

# Bag-of-steps: Predicting Lower-limb Fracture Rehabilitation Length

Albert Pla<sup>1</sup>, Beatriz López<sup>1</sup>, Cristofor Nogueira<sup>1</sup>, Natalia Mordvaniuk<sup>1</sup>,  
Taco J. Blokhuis<sup>2</sup> and Herman R. Holtslag<sup>2</sup> \*

1- University of Girona - Institut d'Informatica i Aplicacions  
{albert.pla,beatriz.lopez,natalia.mordvaniuk}@udg.edu - Girona, Spain

2- University Medical Center Utrecht  
{t.j.blokhuis,h.r.holtslag}@umcutrecht.nl - Utrecht, The Netherlands

**Abstract.** This paper presents bag-of-steps, a new methodology to predict the rehabilitation length of a patient by monitoring the weight he is bearing in his injured leg and using a predictive model based on the bag-of-words technique. A force sensor is used to monitor and characterize the patient's gait, obtaining a set of step descriptors. These are later used to define a vocabulary of steps that can be used to describe rehabilitation sessions. Sessions are finally fed to a support vector machine classifier that performs the final rehabilitation estimation.

## 1 Introduction

Hip and lower-limb fractures are some of the most common lesions amongst the elderly population. This kind of injuries produces a high morbidity and mortality [1] due to both the direct impact of the injury and the often fragile health status of elder patients. In addition, many patients with lower-limb fractures become highly dependent, meaning that they need to be constantly assisted by care providers or need to move into residential care institutions. Therefore, when patients are not able to recover their pre-injury degree of mobility, they experience a loss of autonomy. Moreover, this situation causes an increase of costs for the health care systems [2]. Given the current aging of the population, tackling the issues regarding mobility reduction due to hip and lower-limb lesions becomes a key factor for improving the autonomy and quality of life of patients while reducing the associated costs of the after-surgery processes.

Currently, there is no general standard for mobilization after hip and lower-limb fractures. Nevertheless, an important aspect during the mobility recovery process is the proper weight load on the affected lower extremity. This procedure needs to be guided by a therapist as an inappropriate weight load might harm the patient. Therefore, qualitative monitoring of the patient's weight distribution is useful to assist therapists and patients during the rehabilitation period. Such monitoring also offers the opportunity to assess the data in order to analyze

---

\*The work described in this paper was carried out as part of the MoSHCA project funded by the spanish INNPACTO program and the European ERDF funds (Ref. EUREKA ITEA 2 n° 11027 - IPT-2012-0943-300000 / Ref. DPI2013-47450-C2-1-R). Evalan has had no influence on the interpretation of data and the final conclusions drawn. The Sensistep device is available for the European market ([www.sensistep.com](http://www.sensistep.com))

the recovery of the lesion; in this sense, recent studies show that there exists a correlation between the weight loading pattern of the patient and the speed of rehabilitation [3]. Thus, we expect that the use of machine learning techniques can allow the prediction of the rehabilitation length of a patient; On the one hand, this information will allow therapists to know if a patient is recovering as expected or if the therapy is being effective. On the other hand, having an estimation of patients' rehabilitation length will allow hospitals and health institutions to optimize and adapt the hospital resources and agenda according to the patients' needs.

In this work we present a methodology to estimate the rehabilitation length of elderly patients who suffered an injury or surgery in a lower limb. Our approach proposes to use a force sensor (*Sensistep* [4]) to gather data of the patients activity' and to analyze it using *Bag-of-Steps*, a new reasoning method based on the bag-of-words pipeline [5]. The sensor registers the force that a patient is exerting on his leg during therapy; following, the gathered information is analyzed using a reasoning engine that, following a bag-of-words pipeline, estimates if the patient is evolving in a proper way and how much time he will need to recover his mobility. The proposed methodology is tested using real data.

## 2 Related Work

Despite the problem proposed in this paper still has not been addressed by the research community, gait analysis has been the focus of different studies during recent years . The analysis of the human gait cycle has proven to be a valuable tool for dealing with problems related with the aging of the population such as the treatment of Parkinson [6] or risk of falling evaluation [7]; these works mostly focus on analyzing how the gait cycle is related with the evolution of the diseases and how the gait needs to be represented. In our work we use artificial intelligence techniques to analyze such descriptors in order to help therapists during the rehabilitation process. Similar approaches have proven to be successful for aiding in the treatment of chronic diseases [8] and in veterinary applications [9].

