Accounting Financial Transactions Across Companies and Corporate Sectors

Rasmus Kær Jørgensen[1,2] & Christian Igel [1]

[1] Department of Computer Science, University of Copenhagen
[2] PricewaterhouseCoopers (PwC), Denmark

Contact Information:
Universitetsparken 1
DK-2100 Copenhagen Ø

Phone: +45 20837566
Email: rasmuskj@di.ku.dk

Abstract

Accounting is essential for every company. It is required by international corporate law and characterizes a structured overview of the company's finances, which is reported in the financial statement of the company. An important initial step in accounting is mapping financial transfers to the corresponding accounts. Semi-automatic systems for this task typically rely on company-specific, rule-based systems. In contrast, this study presents a transaction classification system that generalizes across companies, even across different corporate sectors and to new companies. The first step is standardization, where the different account structures of companies are mapped to a unified chart of accounts. The second step involves non-standard feature engineering. We show how text fields in transactions and company registries can be utilized by applying character-level word embeddings. In comparison to simple pattern matching, this induces a proper neighborhood relation between text fields that fosters generalization. The study considered 10,354 real-world transactions from 44 small to medium-sized companies from 28 different sectors that could be mapped to 54 of the accounts in the unified chart. A single classifier trained on historic data of all companies could classify new unseen transactions of these companies with an accuracy of more than 80%. An ablation study showed the importance of including the text fields for achieving this high accuracy. Training on 43 companies and testing on the remaining one achieved an average performance of 64.62%. When we restricted our study to the single sector for which we had the most training data, this rate increased to nearly 70%. That is, we showed that it is possible to classify transactions for a new company without historic data with superior performance compared to contemporary operations.

Task and Data

This study presents a machine learning based system for financial firms that learns to classify transactions across companies and corporate sectors. This allows for the mapping of transactions from new companies for which little or even no historical data exists.

No.	Date	Transaction Text	Amount	Account Code
1	17/10-2018	DK4095 item xzy 777	799	100 (Sales)
2	08/11-2018	BS-123 Housing Company	-11,943	210 (Rent)
3	11/11-2018	DK7045 fabric store A/S	-31,586	230 (Supplies)

Table 1: Transactions and their mapping to account codes.

Data 10,354 real-world transactions collected from 44 Danish small to medium-sized companies from 28 different sectors.

Unifying Charts of Accounts The first challenge when designing a classifier is that the chart of accounts is not standardized among the companies. A company is free to design a list of accounts as long as it follows regulations and laws. This difference among companies leads to different output spaces, which complicates the use of machine learning. We map the individual charts of accounts to an *Unified Chart of Accounts*. We are hence considering a 54-class classification problem instead of a 3,549-class problem.

Table 2: Simplified chart of accounts.

Simplified Chart of Accounts	
Account Code	*Account Description*
Revenues (100-199)	
100	Sales
…	…
Expenses (200-299)	
210	Rent
230	Supplies
250	Salaries
…	…
Assets (300-399)	
330	Inventory
360	Equipment
…	…
Liabilities (400-499)	
400	Tax
440	Interest
…	…
Equity (500-599)	
500	Capital
…	…

Feature Engineering The second challenge is to find a representation of the input that allows for highly accurate classification and generalization across companies. We show how to engineer suitable features by mimicking the use of background knowledge by accounting and bookkeeping clerks.

We advanced from 3 features (date, text and amount) to 28 features per financial transactions, see table 5. Besides including publicly available data and deriving basic features as *Distance, Payment Type and Difference in Days*, we utilized a **general-purpose word embedding with character-level features** to create a low dimensional representation of the unstructured free text fields: *Transaction Text* and *Sector Descriptions*.

Experiments and Results

Classification of New Company Transaction models the real-world scenario of classifying financial transactions of new, previously unseen companies when no historical labeled transactions are at the disposal. Generalization across companies and corporate sectors can be viewed as a multi-source domain adaptation problem.

- Settings B.I and B.II address this task by an *aggressive approach*, which combines all available data – all companies – into one source domain.
- Setting B.III follows a *selective approach* that only combines selected sources, the within-sector companies, into a source domain.

The main questions we wanted to answer were: Can a classifier built using our approach generalize to a new company? Can it generalize to a new corporate sector? Does the task become easier if we consider less but more homogeneous data?

	Leave-One-Company-Out	Leave-One-Sector-Out	Leave-One-Pharmacy-Out
Experiment	B.I	B.II	B.III
Number of Sectors/ Companies	44 companies	28 sectors	6 pharmacies
Source Domain: Target Domain	43 companies: 1 company	27 sectors : 1 sector	5 pharmacies : 1 pharmacy
Overall Correctly Classified	6,691	5,378	3,257
Overall Incorrectly Classified	3,663	4,976	1,399
Overall Classification Accuracy	**64.62%**	**51.94%**	**69.95%**

Table 3: Results of Classification of New Company Transaction.

