Martin FRIDRICH, PhD Candidate
E-mail: fridrichmartin@yahoo.com
Department of Informatics
Faculty of Business and Management
Brno University of Technology

UNDERSTANDING CUSTOMER CHURN PREDICTION RESEARCH WITH STRUCTURAL TOPIC MODELS

Abstract. *Customer churn prediction is showing a growth in attention from both researchers and practitioners, creating a vast body of scientific works while being recognized as an indispensable tool of corporate retention activities. Thus, we aim to demonstrate the potential of structural topic models to navigate through the research articles and to identify essential themes and trends within the field of customer defection prediction. We apply a modified modeling procedure to journal articles focused on customer churn. As a result, the structural model of 38 topics is formed and examined considering topic prevalence, its changes over time, and the scientific impact (citations). We see prevailing themes tackling broad perspectives such as modeling, evaluation, and performance metrics. Furthermore, we recognize a slow decline in business & marketing aspects of churn prediction coupled with rising of more nuanced topics. At last, we discuss possible future steps in topic modeling within the domain.*

Keywords*: Customer Churn Prediction, Natural Language Processing, Topic Modeling.*

JEL Classification: M31, C60

1. Introduction

Over the past decades, both marketing researchers and practitioners acknowledged the importance of customer retention as a critical part of customer relationship management (CRM) aimed at the surge of total customer value. Retention management is generally reported as an enterprise priority; however, a considerable part of top executives is not satisfied with the ability to aid retention ambition (Daunis & Iwan, 2014) , nor seem to be the customers (Handley, 2013). We come to believe this dichotomy fuels thriving interest in the research fields of retention management and churn prediction. With the rapid growth in academia works, it becomes inevitably challenging to identify relevant tendencies and patterns.

DOI: 10.24818/18423264/54.4.20.19

Hence, the paper aims to demonstrate the potential of structural topic models through exploratory analysis of an ample body of works in the field. Three suitable perspectives are examined (1) topic prevalence in customer churn prediction, (2) changes in topic prevalence over time, (3) identification of highly cited topics. The paper does not aim to present an exhaustive analysis of the whole scientific domain, rather than a reproducible study based on available data, which limits the generalization of the findings.

2. Topic modeling

In the field of natural language processing (NLP), topic modeling is commonly recognized as a generative statistical procedure, which allows similarity amongst a set of observations (documents) to be partly explained by unobserved structural groups (topics). It enables organizing, summarizing, annotating, and understanding a vast collection of textual information. The task of topic modeling can be tackled with various algorithms such as latent semantic analysis, latent Dirichlet allocation (Blei et al., 2003; Blei, 2012), correlated topic model, structural topic model (Roberts et al., 2014; Roberts et al., 2019), and others. We focus solely on the structural topic method, as it allows us to incorporate covariates (document metadata) into the topic model.

The applications of structural topic models intersect with fields of economics, finance, political science (Cerchiello & Nicola, 2018;Rothschild et al., 2019; Shirokanova & Silyutina, 2018), education (Reich et al., 2015), new media (Rodriguez & Storer, 2019), and others (Kuhn, 2018; Zafari & Ekin, 2019). However, there is a limited number of articles dedicated to a meta-analysis of academic research (i.e. Bohr & Dunlap, 2017). Thus, we aim to examine the topic model capability through an investigation of an extensive body of works in the field of customer churn prediction.

2.1. Latent Dirichlet allocation

Latent Dirichlet allocation (LDA) is the most prominent non-linear generative probability topic model. The assumed process of document formation consists of (1) generating topics as distributions over fixed vocabulary, (2) choosing a random topic distribution for each document, and (3) selecting a random topic and corresponding vocabulary distribution for each word in a document. The generative process determines a joint probability distribution over observed (words in documents) and latent random variables (topics). The resulting joint probability distribution is used to compute the posterior distribution of latent random variables given observed ones.

We can describe the posterior distribution formally with the following notation. Let us have topics $\beta_{1:K}$, where each β_k is a vocabulary distribution. The topic mixture for document d is defined as θ_d. The topic label for the document d

DOI: 10.24818/18423264/54.4.20.19

is then z_d, the topic assignment for the n word in document d is $z_{d,n}$. The seen words for document d are f_d, n word in document d is $f_{d,n}$. Finally, the posterior distribution can be outlined as follows (Blei et al., 2003):

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D} | f_{1:D}) = \frac{\prod_{i=1}^{K} p(\beta_i) \prod_{i=1}^{D} p(\theta_i) (\prod_{n=1}^{N} p(z_{d,n} | \theta_d) p(f_{d,n} | \beta_{1:K}, z_{d,n}))}{p(f_{1:D})}. \tag{1}$$

The numerator represents the joint distribution of latent variables and defines several dependencies amongst underlying elements. It is straightforward to compute. The denominator, on the other hand, stands for the probability of witnessing the observed corpus under any topic model. In practice, it is infeasible to compute. Thus, the posterior distribution is usually approximated with sampling or variational algorithms (Blei et al., 2012).

