# Using Information-Seeking Argument Mining to Improve Service

**Bernd Skiera[1]** ⓘ**, Shunyao Yan[1]** ⓘ**, Johannes Daxenberger[2,***,
Marcus Dombois[2]** ⓘ**, and Iryna Gurevych[2]**

## Abstract

If service providers can identify reasons users are in favor of or against a service, they have insightful information that can help them understand user behavior and what they need to do to change such behavior. This article argues that the novel text-mining technique referred to as information-seeking argument mining (IS-AM) can identify these reasons. The empirical study applies IS-AM to news articles and reviews about electric scooter-sharing systems (i.e., a service enabling the short-term rentals of electric motorized scooters). Its results point to IS-AM as a promising technique to improve service; the data enable the authors to identify 40 reasons to use or not use electric scooter-sharing systems, as well as their importance to users. Furthermore, the results show that news articles are better data sources than reviews because they are longer and contain more arguments and, thus, reasons.

## Introduction

Service providers constantly aim to improve their offerings (Dotzel et al., 2013; Edvardsson and Olsson, 1996). Knowledge about users' reasons for using or not using a service helps in this pursuit because these reasons often point to service attributes relevant to its users that require improvement. Several quantitative techniques are available to accomplish this aim, such as conjoint analysis and customer satisfaction studies (Bacon, 2012; Baltas et al., 2013; Danaher, 1997; Schlereth, Skiera, and Wolk, 2011). Various qualitative approaches are available as well. These approaches rely on techniques such as design scenarios, storytelling, or customer journey maps, which usually require surveying or interviewing users (for an overview, see Vink and Koskela-Huotari, 2021).

However, both types of techniques have shortcomings. Some quantitative techniques, particularly conjoint analysis, require identifying the relevant attributes before conducting the survey, which is often challenging (Rao, 2014) and can vastly constrain the available service attributes space. As for qualitative techniques, the cost can be prohibitive; even if service providers were able to conduct surveys or interviews at moderate costs, their data could still suffer from users' dishonest answers or nonresponse bias (Wertenbroch and Skiera, 2002). Moreover, it is difficult to accurately describe specific levels of intangible attributes of a hypothetical service (Bacon, 2012; Baltas et al., 2013).

Today's world is becoming increasingly digitized, yielding an explosion of available data, most of which are unstructured, among them textual data. These data are typically available quickly, at a large scale and low cost. Harnessing these data to improve service seems promising, and service providers have already done so with textual analysis techniques, capturing users' emotional and cognitive reactions (Huang and Rust, 2021; McColl-Kennedy et al., 2019; Rust et al., 2021), topics (Antons and Breidbach, 2018), or service attributes (Chakraborty, Kim, and Sudhir, 2022; Dhillon and Aral, 2021; Toubia et al., 2019). However, these techniques fall short of automatically identifying linguistic relationships (Berger et al., 2020) (i.e., the reasons behind changes in emotions, e.g., as measured by sentiment, topics, or attributes), such as the problem behind a service attribute that receives customer complaints.

This paper addresses these gaps in the service literature by proposing information-seeking argument mining (IS-AM) as a technique to identify reasons users are in favor or against a service. Understanding these reasons can enable service providers to better understand why users behave in a certain way and use this

---

[1]Department of Marketing, Faculty of Economics and Business, Goethe-University Frankfurt am Main, Frankfurt, Germany
[2]Ubiquitous Knowledge Processing (UKP) Lab, Faculty of Computer Science, Technical University Darmstadt, Darmstadt, Germany

*Summetix GmbH, Aschaffenburg, Germany.

**Corresponding Author:**
Bernd Skiera, Department of Marketing, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, Frankfurt 60323, Germany.
Email: skiera@skiera.de

knowledge to improve their offerings. Suppose, for example, that a user notes that she "stopped going to a particular hairdresser because the hairdresser closes too early." In that case, the hairdresser learns that hours of operation constitute an important attribute and that expanding these hours could help win back this customer. The hairdresser can learn because the user provided an argument. The argument contained a reason ("closes too early") and a claim ("stopped going to the hairdresser").

