Master Thesis Proposal

Variable importance and feature selection in machine learning: A game-theoretical approach

by Firstname Lastname
(matriculation number)

under the supervision of

Prof. Eyke Hüllermeier
Intelligent Systems
Department of Computer Science
Paderborn University

January 25, 2017

1 Motivation and Background

In recent years, the use of the Internet, mobile devices and databases have significantly increased the volume of data available for commercial and research purposes such as in the domain of bioinformatics, geomatics and search engines. However, the recent proliferation of large data, with hundreds to thousands of features, possess unprecedented challenges. In an available data, not all variables or features are useful for an application as they may have a negative effect on performance which requires selecting only suitable subset of features, called feature selection.

Feature selection is a preprocessing step in the field of machine learning as it significantly reduces the running time of a learning process by removing irrelevant and redundant features. After selecting a suitable set of features, learning algorithms can focus on the most important aspects of data, without the intervention of irrelevant, redundant and noisy features, improving the overall classification performance. The process of feature selection assists in building a simpler and a generalized learning model. It helps in increasing the prediction accuracy and reduces training times of algorithms [1]. The objectives of feature selection are mentioned below.

1. Find a necessary and a sufficient minimum subset of features from a group of features [3].
2. Select a subset of features of size S such that the value of a criterion function is optimized over all subsets of same size.
3. Decrease the size of the feature set without significantly decreasing the prediction accuracy [4].

Real datasets comprise of limited number of features. In order to find an optimal subset of features from a group of features, all feature subsets must be evaluated. The requirement for evaluation arises as one attempts to estimate the contribution of each and every feature in classification. The significance of each feature in classification process can be established by considering all possible coalitions with other features. The Shapley value, a concept from coalitional games, accommodates all contributions of a feature for classification. It is a well-known concept from game theory, where it is utilized to calculate average contribution of each player involved in the game by taking into account their interactions in other subsets as well [8].

There is a serious limitation associated to the Shapley values calculation as it demands calculating the marginal contribution of all subsets and hence requires exponential computation time. This method is possible only for a small number of features and if the number of features are increased then computation becomes very costly. It demands introducing an approximation technique for Shapley values that will reduce the computation time for a large number of features. Shapley value approximation with low computation time is vital for the overall machine learning process, as feature selection itself adds an additional layer of computations to the overall learning process and if this layer is computationally expensive, then the overall process of learning will become computationally expensive.

2 Goals of the Thesis

The overall aim of this thesis is to address the approximation of Shapley values, theoretically or empirically, utilizing approximation approaches e.g., Monte Carlo simulations which are used to

approximate solutions of computationally hard problems. A good approximation approach is to come up with an adaptive algorithm of estimating Shapley values in a way that approximation becomes more certain and more accurate with time. Reducing the computation time of overall feature selection process using this approximation and eventually reducing the overall time of learning will be targeted as well. Different approximation techniques for estimating Shapley values will be tried and their results will be recorded both in terms of computation time and accuracy. A trade-off between accuracy and time complexity will be investigated for Shapley values in a way that average error must decrease with time.

It will also be investigated how Shapley value approximation can be integrated with heuristic feature selection approaches such as forward or backward feature selection in order to reduce their greedy nature. Investigating this problem may assist to resolve the problems associated with greedy algorithms such as overfitting and chances of selecting a local optimum. Explicit goals of the thesis have been mentioned below.

1. Develop an adaptive approach of approximating Shapley values.
2. Implementing a Shapley value based framework for feature selection, producing intermediate or final results with associated uncertainties in a fixed budget of time.
3. Integration of Shapley value approximation approach with forward and backward feature selection approaches to reduce their greedy nature.
4. Extensive empirical evaluation of the integrated algorithm defined above with forward and backward feature selection approaches.
5. Empirical comparisons of the adaptive Shapley value approximation method with existing random Shapley value approximation approaches of feature selection on real datasets.

3 Related Work

In the literature several feature selection techniques exist such as wrapper, filter or embedded approaches but there are associated advantages and drawbacks with each approach. Wrapper approaches of feature selection have been discussed by Khavi and John in which a subset is generated and its performance is evaluated by training and testing on a particular learning algorithm. However, the feature space of subsets increases exponentially with increase in the number of features and becomes computationally very expensive and prone to overfitting [10]. Filter approaches of feature selection focus on intrinsic properties of the data without taking into account the classification algorithms which makes them simple, fast and scalable. A visible disadvantage of filter approach is that they ignore the interaction and dependencies of features in order to calculate optimal subset of feature [1]. Shapley values has also been utilized as a feature selection method by Karlson Pfannschmidt and Eyke Hllermeier in [8].

Previously, many approximation techniques have been utilized to approximate Shapley values. Mann and Shapley proposed an approximation method [6] using Monte Carlo simulations which approximates the Shapley value from randomly sampled coalitions. The advantage of this method is linear time complexity but the disadvantage lies in the lack of information about the sampling technique. To overcome the problem of the sampling approach, Bachrach and Rosenschein provide an analysis of an error bound and the minimum number of samples required to achieve a given accuracy. Owen proposed a Multi Linear Extension (MLE) approximation

method [7] which is also linear time but has been implemented on weighted voting games by considering the weights as normally distributed. A modified MLE method [5] has been proposed by Leech which has better approximation than MLE but with exponential time complexity. Zlotkin and Rosenschein proposed a linear time random permutation mechanism [11] in which a full coalition is established by adding one player after another.