2.2. Structural topic model

The structural topic model (STM), like other generative probability topic models, is based on a random generative process and use of observed properties to estimate internal parameters of the model. However, the generative process of STM allows for topical prevalence (document-topic distribution) and topical content (topic-word distribution) to be a function of external covariates (document metadata). The structural modeling approach consists of three elements, specifically (1) topic-word allocation as a function of topic prevalence covariates, (2) term-frequency estimation as a function of topic content covariates, and (3) an observational model, which reconciles the two preceding sources of variation to compose the words in each document (Roberts et al., 2014).

Formally, we can describe the numerator of posterior distribution with expanding LDA notation. X and Y denote matrices of topical prevalence and content covariates. Γ and Κ represent a coefficient matrix of the prevalence and content model. Σ is a hyperparameter affecting θ distribution (Roberts et al., 2014).

$$p(\eta, z, \kappa, \gamma, \Sigma | f, X, Y) \propto \left(\prod_{d=1}^{D} Normal(\eta_d | X_d \gamma, \Sigma) \left(\prod_{n=1}^{N} Multinomial(z_{n,d} | \theta_d)\, Multinomial(f_d | \beta_{d,k=z_{d,n}}) \right) \right) \times \prod p(\mathrm{K}) \prod p(\Gamma) \tag{2}$$

Evident contrast between LDA and STM posterior distributions is that STM enables for differentiation of prevalence and vocabulary distribution across all documents. Similarly, to LDA, the posterior distribution is intractable. Thus, Roberts et al. (2019) suggest approximation with nonconjugated variational expectation maximization (EM).

3. Research methodology

This section paints the elements of the research methodology. Concretely, we outline the data collection process, synthetic performance metrics used for topic models, and specifics of experimental design.

3.1. Dataset

The research aims at peer-reviewed articles published between the years 2009 and 2019, in journals indexed in scientific databases Web of Science or Scopus and written in the English language. Other types of documents, such as conference proceedings, book chapters, or thesis, are omitted. The domain of interest is defined with the use of Boolean search and relevant keywords such as customer churn prediction, customer churn, or churn prediction. For each article, we obtain a digital object identifier (DOI), authors, title, abstract, publication year, and citation count.

3.2. Performance metrics

Exclusivity – For each model and topic, the FREX metric is used to evaluate quality concerning both exclusivity and word frequency. Balancing these two perspectives is necessary as frequent words are often not exclusive to a specific topic, while singular words are not informative. To prevent high rank along one dimension to compensate for low rank in another. Work of Airoldi & Bischof (2017) suggests reconciling the aspects with harmonic mean. For f as a word in topic k, the metric is defined as follows:

$$FREX = \sum_{k=1}^{K}\sum_{f=1}^{F}\left(\frac{w}{F_{f,k}} + \frac{1-w}{E_{f,k}}\right)^{-1}, \qquad (3)$$

where w stands for weight; F is the frequency score given by empirical cumulative distribution function (ECDF) of the word f in k topic distribution; exclusivity E comes from the conditional probability of observing topic k given word f and respective ECDF. High exclusivity is easily attained by having a large number of narrow topics. Thus, we follow the suggestion of Roberts et al. (2019) and evaluate topic quality not only with exclusivity but also with semantic coherence.

DOI: 10.24818/18423264/54.4.20.19

Semantic coherence – The idea of semantic coherence comes from the concept of pointwise mutual information and assumes highly probable words incoherent topics should co-occur within the same document. Furthermore, Mimmo et al. (2011) suggest that the metric at hand can approximate the human perception of topic quality. To describe the metric formally, let us have words f_i and f_j, then for the set of M most probable words in topic k, the semantic coherence for the topic k is defined as:

$$C_k = \sum_{i=2}^{M} \sum_{j=1}^{i-1} \log\left(\frac{D(f_i, f_j) + 1}{D(f_j)}\right) \tag{4}$$

where $D(f_i, f_j)$ represents co-occurrence of words f_i and f_j; similarly, $D(f_j)$ stands for the occurrence of a word f_j. As accurately noted by Roberts et al. (2019), semantic coherence tends to be high for models with a small number of topics dominated by common words. The contrasting characteristics of both exclusivity and semantic coherence are exploited in the quantitative part of model evaluation.

3.3. Experimental design and implementation

The application of structural topic modeling techniques is explored through an iterative topic modeling pipeline, which consists of four main blocks (1) data acquisition, (2) data processing, (3) model estimation & evaluation, and (4) model understanding. Figure 1 denotes the relationship amongst the blocks. We implement the pipeline in the R language for statistical programming, specifically in Microsoft R Open 3.5.3.

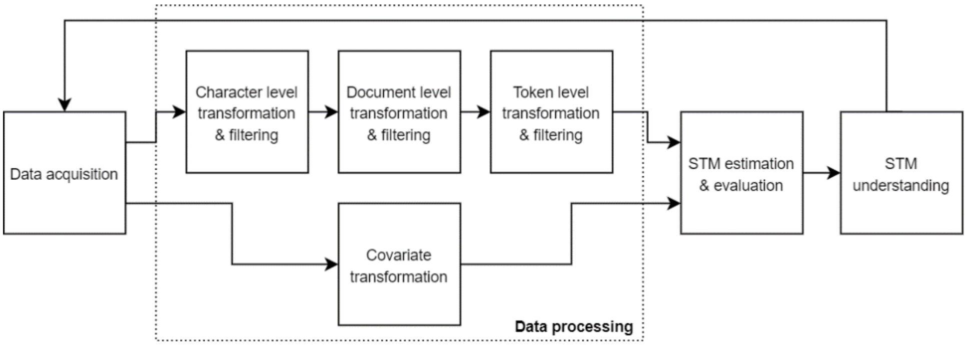
Data acquisition
Character level transformation & filtering
Document level transformation & filtering
Token level transformation & filtering
Covariate transformation
Data processing
STM estimation & evaluation
STM understanding

Figure 1. Topic modeling pipeline

Data acquisition – The appropriate documents and topic prevalence covariates are retrieved from Web of Science and Scopus databases. We unify data representation across exported files with respect to encoding, naming conventions, and DOIs.