The bag of words methodology is well known for its usefulness in text and document classification. In it, each piece of text is represented as a set of words belonging to a vocabulary; then, each text is classified using the appearance frequency of each word. This methodology has been broadly adopted in the computer-vision field, where it is mainly used for object and scene recognition [5, 10], but it has also been explored in other domains such as bioinformatics [11] or sound recognition [12]. In this work we inspire in such previous research and we apply this methodology for gait recognition and rehabilitation length prediction.

## 3 Bag-of-Steps

Our approach is based on the bag-of-words methodology [5] used in the text recognition and computer vision domains: identify and describe the features that describe a rehabilitation session (the steps); group steps that share similar

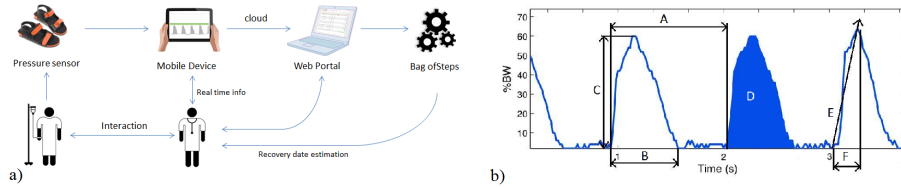


Fig. 1: Left: Methodology overview. Right: Step characterization features: a) Stride b) Heel stance time c) Gait peak d) Impulse e) Loading response f) Peak instant.

characteristics into step stereotypes or *words* in order to create a vocabulary; and, finally, build a classifier relating the stereotypes appearing in a particular rehabilitation session and the rehabilitation length of the patient.

**Data gathering** is performed using the SensiStep technology developed by Evalan. The system consists of an axial force sensor that is placed inside a special sandal. The sensor registers how much axial force is exerted on the leg [3]. This provides information about the intensity of movement and the amount of weight borne on the leg. The sensor continuously streams the recorded data to a special feedback module. This module compares the actual weight loading with the instructions set by the therapist in order to guide the patient during his exercises.

Once the rehabilitation session finishes, the signal recorded by the sensor is smoothed using a Gaussian filter. Then, steps are segmented according to the two following criteria provided by therapists: in a step the patient's leg bears more than a 10% of its weight between 0.2 and 2 seconds; a step has a loading peak of at least the 20% of the patient's weight. Finally, the segmented steps are described using six characteristic features of gait analysis [3]: stride, heel-stance time, impulse, loading response, peak and peak instant (see Figure 1). Once a patient finishes his rehabilitation session, the session is represented as a collection of steps with different features.

The next step to follow is the **generation of the vocabulary**. In text and document classification, the vocabulary or the bag-of-words methodology is created by identifying the most repeated and significant words that appear in the training documents. In the problem we are dealing with, each rehabilitation session is composed by several records of steps; thus each step recorded could be considered a word. Nevertheless, contrary to the text analysis domain, it is unlikely to find two steps with the exact same feature descriptors. Therefore, we propose to construct the vocabulary of the *bag-of-steps* in a similar way that is done in computer vision.

The vocabulary of the *bag-of-steps* can be constructed by clustering all the steps available in the training data and using each cluster centroid as the code-word. Thus, each centroid represents a step stereotype and we can say that the *bag-of-steps* vocabulary is composed by step stereotypes. The selection of the clustering algorithm and its parameters can have an important influence in the solution of the problem. Given that the step stereotypes will act as histogram

bins, an inappropriate number of stereotypes can harm the efficiency of the *bag-of-steps*. A too small number of clusters will result in a low discriminative power whilst too many clusters would result in noise introduction and overtraining. To avoid such issues, we propose to use clustering algorithms where the number of clusters to be found ( $K$ ) is determined by the user (e.g. *K-means* or *K-medoids*), and to define the  $K$  value in concordance with the size of the training data ( $n$ ) by means of the thumb rule ( $\sqrt{n/2}$ ) or the Rice's rule ( $2n^{1/3}$ ).

Once the vocabulary is defined, each step in the training data is labeled with a particular step stereotype. Therefore, each rehabilitation session can be described as a step stereotype histogram of  $K$  bins.