Information-seeking argument mining is a subfield of argument mining. Argument mining, which is, in turn, a subfield of computational linguistics, refers to the automatic and machine-aided identification of arguments and the argumentative structure in texts or, more broadly, in natural language (Lawrence and Reed, 2020; Stede, 2020). IS-AM automatically (1) searches for documents about a specific topic, (2) extracts arguments from documents, (3) classifies the claims in those arguments into supportive ("pro") and attacking ("con") claims, and (4) classifies the reasons in those arguments. In sum, it represents a technique to quickly and automatically extract and classify reasons for and against using a particular service from many documents, which could help providers improve their service. This article aims to examine whether IS-AM is actually a useful technique to identify ways to improve service. If it is, it will complement existing methods of service design (Kurtmollaiev et al., 2018; Vink and Koskela-Huotari, 2021), new service development, and service innovation (Biemans, Griffin, and Moenaert, 2016; Ordanini and Parasuraman, 2011; Sudbury-Riley et al., 2020). This research also addresses a call to use big, unstructured data in service research (Ostrom et al., 2015) and focuses on service innovation (Antons and Breidbach, 2018; Gustafsson, Snyder, and Witell, 2020).

## Description of Argument Mining

### Description of an Argument

Research on argumentation dates back to 2000 years ago, when Aristotle recognized arguments as a means to persuade (Aristotle, 1984). An argument has two core components: claims and reasons (Stab et al., 2018b). A claim, also referred to as a conclusion, is a defeasible statement that one should not accept without additional support. A reason, also referred to as a premise, justification, or evidence, supports the claim.

Table 1 illustrates the differences between arguments and nonarguments around the aforementioned hairdresser service. The first sentence claims that "Hairdressers are worth the money" and provides the reason: "Hairdressers are trained to know what styles will look good with different face shapes and coloring." The second sentence's claim is "It is not necessary to go to a hairdresser," and the reason is that "it is easy to use hair clippers." Both sentences contain arguments because they include a claim and a reason for drawing the claim. In contrast, the third sentence is not an argument because it only includes a conclusion ("He is not a good hairdresser") but not a reason. Adding a reason (e.g., "He is not well-trained") would yield an argument.

Arguments can be supportive or attacking. A supportive argument contains a reason that supports a claim; loosely speaking, it shows why a claim is true. In our setting, the claim is that people should go to a hairdresser. In contrast, an attacking argument contains a reason that rebuts a claim; it outlines why a claim is not true. So, the first sentence in Table 1 is a supportive argument, and the second is an attacking argument. Both arguments contain reasons that hairdressers could use to improve their service. For example, hairdressers could emphasize that their training allows them to identify the best colors that match a particular person or that the ability to use hair clippers only constitutes a part of a hairdresser's job.

### Overview of Argument Mining

Argument mining primarily examines the argumentative structure of a text, such as a debate. Relevant tasks are decomposing complex language into its argumentative units, identifying and classifying the function of these (discourse) units (i.e., claims and reasons), and understanding the argumentative structure in a text by recognizing the relations between argumentative units (Stab et al., 2018b; Stede, 2020). Examples of argument mining can be found in AI debaters such as IBM's project debater (Slonim et al., 2021), applications for writing support (Stab, 2017), and the analysis of posts in web forums (Habernal and Gurevych, 2017).

In addition to argument mining, IS-AM goes a step further and focuses on extracting argumentative statements (evidence or reasoning) relevant to an externally defined topic. The central task is to find arguments related to a specific topic within a large set of texts. These texts can come from different sources and represent different points of view (Daxenberger et al., 2020). Most research in IS-AM aims to identify supporting and attacking arguments (Ajjour et al., 2019; Stab et al., 2018b), and relatively few studies address the argument search engines that detect and visualize argument components, such as claims and premises (Chernodub et al., 2019).

**Table 1.** Example of Arguments and Nonarguments.

| Standpoint and Argument | Claim and **Reason** |
| --- | --- |
| Supportive argument | "Hairdressers are **trained to know what looks good with different face shapes and coloring** and are worth the money." |
| Attacking argument | "It is unnecessary to go to a hairdresser **because it is easy to use hair clippers.**" |
| Nonargument | "He is not a good hairdresser." |

## How Information-Seeking Argument Mining Works

Figure 1 provides an overview of how IS-AM derives its results. It proceeds in three steps, which we describe in the following subsections.

*Selection of the documents.* Information-seeking argument mining aims to identify arguments on a particular topic in a large set of documents. Thus, it requires defining the topic and the set of documents—that is, a database. Two primary requirements are that the documents contain arguments and cover a wide range of topics and perspectives on those topics. The field of corpus linguistics (Stefanowitsch, 2020) deals with the specifics of creating representative document collections (i.e., corpora). Recent advances in language technology require enormously large collections, often too large for manual compilation. For example, the famous RoBERTa language model was trained on 160 GB of text, including several web crawls (Liu et al., 2019).

The range of feasible documents is large, including news articles, reviews (e.g., Reddit), the entire web (e.g., as collected by the Common Crawl project), and specified publications (e.g., academic journals). These documents contain the original text in full, and the algorithm must identify those parts that contain arguments. In rare cases, preprocessing has already occurred, such that the documents only contain parts (e.g., sentences) with arguments (Ajjour et al. 2019).