DOI: 10.24818/18423264/54.4.20.19

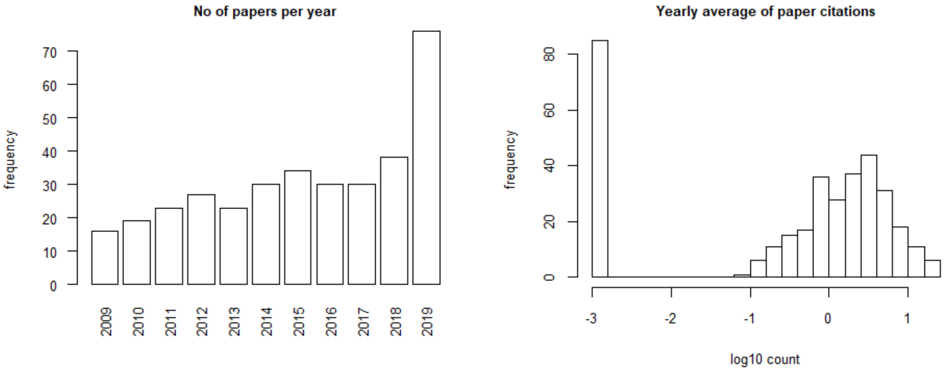

Figure 2. Frequency distribution of processed topic prevalence covariates

Data processing – We introduce two independent branches of processing to reflect the nature of the input data, (1) textual data transformation and filtering, and (2) covariate transformation.

The first branch consists of three subsequent levels, (1.1) character level - we concatenate article title and abstract, convert all characters to lowercase, remove tags, special characters, numbers, and multiple spaces, (1.2) document level - duplicates are eliminated, a language of the text is estimated, only abstracts written in English are selected, document satisfactory character length is chosen, then a restrained variant of automatic spell correction is applied, (1.3) token level - we use UDPipe R package (Straka & Straková, 2017) with English-ParTUT linguistic model for tokenization, parsing, lemmatization, and tagging. Resulting lemmas are filtered on character length, observed frequency, and estimated universal part of speech (UPOS), remaining polluting artifacts are removed with stop-words. The resulting dataset consists of 346 journal abstracts, 709 token corpora.

The second branch deals with covariate transformation. Namely, the annual component from the publication date is extracted. To account for the fact that recent articles have less time to gain citations, we compute a yearly average of paper citations. In Figure 2., we recognize growth in a number of published papers per year, with 2019 being the most fruitful. However, approximately one-fourth of the texts have not been cited. Both features directly affect the prevalence distribution of the structural topic model.

DOI: 10.24818/18423264/54.4.20.19

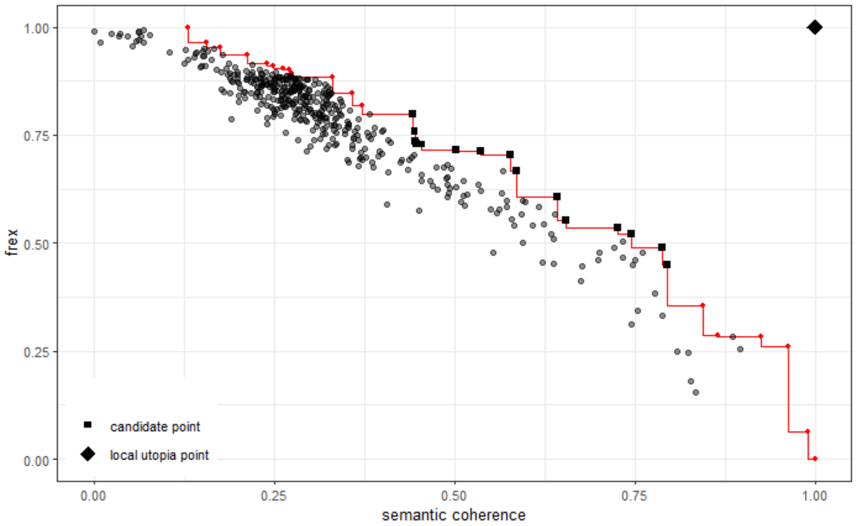

Figure 3. STM computational evaluation – exclusivity and coherence trade-off

STM estimation & evaluation – STM estimation block ingests processed texts, covariates, and token corpus. Subsequently, structural topic models are built, evaluated, and selected. To tackle the problem of model search across numbers of topics, we adopt a hybrid approach based on (1) computational evaluation and (2) human judgment & domain expertise.