Next, a **predictive model** should be trained in order to distinguish between patients with different rehabilitation patterns. Literature offers a great number of classifiers to tackle this stage of the bag-of-words methodology: artificial neural networks, support vector machines, decision trees and forests, etc [11]. In this paper we study the use of 2 different classifiers, nearest neighbour (NN) and support vector machines (SVM), but it is important to note that *bag-of-steps* can be implemented with other classifiers. The NN classifier, one of the simplest classifiers available, will be useful to establish a baseline for the *bag-of-steps* performance. On the other hand, we opt for support vector machines (SVM) [7] as they offer high generalization performance without the need of *a priori* knowledge and regardless of the dimensionality of the inputs (which will be conditioned by the number of clusters used in the previous step).

Once a patient performs a new rehabilitation session, its data will be analyzed in order to estimate its rehabilitation length. The steps to follow are similar to the ones used in the training phase (described above): to identify and characterize the steps of the rehabilitation session; use the created vocabulary to generate the histogram of stereotypes defining the rehabilitation session; and, finally, to use the predictive model to predict the rehabilitation length of the patient.

## 4 Experimentation

The methodology presented in this paper has been tested using data containing information of patients who underwent surgery after a lower-limb fracture. The data is composed of 48 different rehabilitation sessions with different rehabilitation lengths (the time period between his surgery and his official discharge date). On average, each rehabilitation session was composed of 1200 steps. For the experiments, the data was distinguished between sessions belonging to a patient with a long (more than 56 days) or a short (56 days or less) rehabilitation period. The threshold between the two classes was defined according to the therapists' expertise. The number of rehabilitation sessions corresponding to each class was balanced (24 long and 24 short).

Experiments have been performed following a 10 fold cross validation methodology. We have built the vocabulary for the tests using K-medoids; to evaluate the influence of the vocabulary size the experiment has been performed with

Total Rehabilitation Length: 10-fold Cross Validation						
$K$	20	50	81	129	150	200
Mean accuracy (SVM)	63.08%	80.00%	87.69%	81.53%	80.00%	72.30%
Standard dev. (SVM)	3.94%	5.01%	4.07%	3.78%	3.71%	3.36%
Mean accuracy (NN)	56.92%	77.31%	67.69%	70.77%	69.23%	58.46%
Standard dev. (NN)	3.56%	4.84%	3.14%	3.28%	3.21%	2.71%

Table 1: Accuracy when predicting the total rehabilitation length of a patient when using different classifiers (*cls*) and different vocabulary sizes ( $K$ )

different  $K$  parameters (20, 50, 81<sup>1</sup>, 129<sup>2</sup>, 150, 200). In the experiments we have implemented *bag-of-steps* using a SVM with a quadratic kernel function and using NN.

The *bag-of-steps* methodology presented satisfactory results (Table 1), obtaining an accuracy higher than the 80% in the best configurations (using SVM). Nevertheless, the data shows that the size of the vocabulary and the classifier used to build the predictive model have a significant influence on the method performance. If we focus on the classifiers performance, as expected, SVM outperforms NN with all the tested configurations. It is remarkable that with all the SVM configurations but  $K = 20$  and  $K = 200$ , the accuracy of steps is above the 80%; being 87.69% the best result when using a  $K$  value of 81. The best configuration for the NN classifier ( $K = 50$ ) obtains an accuracy of 77.31%, a quite high value for a classifier that can be considered as the *bag-of-steps* baseline; when comparing this result with the SVM it can be seen that this only improves the results of the worst SVM configurations. A Wilcoxon T test ( $\alpha = 0.05$ ) confirms that *bag-of-steps* obtains better results when using SVM than when using NN under any configuration but with  $K = 50$  and  $K = 20$ , in which they obtain similar results.

When analyzing the influence of the  $K$  parameter, the results point that a too small or a too high  $K$  value can decrease the performance of *bag-of-steps* as the worst results for each classifier have been obtained when using  $K = 20$  and  $K = 200$ . On the case of NN classifier the best results are obtained with  $K = 50$  a whilst when using SVM the best results are when using a  $K$  value obtained from the thumbs and the Rice's rule. This fact suggests that the  $K$  value should be carefully tuned according to the classifier used in *bag-of-steps* and that the thumb and the Rice's rule can be useful but are not a silver bullet.