In addition to the database, the user must specify the selection criteria, including geography, language, and time. This involves defining search queries (Schütze, Manning, and Raghavan, 2008) and retrieving documents that match the search query. The search query contains the topic on either a generic level (e.g., the name of the service) or a detailed level for more specific inquiries (e.g., the name of the service provider and a geographic identifier). The retrieval then uses popular ranking functions such as BM25 (Jones, Walker, and Robertson, 2000a, 2000b). It can also use partial or semantic search (for related concepts) instead of searching for exact matches.

*Identification of the arguments.* After selecting documents, IS-AM uses a trained machine learning algorithm (usually a neural network) to identify the arguments from these documents. The input for the algorithm is the selected set of documents, the chosen unit of analysis (e.g., sentence, paragraph), and the search query. The output is a set of pro and con arguments. In summary, the machine learning algorithm must solve a three-class text classification problem that yields for each unit of analysis, whether it is a pro argument, con argument, or nonargument.

*Unit of analysis.* The documents are the input, broken down into the chosen unit of analysis, which can range from a few words over a sentence or a paragraph to a whole document. The advantage of selecting larger units of analysis (e.g., two sentences instead of one sentence) is that multiple sentences could contain the argument (e.g., two sentences: "I like apples" and "The reason is that apples are healthy"). The disadvantage is that larger units of analysis usually contain more content that is less relevant. Larger units may also contain multiple arguments, even pro and con arguments within the same unit, and separating them is challenging (Trautmann et al., 2020). For example, highlighting in one unit of analysis (e.g., a sentence) both the high quality of the service (as a supportive claim) and its high price (as an attacking claim) yields two arguments, one pro argument and one con argument. Identifying them in one unit of analysis requires a more complex argument identification process (Ma and Hovy, 2016).

Information-seeking argument mining typically uses a sentence or a paragraph as the unit of analysis (Ajjour et al., 2019; Shnarch et al., 2018; Stab et al., 2018b), such that it divides documents into the unit of analysis (e.g., a sentence) and eliminates nonsentence fragments like URLs.

*Algorithm used for argument identification.* A three-class sentence-level argument identification task is a supervised machine learning problem. It uses "ground-truth" (typically human-generated training data) to train a function. These training data are often difficult to get because detecting evidence or reasoning in texts is challenging. In addition, deciding what constitutes an argument always depends on the query (i.e., the topic). Therefore, IS-AM needs to classify different inputs for different queries such that the machine learning model eventually learns *across* topics. As a result, IS-AM researchers typically use crowdsourcing to manually generate high volumes of ground-truth training data (Ein-Dor et al., 2020; Stab et al., 2018b).

Transformer networks such as the popular BERT method have increased the quality of text classification tremendously, and argument mining has also successfully implemented them. Reimers et al. (2019) show that a BERT-based architecture outperforms previous work on three-class cross-topic IS-AM. BERT (using either base or large) word embedding maps words into numerical vector spaces that incorporate contextual information about words, which serves as the input of the model.

*Argument score.* The algorithm's outputs are usually lists of pro and con arguments. The order of the arguments depends on the retrieval score itself for offline argument identification (Wachsmuth et al., 2017) or on the score from the argument classification step for online argument identification. The latter is typically some form of confidence level in the machine learning model (e.g., the likelihood that the argument belongs to a particular class (in this case, a pro argument or con argument). As a result of the ranking, the most important arguments appear at the top of each list.
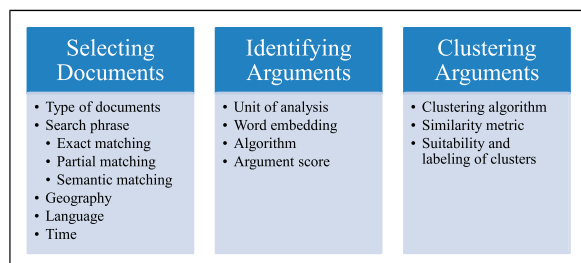


| Selecting Documents | Identifying Arguments | Clustering Arguments |
| --- | --- | --- |
| • Type of documents<br>• Search phrase<br>  • Exact matching<br>  • Partial matching<br>  • Semantic matching<br>• Geography<br>• Language<br>• Time | • Unit of analysis<br>• Word embedding<br>• Algorithm<br>• Argument score | • Clustering algorithm<br>• Similarity metric<br>• Suitability and labeling of clusters |