Computational evaluation (1) modifies the selection procedure proposed in Roberts at al. (2019). For each number of topics, 100 STM models are randomly initialized with the LDA during the cast-net stage (~ 5 EM iterations). Subsequently, 5 STM models with the highest marginal likelihood bound are estimated with additional ~ 200 EM iterations. The model search is conducted in the range of 10 to 100 topics resulting in 455 STM models. Their quality is then assessed with exclusivity and semantic coherence. We approach the assessment as a no-preference multi-objective problem with the local utopia point. Non-dominated observations closest to the utopia point are picked for a human reader to examine. The candidate subset of STM models consists of 15 cases where 13 to 50 topics are formed. To illustrate the evaluation procedure, we construct Figure 3; the dimensions are lin-scaled for comparability, the Pareto boundary is drawn in red. Finally, human judgment & domain expertise (2) are employed to evaluate the interpretability of candidate models.

Model understanding – After an STM model is estimated, it must be interpreted by a human reader. We follow the strategy proposed by Roberts et al. (2019) and break down the analysis into the subsequent steps, (1) identifying

DOI: 10.24818/18423264/54.4.20.19

documents and tokens associated with particular topics, labeling topics, (2) assessing links amongst topics with correlation analysis, and (3) determining relationships between topics and covariates. The method is used to explore, evaluate, and select a model from the candidate subset with the use of human comprehension & knowledge. We pick a model forming 38 topics with semantic coherence ~ -106.2 and FREX ~ 9.74. Details of the analytical procedure are presented in the following paragraphs.

4. Results

In this section, we cover the understanding of the STM outcomes. Notably, we examine the topic prevalence and labels, review the significant associations amongst latent factors, and explore the relationships between covariates and expected topic prevalence.

4.1. Topic labels

We employ several tools to visualize and explore the model results. We begin with understanding topics through a collection of prominent lemmas and related documents. The association between lemmas and topics is ranked with the conditional probability of occurrence, lift, and FREX. The last two measures account for the observed frequency across the corpus, hence value more distinct lemmas. The link between relevant documents and associated topics is also studied.

Table 1. Articles with high expected prevalence towards the Social Networks topic

Topic label	Authors	Title	Expected topical prevalence
Social Networks	Al-Molhem et al. (2019)	Social network analysis in Telecom data	72.64 %
	Mitrovic et al. (2019)	tcc2vec: RFM-informed representation learning on call graphs for churn prediction	71.56 %
	Olivieira & Gama (2012)	An overview of social network analysis	68.20 %

To understand and label latent factors, we employ both topic-word and document-topic relationships. We illustrate the approach using a sample topic. The strongly associated documents with the respective topical prevalence are outlined in Table 1. We notice that prominent lemmas and article titles are considerably aligned (see Table 1 and Figure 4). Moreover, we remark that Al-Molhem et al. (2019) and Mitrovic et al. (2019) focus on understanding customer behavior in the telecommunication industry reconciling graph (social networks) and standard relational data representations (transactions). Oliviera & Gama (2012), on the other

DOI: 10.24818/18423264/54.4.20.19

hand, summarize best-suited graph properties and techniques for social network analysis. The abstract outlines the author's motivation and the article content as follows:

> *"Data mining is being increasingly applied to social networks. Two relevant reasons are the growing availability of large volumes of relational data, boosted by the proliferation of social media web sites, and the intuition that an individual's connections can yield richer information than his/her isolate attributes. This synergistic combination can show to be germane to a variety of applications such as churn prediction, fraud detection and marketing campaigns. This paper attempts to provide a general and succinct overview of the essentials of social network analysis for those interested in taking a first look at this area and oriented to use data mining in social networks."*

Therefore, it is concluded that the sample topic deals mostly with the social network analysis and ought to be perceived and labelled accordingly. The introduced procedure is exercised to recognize and label all topics.

As a result, 38 topics are labelled and depicted in Figure 4 (with the most frequent lemmas associated with each topic), ordered from the most to the least prevalent. We recognize groups of high-level latent factors based on different aspects of customer churn prediction such as business & marketing (CRM, Churn & Retention, Customer Loyalty), machine learning & technology (Model & Evaluation, Data Mining, Classification Performance, Neural Networks), industry (Telco Industry, E-commerce Industry, Banking Industry), blended (Economic Performance, Decision Factors) or others (Social Networks).