## 5 Conclusions

This paper presented *bag-of-steps*, a methodology for predicting the rehabilitation length and the discharge date of patients who suffered a limb-related surgery and are being rehabilitated. The methodology, uses force sensors which gathers the weight load of the leg of the patient during a rehabilitation session. The recorded data is then analyzed in order to detect the gait of the patient. The

<sup>1</sup>Number of clusters suggested by the Rice's rule.

<sup>2</sup>Number of clusters suggested by the thumb rule.

gait is then characterized using step descriptors that are compared with other step stereotypes to finally build a step histogram for the rehabilitation session. Using a classifier (SVM or k-NN), the histogram is then used to estimate the rehabilitation length of the patient.

The methodology has been tested using data from real patients. Preliminary experimentation has shown that *bag-of-steps* can estimate the rehabilitation length with an accuracy higher than the 80%. Nevertheless, further experimentation should be carried out in order to determine if the method can also perform more complex predictions such as regressions. To this end, further works include collecting information from more patients as it is expected that by including new patients into the model, the predictive power of the methodology will also increase. In addition, testing alternative clustering and classification techniques would also be an interesting work to study the flexibility of the methodology.

## References

- [1] J.J.W. Roche, R.T. Wenn, O. Sahota, and C.G. Moran. Effect of comorbidities and postoperative complications on mortality after hip fracture in elderly people: prospective observational cohort study. *Bmj*, 331(7529):1374, 2005.
- [2] M. Bonafede, D. Espindle, and A.G. Bower. The direct and indirect costs of long bone fractures in a working age us population. *J. of medical economics*, 16(1):169–178, 2012.
- [3] A. Bakker, T.J. Blokhuis, M.D. Meeks, H.J. Hermens, and Herman R Holtslag. Dynamic weight loading in older people with hip fracture. *Journal of rehabilitation medicine (UEMS European Board of Physical and Rehabilitation Medicine)*, 2014.
- [4] M Raaben and TJ Blokhuis. De technisch geneeskundige: een nieuwe professional in de zorg. *Nederlands Tijdschrift voor Traumatologie*, 23(5):109–112, 2015.
- [5] Hanna M Wallach. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning*, pages 977–984. ACM, 2006.
- [6] L. Rochester, A. Nieuwboer, K. Baker, V. Hetherington, A-M Willems, F. Chavret, G. Kwakkel, E. Van Wegen, I. Lim, and D. Jones. The attentional cost of external rhythmical cues and their impact on gait in parkinson’s disease: effect of cue modality and task complexity. *Journal of neural transmission*, 114(10):1243–1248, 2007.
- [7] N Shibuya, BT Nukala, AI Rodriguez, J Tsay, TQ Nguyen, S Zupancic, and DYC Lie. A real-time fall detection system using a wearable gait analysis sensor and a support vector machine (svm) classifier. In *IEEE ICMU 2015*, pages 66–67, 2015.
- [8] A Sama, C Perez-Lopez, J Romagosa, D Rodriguez-Martin, A Catala, J Cabestany, DA Perez-Martinez, and A Rodriguez-Moliner. Dyskinesia and motor state detection in parkinson’s disease patients with a single movement sensor. In *Engineering in Medicine and Biology Society (EMBC 2012)*, pages 1194–1197. IEEE, 2012.
- [9] M. Kaijima, T. L Foutz, R. W McClendon, and S.C. Budsberg. Diagnosis of lameness in dogs by use of artificial neural networks and ground reaction forces obtained during gait analysis. *American journal of veterinary research*, 73(7):973–978, 2012.
- [10] J. Yang, Y. Jiang, A.G. Hauptmann, and C. Ngo. Evaluating bag-of-visual-words representations in scene classification. In *Int. Workshop on multimedia information retrieval*, pages 197–206. ACM, 2007.
- [11] J. Wang, P. Liu, M. FH She, S. Nahavandi, and A. Kouzani. Bag-of-words representation for biomedical time series classification. *Biomedical Signal Processing and Control*, 8(6):634–644, 2013.
- [12] S. Chu, S. Narayanan, and C.C.J. Kuo. Environmental sound recognition with time-frequency audio features. *Audio, Speech, and Language Processing, IEEE Transactions on*, 17(6):1142–1158, 2009.