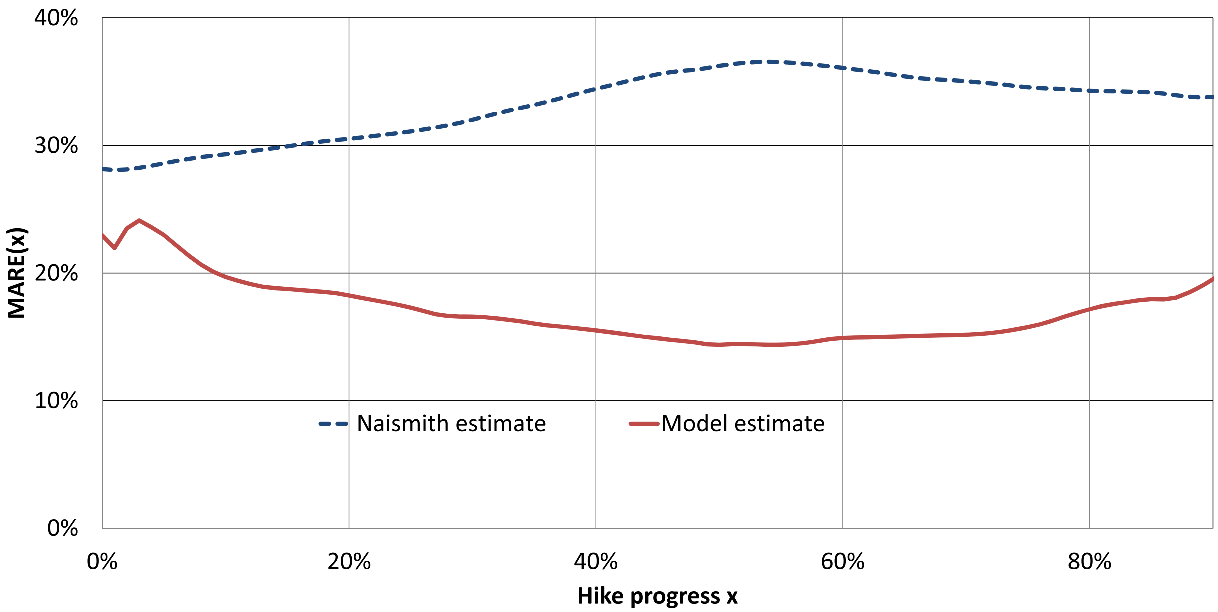
Fix Count				
Average	0 – 25	26 – 50	51 – 75	> 75
54.5	85	88	109	78
Distance (km)				
Average	0 – 5	5 – 10	10 – 15	> 15
7.59	111	160	71	18
Active Time (hours)				
Average	0 – 1	1 – 2	2 – 4	> 4
3.03	36	85	141	98
Pause Time (hours)				
Average	0 – 1	1 – 2	2 – 3	> 3
1.43	186	97	48	29
Average speed (km/h)				
Average	0 – 2	2 – 3	3 – 4	> 4
2.59	69	209	68	14
Total Ascent (m)				
Average	0 – 500	500 – 1000	1000 – 1500	> 1500
564	172	144	38	6
Total Descent (m)				
Average	0 – 500	500 – 1000	1000 – 1500	> 1500
543	178	133	46	3

Importantly, b may be estimated once a hike is in progress by comparing an $\hat{r}$ for the elapsed distance with the associated measured r:

$$\hat{b} = \frac{r}{\hat{r}} \tag{8}$$

As each hike contains a different number of points, points

Fig. 2. Model performance compared to Naismith's estimate



were randomly duplicated such that each hike contributes the same number of feature vectors and thus has the same bearing on the model.

The vector A was found using a batch gradient descent optimizer on feature vectors extracted from the learning set. The algorithm conducted 20 random restarts which produced only minimal variations in the coefficients. The resulting model is summarized below in Equation 9 and plots of the individual component functions v_θ, v_x, v_a, v_d and v_l are presented in Figure 1. Of particular interest is v_θ, which closely corresponds to functions proposed in the literature (see [16] for example). It can be observed that a slight descent increases speed, while steeper ascents or descents decrease speed significantly. In general, speed increases with hike length, presumably due to the fact that only experienced hikers attempt longer hikes. An initial value of $b = 0.82$, for use before a personalized $\hat{b}$ is available, was also determined from the learning set.

$$\begin{aligned} v(\theta, x, a, d, l) = &+ 0.87 \\ &- 0.40\theta - 4.02\theta^2 + 3.41\theta^3 + 10.24\theta^4 \\ &- 9.36\theta^5 - 7.33\theta^6 + 6.94\theta^7 \\ &- 0.86x + 1.63x^2 - 0.96x^3 \\ &+ 0.71a - 0.83a^2 + 0.27a^3 \\ &+ 0.53d - 1.65d^2 + 1.17d^3 \\ &- 0.009l + 0.068l^2 - 0.048l^3 \end{aligned} \tag{9}$$