DOI: 10.24818/18423264/54.4.20.19

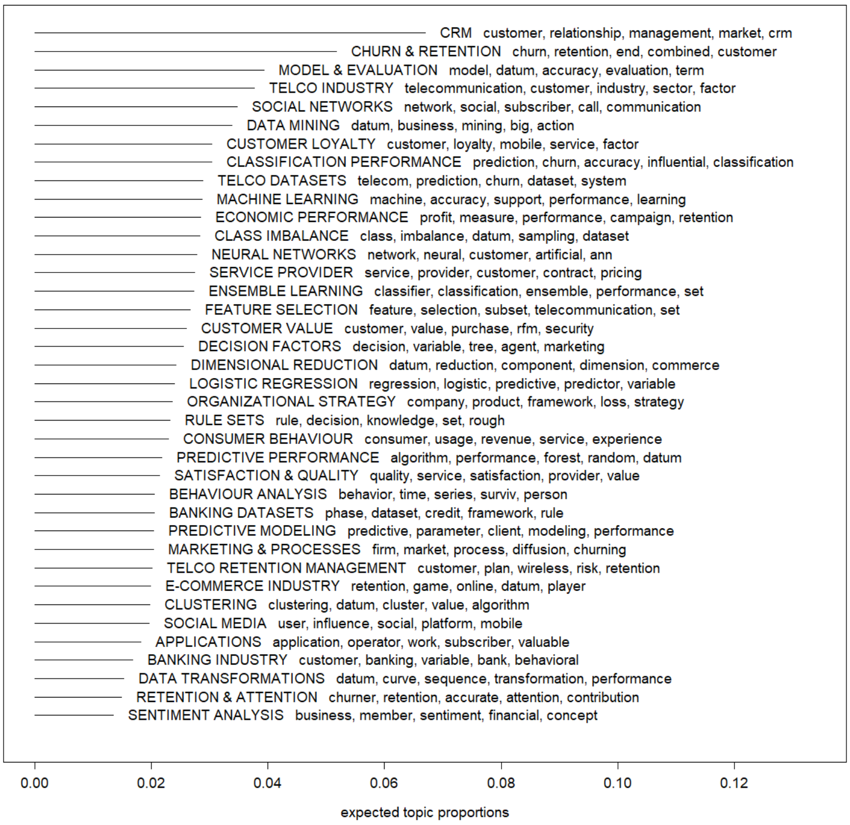

Figure 4. Topic labels, prevalence and prominent lemmas

Customer Relationship Management appears to be the most prevalent topic, accounting for ~ 6.7 % of topic proportions among article abstracts. This is to be expected, as customer churn prediction is an inherent part of the customer relationship management. Moreover, we acknowledge the problem context and its understanding as one of the critical determinants of a successful machine learning/research project. Other popular topics are Churn & Retention with ~ 5.2 %, and Model & Evaluation with ~ 3.9 % of topic proportions. The most prevalent factors cover broad perspectives of the customer churn prediction, such as problem context, modeling, evaluation, and metrics. Furthermore, we see variations in terminology, but this remark is not limited to the most wide-spread topics.

4.2. Correlations

In addition to topic labeling, STM allows us to analyze relationships between latent factors with positive Pearson's correlations. Unlike other authors, we cut-off topic links concerning asymptotic p-values and unadjusted alpha = 0.05.

The significant link then indicates that both topics are likely to be discussed within a document. The resulting correlation network of 26 topics and six communities is pictured in Figure 5. Node size indicates the topic prevalence, and its color displays community membership. Edge width reflects the correlation strength. We use a spring-based algorithm to make the graph layout and form communities with a fast-greedy modularity optimization algorithm. Orphan nodes are omitted for readability.

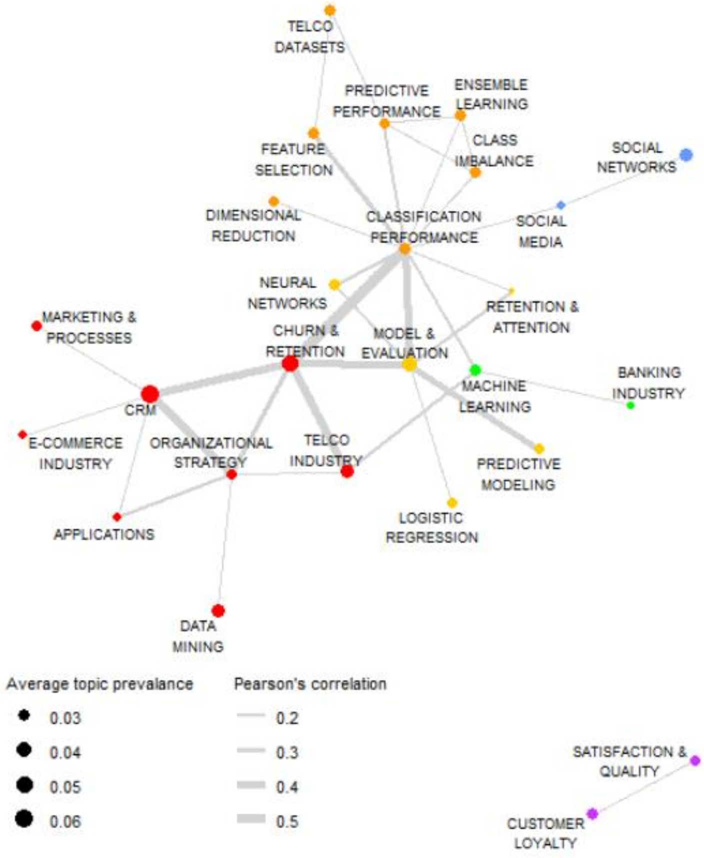

Figure 5. The correlation network of topics and respective communities

The depicted communities share the following concepts (from top to bottom): (1, orange with cliques) aspects of machine learning, (2, blue) social networks, (3, yellow) classification algorithms, (4, red with cliques) aspects of business/marketing, (5, green) machine learning, and (6, violet) customer loyalty/satisfaction. The resulting graph consists of two isolated components. The large network covers a broad range of topics, from business/marketing to social networks. Hence, the said component reflects most relationships amongst topics in customer churn prediction literature. Besides, we recognize groups of high-level latent factors: business/marketing (4), and machine learning (1, 3, 5). We expect blended topics (Economic Performance, Decision Factors) to serve as a bridge between said high-level graph components, which is not the case. Besides, the customer loyalty/satisfaction community (6) is not positively correlated with the business/marketing context of customer churn prediction research, nor is the social network's community (2). We suspect vast diversity in corpus structure amongst the topics to be responsible for outlined discrepancies.

