

A hierarchical multi-level product classification workbench for retail

Maximilian Harth, Christian Schorr, and Rolf Krieger

Trier University of Applied Sciences, Environmental Campus Birkenfeld,
55761 Birkenfeld, Germany
c.schorr@umwelt-campus.de

Abstract. Exploratory data analysis and especially model evaluation get difficult when facing the challenge of classifying product data according to a hierarchical classification system (HCS) containing thousands of categories as used in the retail industry. The identification of incorrectly classified products and the optimization and monitoring of automatic classification algorithms is very time-consuming. To solve this problem we propose a workbench which provides an interactive graphical user interface (GUI) for exploratory product data analysis and model evaluation taking into account the structure of an HCS and supplying statistical insights. In addition, the workbench offers an integrated machine learning based product classification module which can classify products on the fly using their names only.

Keywords: Product classification, machine learning, data exploration, hierarchical classification systems

1 Introduction

The management of product data is an important task in retail companies. Products must be classified based on a hierarchical product classification system (HCS) which defines categories of products and relations between them. The assignment of products to the categories is based by either implicitly or explicitly defined attributes [1]. A well-known product classification standard for retail is the Global Product Classification (GPC) described in [2]. It defines a four-level hierarchy consisting of 38 segments, 118 family, 823 class and 4226 brick codes to describe products. It is used by 60.000 German companies and over 1.5 million companies world-wide. In many cases standardized HCSs coexist with company-specific ones that have a similar number of categories and hierarchical levels. Consequently, new products have to be continuously classified into different hierarchical classification systems.

Product data is of central importance in retail companies. There are companies having millions of product data records. Correct assignment of products to an HCS has a

decisive influence on data quality, the execution of business processes and on a seamless data exchange between business partners. Consequently, consistent (re)classification between different HCSs is a complex, time-consuming and error-prone task, which must often be performed manually. Due to the large number of categories it is difficult to classify the products consistently. It is therefore necessary to identify and correct misclassifications.

In addition, to reduce the effort required for the classification, there are numerous approaches for automatic classification based on machine learning. The typical data science project can be described by an iterative process encompassing the steps business and data understanding, data exploration and cleaning, modeling and deployment. Widely-used process models are the cross-industry-standard for data mining [3] or Microsoft's Team Data Science Process [4]. Modeling encompasses feature engineering, model training and evaluation. For all these steps a plethora of software and libraries is available. Metrics like precision, recall or F_1 -score can summarize the model quality in a few numbers but understanding the results that way and learning from incorrectly classified products is very time-consuming. Confusion matrices are also often employed to visually determine systematic misclassifications at a glance. Keeping in mind that a HCS may have thousands of different classes at the lowest level, though, the resulting confusion matrix has several million entries and is not easy to interpret. For a retail company a tool which allows to interactively explore a HCS and the predicted classes would be a valuable tool.

Consequently, we propose a workbench which provides an interactive graphical user interface (GUI) for exploratory product data analysis, identifying classification errors and model evaluation taking into account the structure of a HCS and supplying statistical insights using dendrograms, tree maps and box plots at the different levels of an HCS, for example. It also features a machine learning module to (re)classify products on the fly according to the underlying HCS. To summarize the workbench supports the process of the manual and automatic product classification based on several hierarchical classification systems.

2 Related Work

Much research is undertaken in the field of product classification with machine learning - a recent and comprehensive comparison of classification algorithms can be found in [5].

Sun et al. suggests a hybrid algorithm called Chimera. A mix of crowd outsourced manual classification, machine learning and data quality is used in addition to rules formulated by in-house analysts [6]. Based on product name, product description and several other attributes, several tens of millions of products are classified into more than 5000 categories. Ha et al. suggest a deep learning-based strategy employing multiple recurrent neural networks (RNNs) [7]. In addition to the product name, brand, manufacturer and the top level category – all in Korean - are used. The data consists of more than 94 million products with 4016 low-level categories.

Several classification methods on a hierarchical data set of product descriptions are studied by Ding et al. [8]. A hierarchical classification approach using the UNSPSC categories leads to a significantly worse result, contrary to intuition. Cevahir and Murakami present a classification model assigning products to one of 28.338 possible categories on a five-tier taxonomy using 172 million Japanese and English product names and descriptions [9]. The model combines deep belief nets, deep auto-encoders and k-Nearest Neighbour-classification (kNN). In [10] the authors present a classification model for GPC categorized products using partially abbreviated German product names only.