**Figure 1.** Workflow of information-seeking argument mining.

A straightforward approach to summarizing the results is to calculate the number of pro or con arguments—in other words, an argument score. The argument score could equal the unstandardized difference in the number of pro arguments and con arguments or the standardized difference (i.e., the difference divided by the total number of arguments). As expressed in equation (1), we prefer the latter because the number of arguments can vary across different periods $t$.

$$\text{Argument Score}_t = \frac{\text{Number of pro arguments}_t - \text{Number of con arguments}_t}{\text{Number of pro arguments}_t + \text{Number of con arguments}_t} \tag{1}$$

The standardized difference returns a single, interpretable score for each query between $-1$ and $1$. A positive score shows that more positive arguments are present than negative arguments. However, we are not proposing a "best argument score" herein; we merely follow the spirit of deriving a sentiment score in the literature (for a review, see Hartmann et al., 2019) to provide a reasonable starting point for deriving an argument score.

So, we can examine the relation of the argument score with other aspects, such as time, as identified by different periods $t$. The index $t$ could also refer to other dimensions such as geographical units like countries, data sources, or properties of the arguments themselves (e.g., whether it mentions a particular service provider or a particular topic).

We then compare the argument scores for different settings (e.g., we might retrieve a higher share of positive arguments for service A in Germany than in the UK). We can also plot arguments from a single query along a timeline (i.e., in a line chart where the $x$-axis is the time dimension and the $y$-axis is the argument score or the share of positive arguments). These charts would show trends or peaks and valleys for the search topic.

*Clustering of the arguments.* Even if the arguments are ranked, going through a potentially long list of results can still be cumbersome. Therefore, IS-AM usually summarizes all pro arguments into clusters containing similar arguments (Ajjour et al., 2018; Bar-Haim et al., 2020) using, for example, agglomerative hierarchical clustering, such as using the average linkage criterion. Such clustering builds on a pairwise similarity (Reimers et al., 2019). As a result, IS-AM derives clusters with similar pro arguments. It then follows the same procedure for all con arguments.

## Description of the Empirical Study

### Aims of the Empirical Study

Our empirical study aims to illustrate how IS-AM helps improve service. More precisely, we look at the service of renting electric scooters. Electric scooters are stand-up scooters with electric motors that support micro-mobility. A heated public debate has arisen about their usefulness, and we intend to

exploit the information provided in news outlets. We apply IS-AM to news articles and validate them with online user reviews. Even though user reviews are the most common textual source in business research, service research emphasizes the need for an "observer view" (an entity other than the customer and the firm; see Grégoire and Mattila 2021). News articles provide such a valuable observer view and thus serve as our first data source.

### Selection of the Documents

Our database contains approximately 650,000 news articles in RSS feeds of newspapers and tech magazines that the platform ArgumenText provides in several languages, such as English and German. We specified the query for relevant documents (i.e., news articles) from July 2018 to December 2019 (to avoid interference with the COVID-19 pandemic) written in English containing the search terms "e-scooter" and "electric scooter." The two queries resulted in 560 relevant documents. Figure 2 outlines how the documents are distributed across time, and Figure 3 shows the distribution of the sources (i.e., publishers) of the documents.

## Results of Empirical Study

### Identification of the Arguments

We use the platform ArgumenText (www.argumentsearch.com; Daxenberger et al., 2020) to derive our results. This publicly available software comprises publicly accessible interfaces and private backends for IS-AM. The core components of ArgumenText center around its argument detection system that builds on a transformer network to account for the dependency between the query (topic) and the argument candidate sentence (for more details, see Stab et al., 2018a). ArgumenText's current argument detection component builds on the BERT-based architecture described in Reimers et al. (2019). Stab et al. (2018a) describe the system itself, including the training data and the searchable document collection.

We find 5855 arguments in 560 documents (on average, 10.45 arguments per document). Figure 4 outlines how the number of arguments per document distributes across documents.

Of the 5855 arguments, 3669 are pro arguments (63%) and 2156 are con arguments (37%). They yield an argument score of 0.26 (= (3699 − 2156)/5855). The positive value indicates more supporting arguments (pro arguments) than attacking arguments (con arguments).

Equation (1) defines the argument score that varies over time, as Figure 5 outlines. Although the argument score was over 0.7 in mid-2018, showing very positive attitudes toward the electric scooter industry, it had trended down to less than 0.2 at the end of 2018. At the beginning of 2019, the argument score moved up to slightly over 0.4 but then dropped again.