4.3. Covariates

Furthermore, STM permits us to explore relationships between covariates and expected topic prevalence. We focus on variations in topic prevalence over time and prominent areas of research.

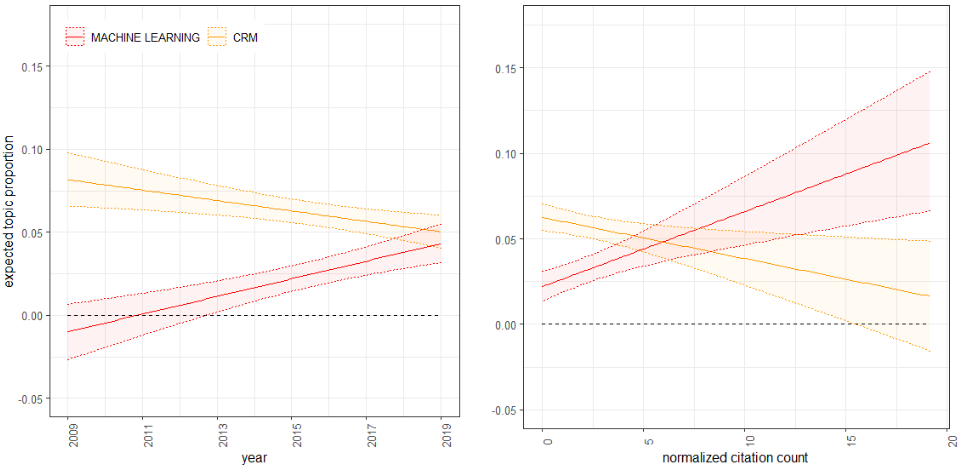

Figure 6. Expected topical prevalence by year (left) and normalized citation count (right), with 95 % confidence intervals

Figure 6 illustrates the contrast between Customer Relationship Management and Machine Learning. We observe a steady decline in the prevalence of CRM and recognize it as a sign of a domain maturity. Prevailing concepts such as CRM are well understood and documented; thus, the attention of researchers shifts towards more specific aspects of churn prediction. Journal articles with a high prevalence of CRM tend to have a lower number of citations. Machine Learning is, on the other hand, the fastest-growing topic. It may surpass CRM in the expected topic proportion by the end of 2020. Additionally, it is the most prominent area of research. Examples of works relevant to the topic include Sabbeh (2018), Vafeiadis et al. (2015), or Vijaya & Svasankar (2019) and concentrates around the technical aspects of churn modeling.

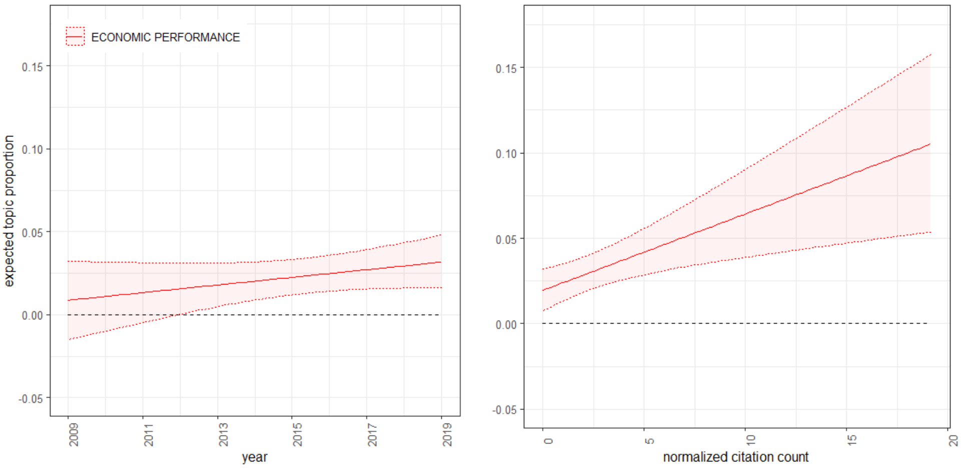

Figure 7. Expected topical prevalence by year (left) and normalized citation count (right), with 95 % confidence intervals

Figure 7 shows another emerging topic - Economic Performance. It displays continuous growth in the prevalence; moreover, associated texts are cited very often. Articles relevant to the topic introduce the economic context of customer churn to the modeling solution and may encompass works of Verbeke et al. (2012), Verbraken et al. (2012), and Garrido et al. (2018).