In summary, we can say that the field of automatic classification using machine learning is highly topical. Often hierarchical classification systems are considered, which contain thousands of categories at the lowest level. There are numerous models whose quality is also influenced by the field of application. Consequently, the development of suitable models in practice requires an intensive evaluation and optimization. Furthermore, the models must be continuously monitored in their practical application.

Often an evaluation is only carried out at the lowest level of a classification system. However, for the systematic evaluation of models it is advantageous to consider the entire hierarchical structure of a classification system.

Our proposed workbench could be a step to optimize the model by explicitly analyzing the results with respect to the special structure of an HCS.

Several software options for big data visualization exist for commercial use. Usually, a company using an ERP system like SAP², builds its own specific add-ons using what options the ERP system provides for customization. A visual exploratory analysis of different product classifications is only possible to a limited extent.

Beside full-fledged software solutions, numerous frameworks for building custom applications exist. Dash Open Source³ is a Python-based open source framework for building machine learning and data science web apps. While it supports integrated classification the user has to program the actual application for himself. The focus of the JavaScript framework D3.js⁴ lies on complex interactive web graphics applications. It can also be used with languages like R and Python, but does not provide classification algorithms. As with Dash, the user is required to implement what he wants on his own.

In sum we found no ready-to-install software dedicated to visualizing and manipulating hierarchical product data, especially not with added machine learning based classification capabilities. We see the express need of retail companies for such a workbench, though, which in our opinion is crucial for the successful utilization and high acceptance of automatic classification in practice.

² 1 <https://www.sap.com/germany/industries/retail.html> (last accessed 03.08.2020)

³ 3 <https://plotly.com/dash/> (last accessed 03.08.2020)

⁴ 4 <https://d3js.org/> (last accessed 03.08.2020)

3 Hierarchical classification workbench

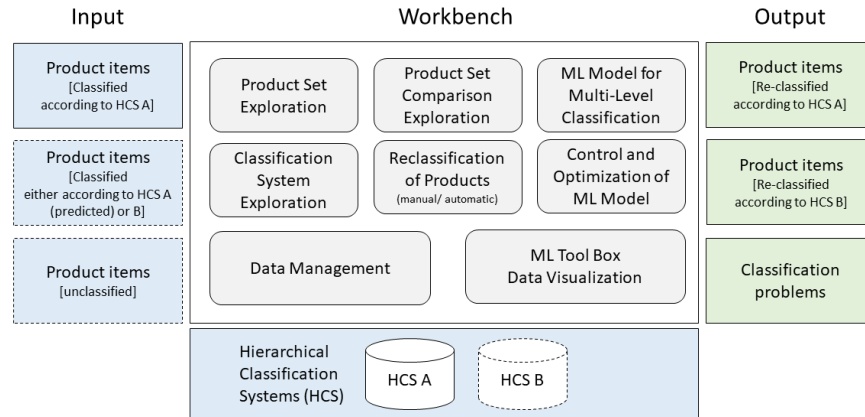


Fig. 1. Schematic overview of the proposed workbench

The main functional modules of the workbench are shown in figure 1. As input the workbench needs at least one hierarchical classification system (HCS A) and a set of classified product data. If the comparison of classification systems is desired a second classification system (HCS B) is also a required input. Using ML methods, the workbench outputs the product items reclassified according either to HCS A or HCS B. If the quality of a classification model is to be evaluated, HCS A and HCS B are identical. At the moment the (re)classification modules have been developed as a stand-alone software, but not yet been integrated into the workbench prototype.

3.1 Data exploration functions

Classification system and product set exploration

Our workbench utilizes diagrams for visual analysis to detect misclassified products in a data set. First basic information about the HCS itself is provided regarding the number of hierarchy levels, the amount of different categories for each level and the number of products in the data set. If more than one HCS is given, the number of products belonging to the same category in both HCSs is determined. In a second step, the distribution of the categories on each hierarchy level can be investigated. To this end, an interactive dendrogram picturing the hierarchical structure of the HCS can be generated in order to provide a first visual aid for further analysis (fig.2). With the help of tree maps and heat maps the distribution of products to the HCS can be visualized. This allows to identify unbalanced categories of the HCS regarding the data set. If a specific category contains only a few products compared to other categories, this could cause the corresponding training data for the subsequent model training to be unbalanced. As a consequence, the prediction quality for this category will degrade. Using our work-

bench, these imbalances can be detected and mitigating actions taken before the prediction model is trained. Additional product data from external data pools could be requested or oversampling algorithms employed to counter the imbalance of the specific categories.