This argument score characterizes the overall attitude toward electric scooter provider services. Managers could compare this score across different markets. However, the score itself usually
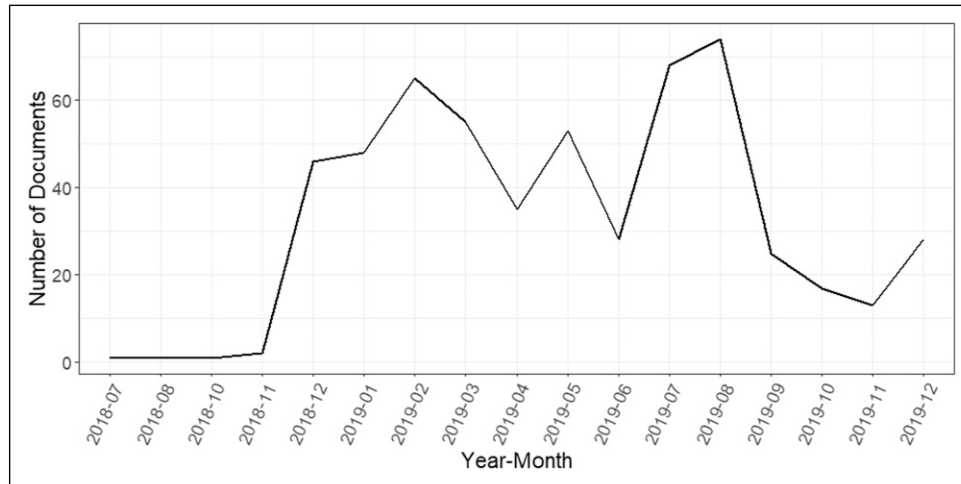
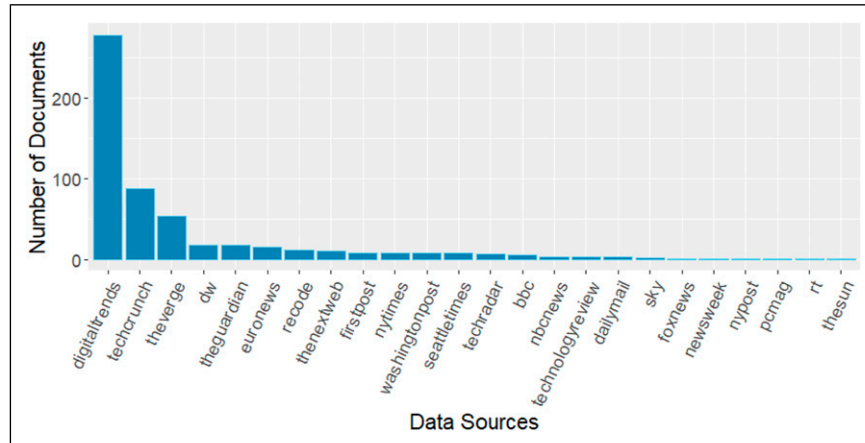**Figure 2.** Distribution of documents across time.



**Figure 3.** Distribution of documents across data sources (i.e., publishers).

does not explain why people use or do not use electric scooters. Therefore, we use clustering analysis to summarize the reasons of the pro and con arguments.

## Clustering of the Arguments

We conducted two cluster analyses to derive groups of pro and con arguments: one for pro arguments and the other for con arguments. We used a hierarchical-agglomerative clustering approach with a similarity function based on cosine similarity to cluster the arguments derived by the query "electric scooter" (which yielded 3083 pro and 1705 con arguments).

To identify the most relevant clusters of pro and con arguments (Table 2 and 3), we iteratively adjusted two parameters: the similarity threshold (measured by an adapted cosine similarity between pairs of arguments) and the minimum cluster size. The resulting clusters are as similar as possible and contain at least 80% of all arguments. To maintain interpretable results, we set the parameters so that the number of clusters ranges
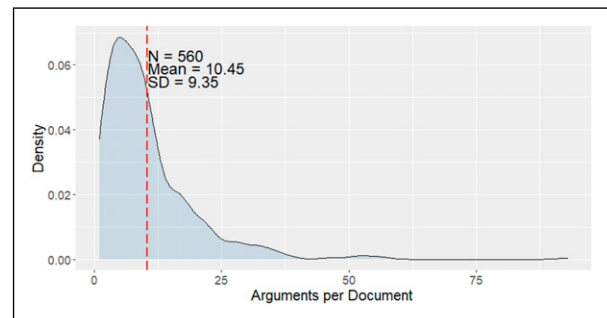


**Figure 4.** Distribution of number of arguments per document.

between 25 and 30. We then used the number of pro (con) arguments to show each cluster's importance. We focus in detail on the three most important clusters, as measured by the number of arguments.