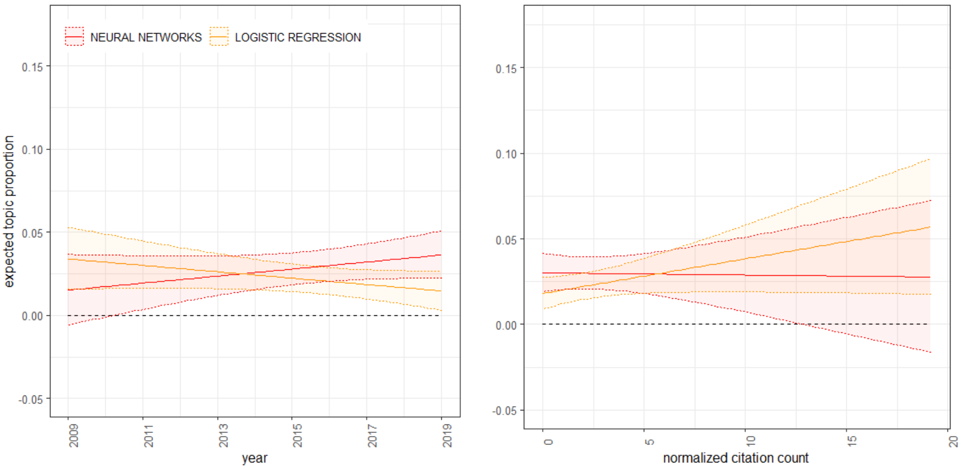

Figure 8. Expected topical prevalence by year (left) and normalized citation count (right), with 95 % confidence intervals

Figure 8 compares two latent factors formed around popular classification algorithms - Neural Networks and Logistic Regression. The prevalence of Neural Networks shows gradual growth, which is aligned with the increasing popularity of

DOI: 10.24818/18423264/54.4.20.19

the connectivism in the machine learning community. Yet, there does not seem to be a meaningful relationship between the topic prevalence and normalized citation count. Representative studies may include texts of Awang et al. (2018), and Tsai & Lu (2009). The predominance of Logistic Regression is, on the other hand, on the moderate decline. Nevertheless, it is one of the latent factors with high impact. We suspect the influence to be driven by the need for transparent machine learning models, which may be documented by contributions of Antipov & Pokryshevskaya (2010), and Caigny et al. (2018).

We recognize that the most prevalent topics do not inevitably grow over time; moreover, they are not having an enormous impact on the field. Mainly, principal latent factors dealing with business & marketing perspectives of churn modeling show a decline concerning both covariates. The reported decline is a sign of research domain maturity and its shift towards more nuanced aspects of the problem. Besides, we see that change in topic prevalence over time is not generally associated with the normalized citation count.

5. Conclusion

This article strives to exhibit the potential of structural topic models through exploratory analysis of a large body of works in the field of customer churn prediction. We focus on abstracts of peer-reviewed articles published between the years 2009 and 2019, in journals indexed in scientific databases Web of Science or Scopus. The experimental design and implementation follow the modeling procedure proposed by Roberts et al. (2019), with original contributions to data processing and estimation & evaluation blocks. The model outcomes are presented from viewpoints of topic labels, prevalence and correlation analysis, shifts in topic prevalence over time, and citations.

We recognize groups of high-level factors based on different aspects of churn prediction such as business & marketing, machine learning & technology, industry, etc. The dominant subjects tackle broad perspectives, such as modeling, evaluation, and performance metrics. Furthermore, we investigate significant associations amongst factors with network analysis and discover communities of subjects loosely aligned with the aforementioned groups. Interestingly, we fail to prove substantial correlations for a few clearly interrelated topics. At last, we review the relationships between topic prevalence and covariates. On the one hand, essential subjects dealing with business & marketing perspectives of churn modeling manifest decline over time and low-moderate impact. On the other hand, more nuanced topics focusing on machine learning techniques and the economic context of modeling are on the rise. We consider both to be signs of a maturing research field.

The outlined insights show topic modeling as a valuable tool for navigation through scientific documents. With the structural topic models, researchers can quickly organize, summarize, annotate, and understand a vast amount of texts, even concerning relevant external determinants. Besides, augmented intuition may lead to new research opportunities. The modeling procedure, however, displays some limitations in identifying relationships between latent factors.

As for future research, we suggest modifying the original dataset – using full texts, include more document types (conference proceedings, book chapters, theses), covariates (authors, document type, journal, publisher), and re-design data processing procedure. Within the topic modeling, we propose to automate hyperparameter tuning and label generation, to consider more complex relationships between formed distributions and covariates, and to exploit the hierarchical topic structure. Besides, general applications may not be limited to the research exploration but may include customer analytics as well.

REFERENCES

[1]**Airoldi, E. M., & Bischof, J. M. (2017), *Improving and Evaluating Topic Models and other Models of Text.*** *Journal of the American Statistical Association*, 111(516), 1381-1403;

[2]**Al-Molhem, N., Rahal, Y. & Dakkak, M. (2019), *Social Network Analysis in Telecom Data.*** *Journal of Big Data*, 6(1), 1-17;

[3]**Antipov, E. & Pokryshevskaya, E. (2010), *Applying CHAID for Logistic Regression Diagnostics and Classification Accuracy Improvement.*** *Journal of Targeting, Measurement and Analysis for Marketing*, 18(2), 109-117;

[4]**Blei, D. M. (2012), *Probabilistic Topic Models.*** *Communications of the ACM*, 55(4), 77-84;

[5]**Blei, D., Ng, A. & Jordan, M. (2003), *Latent Dirichlet Allocation.*** *Journal of Machine Learning Research*, 3(4-5), 993-1022;