Experimental finding:

- Good generalization – one or more within-sector companies in the source domain.
- Lower generalization – no within-sector companies in the source domain.
- Specific sectors are harder to generalize to than others due to individual peculiarities.
- Achieving higher generalization is linked to the similarities between source and target companies.

Transaction Time Series models the scenario in which we have labeled historical data and want to predict the accounts of future transactions. For each company, we ordered financial transactions by time and partitioned by using the first 70% for training and the subsequent 30% for testing.

As we want to build **a single classifier for all companies**, we combine the training and test data of all 44 companies to a joint training and a single test set. In this case, the classifier has to generalize over 28 different corporate sectors.

We consider five scenarios in the general setting A:

A.I All features and all companies were included.

A.II All features were provided, but only the 6 pharmacies were considered.

A.III Only the Transaction Text features were provided and all companies were included.

A.IV All features except the Transaction Text features and all companies were included.

A.V All text field features were excluded and all companies were considered.

Data Setting	All Companies	Subset of data: Pharmacy Sector	All Companies	All Companies	All Companies
Feature Setting	Complete Feature Set	Complete Feature Set	Only Transaction Text	Without Transaction Text	Without Text features
Experiment	A.I	A.II	A.III	A.IV	A.V
Majority Class	23.82%	18.51%	23.82%	23.82%	23.82%
k-Nearest Neighbor	37.74%	47.35%	67.16%	37.70%	37.48%
Logistic Regression	44.99%	26.83%	36.02%	47.53%	29.27%
Random Forest (OOB)	87.72%	88.75%	78.61%	82.21%	69.25%
Random Forest	**80.76%**	**81.50%**	**69.63%**	**78.09%**	**56.72%**

Table 4: Results of Transaction Time Series experiments. All results refer to (average) test errors, only for the random forest; additionally, the out-of-bag error on the training set is reported. The k-nearest neighbor results always refer to the best test accuracy for $k \in \{1, \ldots, 10\}$.

Experimental finding:

- The ablation study exhibits a strong discriminative power of the extracted text features.
- An accuracy of more than 80% makes an adequate basis for industrial adoption.
- We see a marginally higher performance for the simpler scenario in which the classifier has to generalize only over companies of a single corporate sector.

Discussion and Conclusions

We have presented a novel system for supporting financial firms in mapping financial transactions to the corresponding accounts. It may be regarded as a significant step in the development of semi-automatic or even fully automatic financial software because it learns a classifier that generalizes across companies and even to new companies – in contrast to the company-specific classifiers used in industrial systems.

With a *Unified Chart of Accounts* defining the label space, we regard test accuracies over 80% when classifying new transactions from known companies and almost 65% for transactions from new companies without historical data.

The findings of this study are rather applicable to transaction classification systems in general:

1. The system could approximately **save 30-50% time and cost** in comparison to current operations.
2. Simply **retraining the classifiers to update the system** is seen as an advantage compared to rule-based systems where outdated mappings require humans to review the whole rule-base.
3. Through a **standardized label space** and carefully designed input features, we presented a solution to the *cold-start problem* by classifying transactions of a new company without historic data. Today a new company means manually mapping all financial transactions to account codes and constructing rules. We progress from 0% (all manual) to an average accuracy of almost 65% for new companies across all 28 corporate sectors. When we restricted the study to a single corporate sector, we reached almost 70%.
4. The ablation experiments showed the **discriminative power of the engineered features**. Our main focus was on character-level word embeddings trained on a general domain corpus to create a low dimensional vector space representation of the free text fields in the transactions. The embedding turned out to be highly important for the random forest and the ablation studies. Not providing the text features led to a performance drop of more than 20 percentage points.

Table 5: Feature Importance (A.I).

Importance	**Feature**
0.084126	External Sector Text 1st Comp.
0.076574	Transaction Text 1st Comp.
0.071125	External Sector Text 2nd Comp.
0.065961	External Sector Text 4th Comp.
0.063809	External Sector Text 3rd Comp.
0.062806	External Sector Text 5th Comp.
0.056514	Transaction Text 2nd Comp.
0.053486	Transaction Text 4th Comp.
0.052116	Transaction Text 3rd Comp.
0.047529	Amount
0.045169	Transaction Text 5th Comp.
0.038887	Distance
0.036161	VAT Amount
0.027632	Subject Sector Text 1st Comp.
0.027476	Difference in Days
0.024845	External Business Entity
0.021755	Subject Sector Text 4th Comp.
0.021032	Payment Type
0.019611	Subject Sector Text 5th Comp.
0.018122	Subject Sector Text 2nd Comp.
0.018101	Subject Sector Text 3rd Comp.
0.017512	Issued Week
0.016711	Paid Week
0.011049	Subject Business Entity
0.006678	Paid Quarter
0.006569	Issued Quarter
0.005313	Paid Year
0.003329	Issued Year

It is widely acknowledged that machine learning based systems for accounting financial transactions will replace software systems relying on handcrafted rule-bases and will increase the degree of automation in this area – and the results of our study support this prediction.

Our system for classifying transactions is currently evaluated at one of the biggest international accounting firms.