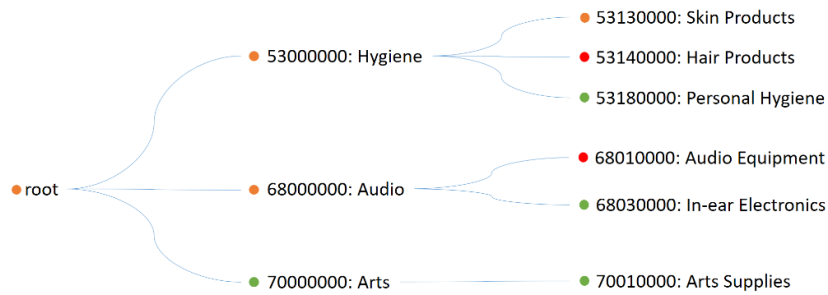


Fig. 2. Section of an interactive dendrogram of a hierarchical product classification as provided by the workbench. Dots denote categories where all (green), some (orange) or no (red) sub-categories contain at least one product.

Product set comparison exploration

The workbench supplies a comparison exploration module to compare two different HCSs in order to detect and identify misclassified products (fig. 3). It has to be kept in mind, that either of the two HCSs may contain misclassified products and that a seemingly wrong predicted classification according to HCS A could also mean that the classification of HCS B is already wrong. These errors can and do happen with real-world data. For example, the classification of a product p according to HCS B is suspect if a high proportion of the products that belong to the same category as p in HCS A belong to another category in HCS B. To identify these products and to check their classification is important since the ML model is based on the assumption that all products in the training data set are correctly classified. A model based on erroneous data usually delivers inferior results.



Fig. 3. Graphical user interface showing the comparison of two classifications based on a heat map. The hierarchical level can be selected by the user.

3.2 Classification functions

ML Model for multi-level classification

One of the main advantages of the workbench is the integrated module for interactive multi-level classification of products. It is planned that the user will be able to call the module from the dendrogram view of the workbench and classify a product on the fly. Depending on the company-specific classification system the user employs, different pre-trained models can be added.

The prototype of the workbench has an integrated machine learning model for multi-level classification of products according to the Global Product Classification (GPC) standard, based on the results of [10]. To measure the model performance we use the weighted metrics precision (P_{wg}), recall (R_{wg}) and F_1 score (F_{1wg}). These weighted versions explicitly take the multi-class structure of a product data set into account by computing a weighted average of the respective micro and macro metrics. This makes it possible to calculate the metrics at different hierarchical levels and to identify main categories, categories or sub-categories where the quality of the model is insufficient.

Control and Optimization of ML models

The usual quality metrics model evaluation are accuracy, precision, recall and F_1 -score, which only assess the quality, but do not provide explanations why the model performs as it does. In multi-level classification a confusion matrix is often used for deeper model evaluation. However, it only shows the distribution of the products of a given category to all possible categories and does not take into account that the categories themselves are ordered hierarchically. If a product is not predicted correctly on a given category, it nonetheless may be correct regarding the next higher hierarchical category level. This would be a less grave misclassification than predicting a wrong category within an also wrong overlying category. We define the severity of such an error according to the hierarchy level on which the error first starts. For GPC with four hierarchy levels, a product with a wrongly predicted category on brick level, but correct on class level is called a “level 4 error”. If the product classification is also wrong on class level, but correct on family level it is a “level 3 error” and so on (fig. 4).

In addition our workbench offers heat maps to show the products and their predicted classification embedded in the overall hierarchy. The user has the possibility to list all misclassified products and to correct their classification manually. Figure 5 shows an example where the product “Breaded cauliflower” belonging to the category “Vegetables – Prepared/Processed (Frozen)” has been wrongly assigned to the category “Vegetables – Unprepared/Unprocessed (Frozen)”. The workbench shows this error by colouring the product in red. Following the hierarchy levels up one can see that the classification has already failed on family level and is thus a level 2 error.