[6]**Bohr, J. & Dunlap, R. E. (2017), *Key Topics in Environmental Sociology, 1990–2014: Results from a Computational Text Analysis.*** *Environmental Sociology*, 4(2), 181-195;

[7]**Caigny, A., Coussement, K. & De Bock, K. W. (2018), *A New Hybrid Classification Algorithm for Customer Churn Prediction Based on Logistic Regression and Decision Trees.*** *European Journal of Operational Research*, 269(2), 760-772,

[8]**Cerchiello, P. & Nicola, G. (2018), *Assessing News Contagion in Finance.*** *Econometrics*, 6(1);

[9]**Daunis, L. & Iwan, E. (2014), *Companies Struggling to Win Customers for Life, Says New Study by Forbes Insights and Sitecore.*** *Forbes Insights.* Retrieved May 03, 2020, from https://www.forbes.com/sites/forbespr/2014/09/10/companies-struggling-to-win-customers-for-life-says-new-study-by-forbes-insights-and-sitecore/;

[10]**Garrido, F., Verbeke, W. & Bravo, C. (2018), *A Robust Profit Measure for Binary Classification Model Evaluation.*** *Expert Systems with Applications*, 92, 154-160;

DOI: 10.24818/18423264/54.4.20.19

[11]**Handley, L. (2013).** ***Customer retention: brave new world of consumer dynamics.*** *Marketing Week.* Retrieved May 03, 2020, from https://www.marketingweek.com/customer-retention-brave-new-world-of-consumer-dynamics/

[12]**Kuhn, K. D. (2018),** ***Using Structural Topic Modeling to Identify Latent Topics and Trends in Aviation Incident Reports.*** *Transportation Research Part C: Emerging Technologies,* 87, 105-122;

[13]**Mimno, D., Wallach, H., Talley, E., Leenders, M. & McCallum, A. (2011),** ***Optimizing Semantic Coherence in Topic Models.*** *In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262-272). Association for Computational Linguistics;

[14]**Mitrović, S., Baesens, B., Lemahieu, W. & De Weerdt, J. (2019).** ***tcc2vec: RFM-informed Representation Learning on Call graphs for Churn Prediction.*** *Information Sciences;*

[15]**Oliveira, M. & Gama, J. (2012),** ***An Overview of Social network Analysis.*** *Wiley Interdisciplinary Reviews: Data Mining And Knowledge Discovery,* 2(2), 99-115;

[16]**Reich, J., Tingley, D., Leder-Luis, J., Roberts, M. E. & Stewart, B. (2015),** ***Computer-Assisted Reading and Discovery for Student Generated Text in Massive Open Online Courses.*** *Journal of Learning Analytics,* 2(1), 156-184;

[17]**Roberts, M., Stewart, B. & Tingley, D. (2019).** ***stm: An R Package for Structural Topic Models.*** *Journal of Statistical Software,* 91(2);

[18]**Roberts, M., Stewart, B., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B. & Rand, D. (2014),** ***Structural Topic Models for Open-Ended Survey Responses.*** *American Journal of Political Science,* 58(4), 1064-1082;

[19]**Rodriguez, M. Y. & Storer, H. (2019),** ***A Computational Social Science Perspective on Qualitative Data Exploration: Using Topic Models for the Descriptive Analysis of Social Media Data.*** *Journal of Technology in Human Services,* 1-32;

[20]**Rothschild, J. E., Howat, A. J., Shafranek, R. M. & Busby, E. C. (2019),** ***Pigeonholing Partisans: Stereotypes of Party Supporters and Partisan Polarization.*** *Political Behavior,* 41(2), 423-443;

[21]**Sabbeh, S. (2018),** ***Machine-Learning Techniques for Customer Retention: A Comparative Study.*** *International Journal of Advanced Computer Science and Applications,* 9(2);

[22]**Shirokanova, A. & Silyutina, O. (2018),** ***Internet Regulation Media Coverage in Russia.*** *In Proceedings of the 10th ACM Conference on Web Science - WebSci '18* (pp. 359-363). ACM Press;

[23]**Straka, M. & Straková, J. (2017),** ***Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe.*** *In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies* (pp. 88-99). Association for Computational Linguistics;

[24]**Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G. & Chatzisavvas, K. C. (2015),** ***A Comparison of Machine Learning Techniques for Customer Churn Prediction.*** *Simulation Modelling Practice and Theory*, 55, 1-9;
[25]**Verbeke, W., Dejaeger, K., Martens, D., Hur, J. & Baesens, B. (2012),** ***New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach.*** *European Journal of Operational Research,* 218(1), 211-229;
[26]**Verbraken, T., Verbeke, W. & Baesens, B. (2013),** ***A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models.*** *IEEE Transactions on Knowledge and Data Engineering,* 25(5), 961-973;
[27]**Vijaya, J. & Sivasankar, E. (2019),** ***An Efficient System for Customer Churn Prediction through Particle Swarm Optimization Based Feature Selection Model with Simulated Annealing.*** *Cluster Computing,* 22(S5), 10757-10768;
[28]**Zafari, B. & Ekin, T. (2019),** ***Topic Modelling for Medical Prescription Fraud and Abuse Detection.*** *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 68(3), 751-769.

DOI: 10.24818/18423264/54.4.20.19