Following the modeling phase, we then applied the model to the withheld testing set and compared its performance to the well-known Naismith estimate as a baseline. Both were required to produce an estimate for the remaining time at 1% intervals throughout each of the hikes and compared using mean absolute relative error (MARE) (Equation 10). Specifically, the metric calculates relative errors as each trail is of varying length and requires a different amount of time. Furthermore, these are considered in absolute terms to penalize over and under estimation equally and averaged over all q trails:

$$MARE(x) = \frac{1}{q} \sum_{k=1}^{q} \left| \frac{\hat{r_{kx}} - r_{kx}}{r_{kx}} \right| \tag{10}$$

The results of the evaluation are presented in Figure 2. The model clearly outperforms the Naismith estimate and is significantly better (by as much as 23%) during the crucial region of 20% to 80%. The problem becomes understandably more difficult during the final stages of a hike — small variations have a large influence on the accuracy of the estimate — however we assume that once a hiker has already completed more than 80% of a trail it is unlikely that the few remaining rerouting opportunities would require accurate time estimates. The slight increase during the first 5% is caused by initial instability in $\hat{b}$, the individualized scaling factor for each hiker and hike, due to a lack of available progress information.

V. Conclusion

In this paper we have constructed an empirical model for estimating the speed of hikers based on the gradient of ascent or descent, the current progress along the route as well as the overall length of the hike. Importantly, an evaluation of the model on historical data demonstrates that its time estimates are significantly more accurate than those produced by the well-known Naismith estimate. Furthermore, we believe that the improvement would be even more pronounced if tested on a broader sample of hikers, as GPS traces harvested from the web are representative of the hiker population as a whole, but rather are biased towards technology savvy and younger hikers. In future work we will extend our model to other types of outdoor activities associated with tourism, such as mountain biking and trail running. In addition, also hope to improve estimates by considering other variables, such as terrain information and weather estimates, and to carry out further evaluations using a larger and more varied dataset. The estimators will also be tested as part of a mobile application that will support hikers by offering personalized recommendations and individualized route planning.

ACKNOWLEDGMENT

The authors wish to acknowledge the financial support of the European Union (EU), the European Regional Development Fund (ERDF), the Austrian Federal Government and the State of Carinthia in the Interreg IV Italien-Österreich programme (project acronym O-STAR).

REFERENCES

[1] F. V. Renatus, *De Re Militari Book I: The Selection and Training of New Leviess*, 390, english translation by John Clarke published in 1767.

[2] J. Letchner, J. Krumm, and E. Horvitz, "Trip router with individualized preferences (trip): incorporating personalization into route planning," in *Proceedings of the 18th Conference on Innovative Applications of Artificial Intelligence*, vol. 2. AAAI Press, 2006, pp. 1795–1800.

[3] A. Garcia, O. Arbelaitz, M. T. Linaza, P. Vansteenwegen, and W. Souffriau, "Personalized tourist route generation," in *Proceedings of the 10th International Conference on Current Trends in Web Engineering*. Springer-Verlag, 2010, pp. 486–497.

[4] P. Vansteenwegen, W. Souffriau, and D. Van Oudheusden, "The orienteering problem : a survey," *European Journal of Operational Research*, vol. 209, no. 1, pp. 1–10, 2011.

[5] A. S. Niaraki and K. Kim, "Ontology based personalized route planning system using a multi-criteria decision making approach," *Expert Systems with Applications: An International Journal*, vol. 36, no. 2, pp. 2250–2259, 2009.

[6] G. Tumas and F. Ricci, "Personalized mobile city transport advisory system," in *Information and Communication Technologies in Tourism 2009*, W. Hpken, U. Gretzel, and R. Law, Eds. Springer Vienna, 2009, pp. 173–183.

[7] L. McGinty and B. Smyth, "Shared experiences in personalized route planning," in *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*. AAAI Press, 2002, pp. 111–115.

[8] C. Lucchese, R. Perego, F. Silvestri, H. Vahabi, and R. Venturini, "How random walks can help tourism," in *ECIR*, ser. Lecture Notes in Computer Science, R. A. Baeza-Yates, A. P. de Vries, H. Zaragoza, B. B. Cambazoglu, V. Murdock, R. Lempel, and F. Silvestri, Eds., vol. 7224. Springer, 2012, pp. 195–206.

[9] H. Tong, C. Faloutsos, and J.-Y. Pan, "Fast random walk with restart and its applications," in *ICDM '06: Proceedings of the Sixth International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 613–622.

[10] W. Naismith, "Notes and queries," *Scottish Mountaineering Club Journal*, vol. 2, p. 133, 1892.

[11] E. Langmuir, *Mountaincraft and Leadership*, 3rd ed. Sportscotland, 1995.

[12] P. Scarf, "An empirical basis for naismith's rule," *Mathematics Today*, 1998.

[13] ——, "Route choice in mountain navigation, naismith's rule, and the equivalence of distance and climb," *Journal of Sports Sciences*, vol. 25, no. 6, pp. 719–726, 2007.

[14] J. M. Norman, "Running uphill: energy needs and naismith's rule," *Journal of the Operational Research Society*, vol. 55, pp. 308–311, 2004.

[15] P. Scarf, "A mathematical excursion in the isochronic hills," *Mathematics Today*, vol. 44, pp. 163–167, 2007.

[16] A. Kay, "Route choice in hilly terrain," *Geographical Analysis*, vol. 44, no. 2, pp. 87–108, 2012.