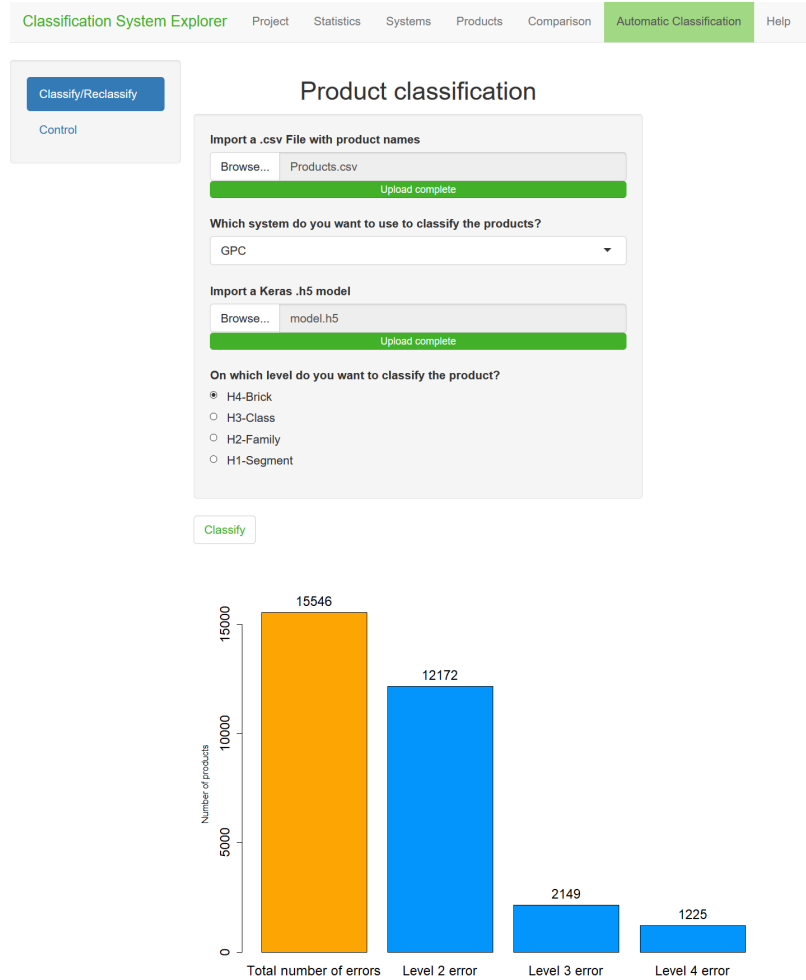


Fig. 4. Prototype of the user interface for bulk product classification. (The number of wrongly classified products regarding the hierarchy level on which the products become correctly classified are displayed.)

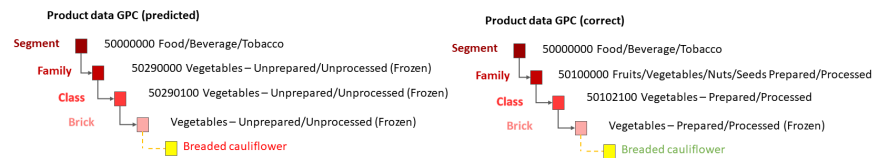


Fig. 5. Predicted / correct GPC brick (level 2 error)

(Re)classification of products

The option to (re)classify a product by its name is one of the main advantages of the proposed workbench. Usually a company uses not only its own custom classification system but is forced to handle also the classification systems its suppliers use. Importing product data from a supplier's data pool necessitates a reclassification into the company classification system. This is often done manually or using mapping tables, both error-prone or time-consuming methods. The proposed workbench offers the option to classify a product according to a given classification system by entering the product name into a field. The integrated ML model returns a suggestion for the appropriate classification regarding the desired classification system. A prototype of this functionality is shown in figure 6.

