

Identifying cross country skiing techniques using power meters in ski poles

Moa Johansson¹, Marie Korneliusson¹, Nickey Lizbat Lawrence¹

Chalmers University of Technology, Gothenburg, Sweden.

moa.johansson@chalmers.se, {marieko, nickey}@student.chalmers.se

Abstract. Power meters are widely used for measuring training and racing effort in cycling, and the use of such sensors is now spreading also to other sports. Data collected from athletes' power meters are used to help coaches analyse and understand training load, racing efforts, technique etc. In this pilot project, we have collaborated with Skisens AB, a company producing handles for cross country ski poles equipped with power meters. We have conducted a pilot study on the use of machine learning techniques on sensor data from Skisens poles to identify which sub-technique a skier is using (double poling or gears 2-4 in skating). The dataset contain labelled time-series data from three individual skiers using four different sub-techniques recorded in varied locations and varied terrain. We evaluated three machine learning models based on neural networks, with best results obtained by a LSTM network (accuracy of 95% correctly classified strokes), when a subset of data from all three skiers was used for training. As expected, accuracy dropped to 78% when the model was trained on data from only two skiers and tested on the third.

1 Introduction

The development of a wide range of sensors and products such as GPS, heart-rate monitors, motion sensors and power sensors have made it possible to record a vast amount of data from athletes, providing a rich source of information to help coaches and athletes measure, analyse and understand training load, racing efforts and technique. Sports like cycling has lead the way among the endurance sports, as it its relatively easy to equip a bicycle with various sensors, for instance, to accurately measure the power in each pedal stroke. Given the relative ease at which large volumes of data can be recorded from sensors, we believe that machine learning has the potential to provide valuable tools for assisting data analysis in sports. In this pilot project, we have collaborated with Skisens AB, a spin-off company from Chalmers University of Technology that is developing a power meter for cross-country skiing, mounted inside the handle of the pole. Unlike cycling where all power comes from the legs via the pedals, in skiing the proportion of power measured in the poles depends on skiing technique. Broadly speaking, the skiing techniques are divided into classical style and freestyle, each regulated by rules in competition. Furthermore, the two styles can each be broken

down into several sub-techniques. The most effective sub-technique will depend on the terrain, the snow conditions and the individual strengths of the skier. In order for an athlete and/or coach to accurately analyse the effort based on data recorded from a race it is therefore valuable to be able to get an automated classification of which sub-technique was used where during the race. This work focuses on free-style technique, however, the methods may be applied also to classical style.

A longer version of this paper is available as a technical report [5].

2 Related Work

There has been several previous works aiming at classifying cross-country skiing technique using a variety of sensors, following the initial experiments with wearable sensors by Marshland et al. [6]. Stöggl et al. used accelerometer data from a mobile phone attached to a belt around the chest of the skier and a Markov chain model to classify strokes [3, 11]. Rindal et al. used wearable inertial measurement units (IMUs) attached to the skiers arms and chest, together with gyroscopes attached to the skiers arms [8]. Sakurai et al. also used data from several IMUs attached to the skis and poles to construct a decision tree classifier both for classical and skating techniques [9, 10]. Recently, Jang et al. conducted a study using wearable gyroscope sensors to identify both classical and skating techniques and a deep machine learning model combining CNN and LSTM layers [4]. The main difference between our work and the above ski technique classifiers is that we do not use any dedicated wearable sensors for the task, but simply explore if we can identify technique using only the sensors already present in the Skisens pole for measuring power. Our sensor data only records the movements of the hands, and does not include any sensors on the body or on the skis, which would make the task easier.

3 The Dataset

The dataset consists of data from three individuals (male, experienced recreational skiers). The data was collected on roller skis on different days, in varied terrain and under varied conditions. There were both uphill and downhill sections as well as turns, with skiers using double poling plus three different skating styles, referred to as Gear 2, Gear 3 and Gear 4, following notation in [7]¹. For each gear there are a number of disjoint data segments, where each segment is a continuous time-series of data during which the skier only uses a specified style. The data collected is summarised in Table 1. Data was recorded at 50 Hz (50 samples per second), hence when we refer to time-steps, these are recorded 0.02 seconds apart. The raw data was pre-processed and longer segments divided into short segments, containing a single stroke each. The resulting dataset contains 1671

¹ We note that the notation varies between different countries, these techniques are sometimes also referred to as V1, V2 and V2a. See [7] for a discussion.

individual strokes, of which 252 strokes in Gear 2, 473 in Gear 3, 360 in Gear 4 and 585 strokes using double poling. Each single-stroke sample is 140 time-steps long, zero-padded when necessary.