The cluster analysis of the 3083 pro arguments in the query "electric scooter" (with a similarity threshold of 0.48 and a
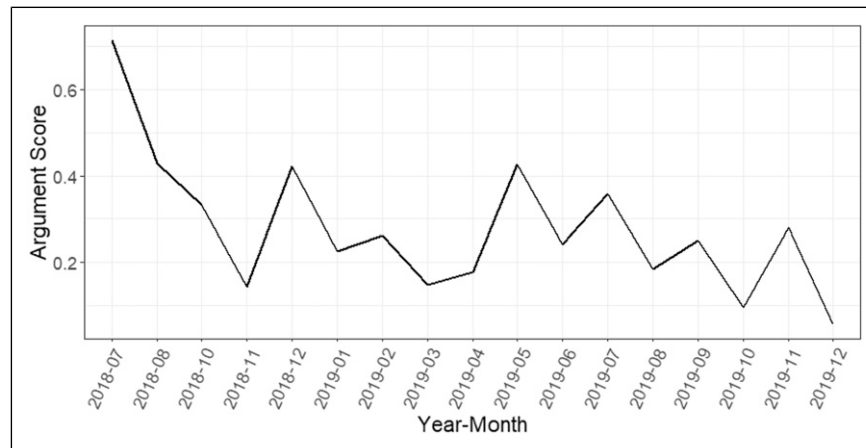
**Figure 5.** Development of argument score over time.

**Table 2.** Three Clusters with the Largest Number of Pro Arguments.

| Cluster Name | Number of Arguments | Examples |
|---|---|---|
| Flexible public transportation | 558 | "'Our investment and partnership in Lime is another step towards our vision of becoming a one-stop-shop for all your transportation needs,' Uber vice president Rachel Holt said in a statement." (Jul 10, 2018; uk.pcmag.com) |
| | | "Add to that, in European cities like Barcelona, where there has already been major investment in public transport infrastructure; there is a clear incentive to funnel residents along existing tracks, including by tightly controlling new and supplementary forms of micro-mobility." (Jan 13, 2019; techcrunch.com) |
| | | "Upsides, downside: E-scooters are billed as an environmentally friendly way to commute that can help fill in gaps in public transportation." (Jan 14, 2019; technologyreview.com) |
| Easy-to-use display | 311 | "It's easy to use, especially for beginners, with its simple menu system and touchscreen display." (Dec 3, 2018; digitaltrends.com) |
| | | "One of the scooter's real standouts is its 7-inch touchscreen display and digital speedometer, which allows you to switch between various performance options (Safe, Econ, Sport) while the vehicle isn't in motion." (Dec 8, 2018; digitaltrends.com) |
| | | "Superpedestrian's main offering is a sturdier scooter with self-diagnostic and remote management capabilities." (Dec 24, 2018; techcrunch.com) |
| High speed and reach | 302 | "With a customizable max speed of 7, 12, or 20 miles per hour, and a running distance of 10 to 20 miles per charge; this scooter will get you where you need to go quickly and reliably." (Dec 8, 2018; digitaltrends.com) |
| | | "It's supposed to be stronger, have a better rider experience and more operational efficiency, with a battery that can last 37.5 miles on a single charge, compared to just 15 miles." (Jan 10, 2019; techcrunch.com) |
| | | "The scooter tops out at 19 mph, and it can carry a max weight of 220 pounds." (May 8, 2019; theverge.com) |

minimum cluster size of 15) results in 26 clusters that contain 2543 arguments (82.48%; 540 omitted arguments). We further use argument aspect detection (Schiller, Daxenberger, and Gurevych, 2021) to automatically assign names to each cluster. We manually adjust some of these assigned names to further improve the meaning of the clusters.

Our clustering analysis also results in a few less suitable clusters. We thus examined each cluster and the respective arguments and deleted 4 clusters that do not directly refer to electric scooters but instead focus on the general debate of smart mobility or related areas and high-tech products. These outliers

occur because of our rather general query ("electric scooter"). In addition, we merged two small clusters into existing clusters. Table 4 presents our 20 pro clusters, which contain 2284 arguments.

The three most important clusters of pro arguments are *Flexible public transportation*, *Easy-to-use display*, and *High speed and reach*. As the examples in Table 2 show, the *Flexible public transportation* cluster contains arguments about electric scooters and their role in personal transportation, especially micro-mobility within a city. The cluster *Easy-to-use display* contains arguments that address, for example, the usability of