Most of the researchers approximated the Shapley value for specific classes of games. Castro et al [2] has attempted to approximate the Shapley value for general classes of games and provided an asymptotic bound on approximation error. On the other hand, where the variance and range of marginal contributions is available, a non-asymptotic bound on estimation error has been presented in [9].

4 Thesis Outline and Time Schedule

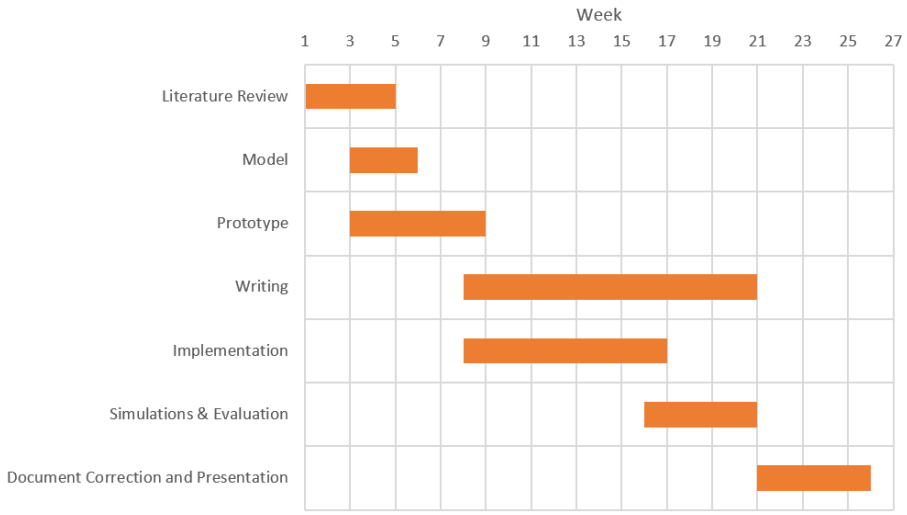
Tentative structure of the thesis

- Introduction
  - Background
  - Problem Statement
  - Objectives and Goals
  - Motivation and Significance
  - Limitations
- Literature Review
  - Filter, Wrapper and Embedded Approaches
  - Correlation Based Feature Selection
  - Shapley Value Approximation Methods
- Theoretical Foundations
  - Co-operative Game Theory and Shapley Values
  - Statistical Distributions
  - Bayesian Statistics and Parameter Estimation
  - Sampling Techniques
  - Greedy Nature of Feature Selection Techniques
- Proposed Feature Selection Methods
  - Non-Parametric Shapley Value Estimation
  - Parametric Shapley Value Estimation
  - Greedy Nature Reduction
- Experimental Results
  - Introduction
  - Medical Dataset

– Other Datasets

– Summary

Time schedule



References

[1] Isabelle Guyon and Andre Elisseeff. *An Introduction to Variable and Feature Selection.* http://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf, 2003.

[2] Javier Castro Daniel Gmez and Juan Tejada. *Polynomial calculation of the shapley value based on sampling.* Comput. Oper. Res., 36(5):17261730, May 2009.

[3] K. Kira and Rendell. *L. A. The feature selection problem: Traditional methods and a new algorithm.* Proceedings of Ninth National Conference on Artificial Intelligence, 129- 134, 1992.

[4] D. Koller and M. Toward Sahami. *optimal feature selection.* Proceedings of International Conference on Machine Learning, 1996.

[5] D. Leech. *Computing power indices for large voting games.* Management Science 49 (6), 831837, 2003.

[6] L.S. Shapley Mann. *Evaluating the electoral college by Monte Carlo techniques, Technical report.* The RAND Corporation, Santa Monica, 1960.

[7] G. Owen. *Multilinear extensions of games.* Management Science 18 (5), 6479, 1972.

[8] Karlson Pfannschmidt, Eyke Hllermeier, Susanne Held, and Reto Neiger. *Evaluating Tests in Medical Diagnosis: Combining Machine Learning with Game-Theoretical Concepts.* Communications in Computer and Information Science, vol 610. Springer, Cham, 2016.

[9] Sasan Maleki Long Tran-Thanh Greg Hines Talal Rahwan and Alex Rogers. *Estimation Error of Sampling-based Shapley Value Approximation with/without startifying.* June 2013.

[10] R.Kohavi and G.H.John. *Wrappers for feature subset selection.* Artif. Intell, vol. 97, no. 12, pp. 273324,, 1997.

[11] G. Zlotkin and J. Rosenschein. *mechanisms foe coalition formation in task oriented domains.* Proceedings of the National Conference on Artificial Intelligence (AAAI-94), 1994.

Agreement

Both the candidate and the supervisor agree on the goals and objectives of the thesis as outlined above.

(Name of supervisor)	(Name of student)