Table 1: Description of the dataset columns used for machine learning. The coordinate system for the vectors of acceleration and angular velocity is relative to the pole with (a) First axis: Pointing right (orthogonal to pole), (b) Second axis: Pointing down (parallel to pole), and (c) Third axis: Pointing forward (orthogonal to pole)

No.	Data	Unit
1	Time	second
2	Force in the left pole	Newton
3	Pole-ground angle of the left pole	degrees
4 - 6	Left angular velocity	rad/s
7 - 9	Left acceleration	m/s ²
10	Force in the right pole	Newton
11	Pole-ground angle of the right pole	degree
12-14	Right angular velocity	rad/s
15-17	Right acceleration	m/s ²

We remark that the data recorded also included the GPS position, but we choose not to include this information as a feature. Different techniques are naturally used at distinct road segments, as some techniques are more natural to use e.g. in uphill terrain. If this was included, the models would end up basing their predictions primarily on GPS-position, ignoring the other features, which would lead to poor performance on unseen data recorded in a different location.

4 Machine learning models

We experimented with three different types of deep machine learning models for stroke classification: a long short term memory network (LSTM) [2], a bidirectional long short term memory network (BLSTM) [1], and a one dimensional convolutional neural network (CNN). The models were implemented in Python using the Keras/TensorFlow libraries². The code is available online³.

The LSTM model in our experiment combines an LSTM cell with two dense layers (see fig. 1). The input of the LSTM model is a sequence of 140 time-steps, corresponding to one pole push. The number of neurons in each layer was chosen experimentally. The two dense layers can be interpreted as a weighted majority vote, weighing the importance of each time-step for giving a result of the most likely class for the entire pole push.

The BLSTM network has the same basic architecture as the LSTM network, but with 64 neurons in the first layer (chosen experimentally). While the LSTM

² <https://www.tensorflow.org/guide/keras>

³ <https://github.com/maajohansson/ai-in-sports>

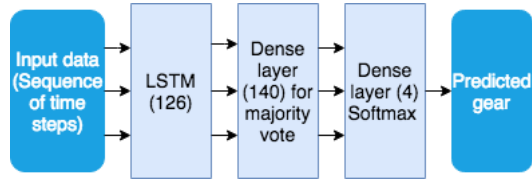


Fig. 1: Network architecture for the LSTM model, with an LSTM cell with two dense layer. The light blue boxes indicates layers in the network, and the number of neurons in each layer is stated inside the brackets in each layer.

network passes information only in the forward direction, the BLSTM network passes information in both the forward and backward direction, thus using twice as many weights.

Our CNN model consists of two one-dimensional convolutional layers and two dense layers (see fig. 2), as well as max-pooling and global max-pooling layers. The latter two layers are used for down-sampling, locally and globally. Number of neurons, filters, kernel- and pool size were decided experimentally.

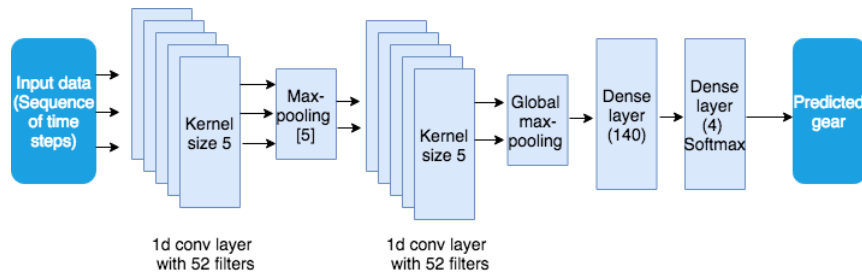


Fig. 2: The network architecture for the CNN model.

5 Experiments and Results

We conducted two experiments to assess classification accuracy. In Experiment 1, both the training set and unseen test set contain data from all three skiers. In Experiment 2, two skiers are used for training and the third for testing.

Experiment 1: We trained the models on a subset of the data containing samples from all three skiers, and evaluated on another, unseen, subset as test data. We suspect that the same person performs strokes in the same techniques in a relatively consistent manner, hence the strokes in the test set are likely to be quite similar to something from the training set. A motivation for this kind of experiment is envisaging an application using Skisens-sensors which is

personalised to the owner, who initially “calibrates” the product by skiing using specified sub-techniques to collect personal training data.

Experiment 1 was performed for all three models described above, using five-fold cross-validation, with each fold containing approximately the same number of strokes and the same proportion of strokes in each sub-technique (folds 1-4 of 329 strokes, fold 5 of 355 strokes, from the total dataset of 1671 strokes).