**Table 3.** Three Clusters with Largest Number of Con Arguments.

| Cluster Name | Number of Arguments | Examples |
|---|---|---|
| Dangerous | 433 | "Safety questions have also been raised, with the death of a Lime scooter rider at the weekend the third reported fatality in the US in the past three months." (Nov 27, 2018; theguardian.com)<br><br>"In September, someone lost their life after a scooter accident." (Dec 23, 2018; techcrunch.com)<br><br>"But as cities across the US have learned this year, they're also vandalization targets, a sidewalk nuisance and an injury risk." (Dec 27, 2018; washingtonpost.com) |
| Disturbing pedestrians on sidewalks | 166 | "Despite the backing of Uber and Google, who invested as part of a $335m fundraising round in July, Lime has fallen foul of the authorities in its native San Francisco, where the city banned the scooters before licensing a rival company to run a similar scheme." (Nov 27, 2018; theguardian.com)<br><br>"Regulatory challenges for these electric scooter companies abounded in Santa Monica, San Francisco, Austin and other cities around the country." (Dec 24, 2018; techcrunch.com)<br><br>"It is currently illegal to ride powered scooters—which can travel up to 30mph—on public roads or pavements, but the government has said the traffic laws are 'a barrier to innovation' and is considering changing them." (Mar 10, 2019; bbc.co.uk) |
| Unstable investment | 100 | "But as they head into year two, investors are losing interest while the business is growing increasingly expensive to operate, according to reports in *The Wall Street Journal* and *The Information*." (Dec 16, 2018; theverge.com)<br><br>"A new focus on profits could well require higher prices for customers." (Dec 27, 2018; seattletimes.com)<br><br>"When Bird launched in Santa Monica, California, in 2017, its fleet was comprised mostly of consumer scooters made by Xiaomi and Segway-Ninebot, which were never intended for heavy fleet use and depreciated quickly." (May 8, 2019; theverge.com) |

electric scooters or the touch displays integrated into many of them. It also contains arguments on the software functionalities of the electric scooters. Finally, the *High speed and reach* cluster includes arguments that address the hardware specifications of different electric scooters, mainly their range and speed.

In addition, we present three example arguments and their original news source for each cluster in Table 2. For example, the *Flexible public transportation* cluster contains arguments highlighting the flexibility for using electronic scooters as a complement to the public transportation, such as "one-stop-shop for all your transportation needs," "new and supplementary forms of micro-mobility," and "environmentally friendly way to commute that can help fill in gaps in public transportation."

The second cluster analysis (with a similarity threshold of 0.46 and a minimum cluster size of 15) uses the 1705 con arguments and yields 25 clusters that contain 1368 arguments (80.23%; 337 omitted arguments). We manually deleted unsuitable clusters, merged redundant clusters, and fine-tuned some of the clusters' names. Table 3 presents the resulting 20 clusters (1295 arguments). The three major clusters with the largest number of con arguments are *Dangerous*, *Disturbing pedestrians on sidewalks*, and *Unstable investment*. The *Dangerous* cluster contains arguments about the safety of electric scooters, accidents, aspects of vandalism, and occasionally the recklessness of the drivers. The cluster *Disturbing pedestrians on sidewalks* summarizes arguments that deal with regulatory problems and bans on electric scooters. Finally, the

*Unstable investment* cluster includes arguments that refer to electric scooter providers, especially regarding the limits of usage and finances.

Table 4 provides the complete list of derived clusters. It shows that other pro arguments include savings in terms of time and money (cluster *Saving money*), environmental friendliness (cluster *Environmentally friendly*), the low weight (cluster *Lightweight*), the thrill of riding electric scooters (cluster *Thrill*), and its attractiveness for urban customers, respectively urban mobility (cluster *Demand in the city*). Other con arguments include its high price (cluster *Expensive*), the short lifespan (cluster *Short lifetime*), the aggressive speed acceleration (cluster *Accelerate too fast*) and its high maintenance cost (cluster *High repair costs*).

## Implications for the Improvement of Service

We identify a positive but declining attitude toward providing electric scooters from the argument score. More importantly, we learn from the cluster analysis about the reasons to use or not use an electric scooter. As shown in Table 4, the five major reasons for using electric scooters are that they (1) enable an easy commute (see arguments in cluster *Flexible public transportation*), (2) are easy to use because of their convenient display (see cluster *Easy-to-use display*), (3) provide long reach and high speed (see cluster *High speed and reach*), (4) are attractive for a lot of countries (see cluster *Appealing to massive markets*), and (5) are inexpensive (see cluster *Saving money*).