The screenshot shows the 'Automatic Classification' interface. The main form is titled 'Product classification' and contains the following elements:

- Navigation:** Classification System Explorer, Project, Statistics, Systems, Products, Comparison, Automatic Classification (active), Help.
- Control Panel:** Classify/Reclassify (button), Control (text).
- Form Fields:**
 - Enter a product name:** Input field containing 'Pizza Siziliana Classic'.
 - Which system do you want to use to classify the product?:** Dropdown menu showing 'GPC'.
 - Import a Keras .h5 model:** File selection area showing 'model.h5' and a green 'Upload complete' button.
 - On which level do you want to classify the product?:** Radio button options: H4-Brick (selected), H3-Class, H2-Family, H1-Segment.
- Classify:** A green button to execute the classification.
- TOP 3 Suggestions:** A table showing the top three results with their respective IDs.

Highest probability	Second highest probability	Third highest probability
10000248	10000249	10000598

Fig. 6. Prototype of the user interface for on-the-fly classification for single products using the product name providing the three predictions with highest probability (Top 3)

4 Conclusion

The proposed workbench is especially designed to support the exploratory analysis of product data classified according to a given hierarchical classification system, the comparison of different product classifications to identify errors and inconsistencies and the evaluation of machine learning models for the automatic classification of products. It supports the whole machine learning pipeline for product classification in an integrated environment. We expect that the usage of the workbench will significantly accelerate and simplify the model evaluation process in practice.

Preliminary evaluation in a major German retail company has been met with success. The workbench was tested on 40.000 products classified according to both a company specific classification system and to GPC managed by an ERP system. Especially, the use of heat maps to compare the product classification at different levels showed hitherto undetected errors in the current classification and was greatly appreciated. The visualization with interactive dendrograms also served to analyze problems in the existing data base. All these features were not available in the company's ERP system but were much valued by the customer.

5 Outlook

Our next step is to integrate all of the already developed ML components into the workbench and to add the option to interactively assign the correct classification to the wrongly classified product directly in the training data set using the dendrogram view. After error correction, the machine learning model can be retrained to improve the classification quality. Using the workbench, users can then monitor the results of the automatic classification. In doing so we hope to increase the acceptance of machine learning methods for automatic product classification in practical applications.

We also plan to evaluate the complete workbench thoroughly with a another major retail company in order to ready it for actual deployment in a productive environment.

Acknowledgements

Part of the research presented in this paper was funded by the German Ministry of Education and Research under grant FKZ 01|S18018.

References

1. Hepp, M., Leukel, J., Schmitz, V.: A quantitative analysis of product categorization standards: content, coverage, and maintenance of eCI@ss, UNSPSC, eOTD, and the RosettaNet Technical Dictionary, Knowledge and Information Systems 13.1, pp. 77–114 (2007)
2. GS 1 homepage, <https://www.gs1-germany.de/gsl-standards/klassifikation/produktklassifikation-gpc/> (last accessed 03.08.2020)

3. Shearer C.: The CRISP-DM model: the new blueprint for data mining, *J Data Warehousing* (2000); 5:13—22
4. Microsoft: Team Data Science Process (TDSP). <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview> (last accessed 03.08.2020)
5. Chavaltada, C., Pasupa, K., Haroon, D.R.: A Comparative Study of Machine Learning Techniques for Automatic Product Categorisation. In: *Proceedings of Advances in Neural Networks - ISNN* (2017)
6. Sun, C., Rampalli, N., Yang, F., Doan, A.. (2014) Chimera: Large-Scale Classification using Machine Learning, Rules, and Crowdsourcing. *Proceedings of the VLDB Endowment*, Vol. 7, No. 13
7. Ha, J.W., H. Pyo, Kim, J. (2016). Large-scale item categorization in e-commerce using multiple recurrent neural networks. *Proceedings of the 22nd ACM SIGKDD*
8. Ding, Y., M. Korotkiy, B. Omelayenko, V. Kartseva, V. Zykov, M. Klein, E. Schulten and D. Fensel (2002). GoldenBullet: Automated Classification of Product Data in E-commerce. *Proceedings of BIS 2002*
9. Cevahir, A., Murakami, K.: Large-scale Multi-class and Hierarchical Product Categorization for an E-commerce Giant. In: *Proceedings of COLING 2016*, pp. 525–535 (2016)
10. Allweyer, O., Schorr, C., Krieger, R., Mohr, A.: Product classification based on partially abbreviated product names in retail. In: *9th International Conference on Data Science, Technology and Applications, Online* (2020)