The results are promising with both the LSTM and BLSTM models reaching an accuracy of 95% on average over the five folds. We note that the CNN model performed slightly worse than the other two, reaching an average accuracy of 90%, with higher variation over the different folds. We suspect that the CNN model suffered more than the LSTM-based models from the relatively small dataset. We note that the LSTM-based models also contains more trainable parameters than the CNN-model, so more experimentation is needed with different CNN architectures. Training took about 6-10 times longer for the LSTM and BLSTM models compared to the CNN model.

For the best-performing model (LSTM), we note that Gear 4 and double poling were the easiest to classify, while Gear 3 was the hardest. This was somewhat surprising, as the arm movements of Gear 4 and double poling are visually quite similar.

Experiment 2: Experiment 1 does not test the capability to generalise to a person not seen before. This was somewhat difficult to test, due to the small dataset. However, we did a second experiment with the best-performing model from Experiment 1 (the LSTM model) where we trained on data from two skiers, and evaluated on unseen data from the third individual. This was expected to be harder, as the model would have to generalise, and ideally learn how an “average” stroke in each sub-technique would be represented by the sensor data. As expected, performance dropped to 78%. We believe that this could be improved by training on a larger dataset with samples from many individuals, and performing a larger study is future work.

6 Discussion and Further Work

We have conducted a pilot study using data from sensors fitted to ski pole handles to predict which technique or gear the skier is using. This experiment aimed at classifying time-series for single strokes, as these are easy to identify from the power data recorded from the poles (near-zero readings indicating when the poles are in the air). We have not yet attempted the task of passing in continuous sequences of skiing strokes and identifying gear changes. This is an interesting problem, as some previous work, e.g. [8], report that mis-classifications of single strokes often happen near change points.

Most other works in cross-country skiing technique classification come from the sports science domain, and often include only a few individuals in the studies (e.g. 10 skiers in [8], four skiers in [4]). Furthermore, these studies often primarily focus on reaching high accuracy for these specific individuals (often elite athletes).

Experiments are often in the style of our Experiment 1, i.e. the training set and test set contain data from the same individuals. On this task, our LSTM model reached an accuracy of 95%, which is similar to other models from the literature [3, 11, 8]. In the setting of Experiment 2, with tests on unseen individuals, Jang et al. [4], reports an accuracy of between 87.2% – 95.1%, compared to ours at 78%, but they had access to more data.

Our dataset, of merely 1671 strokes, is on the small side for deep learning, as seen in Experiment 2. We are however encouraged by the results in this study to gather a larger dataset and perform a larger evaluation. We would like a dataset containing both professional and recreational skiers to investigate whether one can train a model to generalise without taking small individual variations into account. This is particularly relevant from the perspective of Skisens, as they are interested in including technique classification together with their ski-pole sensors in for example a smart sports watch. Ideally, one would like to have a pre-trained model which does an acceptable job out of the box, and possibly then adapts to the individual user, without having to be trained from scratch.

References

1. A. Graves and J. Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 2005.
2. S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 1997.
3. A. Hols and A. Jonasson. Classification of movement patterns in skiing. In *Frontiers in Artificial Intelligence and Applications: Twelfth Scandinavian Conference on Artificial Intelligence*, volume 257, 2013.
4. J. Jang, A. Ankit, J. Kim, Y. Jang, H. Kim, J. Kim, and S. Ziong. A unified deep-learning model for classifying the cross-country skiing techniques using wearable gyroscope sensors. *Sensors*, 18(11), 2018.
5. M. Johansson, M. Korneliusson, and N. Lizbat Lawrence. Identifying cross country skiing techniques using power meters in ski poles. arxiv.org/abs/1904.10359, 2019.
6. F. Marshland, K. Lyons, J. Anson, G. Waddington, C. Macintosh, and D. Chapman. Identification of cross-country skiing movement patterns using micro-sensors. *Sensors*, 12(4), 2012.
7. J. Nilsson, P. Tveit, and O. Eikrehagen. Effects of speed on temporal patterns in classical style and freestyle cross-country skiing. *Sports Biomechanics*, 3(1), 2004.
8. O. Rindal, T. Seeberg, J. Tjønnås, P. Haugnes, and Ø. Sandbakk. Automatic classification of sub-techniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data. *Sensors*, 18(2), 2017.
9. Y. Sakurai, F. Zenya, and Y. Ishige. Automated identification and evaluation of subtechniques in classical-style roller skiing. *Journal of Sports Science and Medicine*, 13, 2014.
10. Y. Sakurai, F. Zenya, and Y. Ishige. Automatic identification of subtechniques in skating-style roller skiing using inertial sensors. *Sensors*, 16, 2016.
11. T. Stöggl, A. Holst, A. Jonasson, E. Andersson, T. Wunch, C. Norström, and H.-C. Holmberg. Automatic classification of the sub-techniques (gears) used in cross-country ski skating employing a mobile phone. *Sensors*, 14, 2014.