**Table 4.** List of 20 Clusters with Pro Arguments (Left) and Con Arguments (Right).

| Cluster Name (Pro) | Number of Arguments | Cluster Name (con) | Number of Arguments |
|---|---|---|---|
| Flexible public transportation | 558 | Dangerous | 433 |
| Easy-to-use display | 311 | Disturbing pedestrians on sidewalks | 166 |
| High speed and reach | 302 | Unstable investment | 100 |
| Appealing to massive markets | 204 | Drunk riders | 96 |
| Saving money | 200 | Expensive | 79 |
| Battery durability | 133 | Sustainable and legal concerns | 49 |
| Multiple functions/design and features | 121 | Dockless scooters blocking sidewalks | 46 |
| Environmentally friendly | 67 | Accelerate too fast | 43 |
| Lightweight | 64 | Insufficient charging infrastructure | 42 |
| Generate revenue | 57 | Bad publicity due to class-action lawsuit | 37 |
| Accessible to everyone | 52 | Short lifetime | 30 |
| Stop everywhere | 44 | Low engine power | 29 |
| Demand in the city | 30 | Poor performance under cold temperatures | 23 |
| Flexible for different road conditions | 26 | High repair costs | 20 |
| Durable and strong mold | 23 | Uncomfortable experience under rain | 18 |
| Noise-filtering | 23 | Poor security systems | 18 |
| Demand in European countries | 21 | Easily stolen | 17 |
| Renewable power transportation | 17 | Bad public relationship | 17 |
| Thrill feeling | 16 | Technical hard to understand | 16 |
| Look different | 15 | Loud background noise | 16 |

From the clusters of pro arguments, providers can learn that the flexibility of electric scooters provides a major advantage. Thus, they can conclude that because electric scooters require using the existing infrastructure, their design must be such that they can use streets (e.g., by having proper lighting and a minimum size) or sidewalks (e.g., not being too fast or too noisy) and can be carried on public transportation (e.g., not being too large or too heavy). Filling gaps in public transportation also requires careful thinking about where to offer electric scooters (e.g., not only at railway stations, where public transportation is likely to already be sufficient). Providers can also learn that a 7-inch display offers users comfort and helps beginners better engage with the service. Moreover, the data show that users appreciate a high speed and reach, so electric scooter providers should conclude that these features are necessary and design their scooters with, for example, a higher battery capacity. Finally, the data show that electric scooters seem to fulfill a widespread user need. So, electric scooter providers should conclude that economies of scale matter and understand that the size of the served markets represents a competitive advantage.

Interestingly, the data show that low cost is the fifth most important pro and high cost is the fifth most important con argument, indicating varying price perceptions of whether prices for renting electric scooters are too high. Electric scooter providers could conclude that customers' willingness to pay varies enormously, implying that market segmentation should enable providers of electric scooters to target less price-sensitive customers.

The most important con argument refers to the safety of electric scooters. Thus, electric scooter providers need to ensure that they are safe (e.g. high-quality brakes, extensive riders' education). The data show that electric scooters also annoy pedestrians, so providers should invest in ideas that make electric scooters more compatible with pedestrians (e.g., lower speed on sidewalks) or think about how electric scooters could better use streets. Furthermore, initiatives aimed to increase intensive use of electric scooters should acknowledge that investors in electric scooter providers might lose interest because of unprofitable short-term returns or at least need to have "big pockets," considering the long period of time before they might see a profit. Scooter providers should aim for incremental innovation that they can launch quickly at a low cost. Finally, drunk drivers represent a severe problem. Providers might avoid this issue by incorporating features to detect intoxicated drivers (e.g., detecting a driving behavior indicating an intoxicated driver).

## Validation of the Results of Information-Seeking Argument Mining

In this section, we confirm the validity of our results by demonstrating the convergent validity of our results across methods and data sources. First, we compare the argument score with a sentiment score. Second, we use Reddit reviews as an alternative data set to identify reasons.

### Comparison of Information-Seeking Argument Mining with Sentiment Analysis

We introduce a novel textual analysis method that extracts the reasons behind positive and negative sentiments toward a topic. For such new methods, Berger et al. (2020) suggest providing

evidence of concurrent validity of the new method by comparing how the results derived from the new method relate to the results derived from a prior and well-validated method. The most related textual analysis method widely used in marketing and service research is sentiment mining (Berger et al., 2020). We thus provide concurrent validation of IS-AM by comparing its result with sentiment mining.

Researchers have mainly used dictionary-based sentiment mining (Berger et al., 2020; Rust et al., 2021), which assigns a polarity of sentiment (i.e., positive or negative) to a predefined list of words (i.e., dictionary). A simple approach calculates a sentiment score for the text based on the frequency of these positive or negative words. The most widely used dictionary is Linguistic Inquiry and Word Count (LIWC) (Blaseg, Schulze, and Skiera, 2020; Kübler, Colicev, and Pauwels, 2020), which

has been well validated, as indicated by traditional brand tracking surveys and stock prices (Schweidel and Moe, 2014).

Using LIWC, we classify the argument sentences into three kinds of sentiment: positive, negative, and neutral. If the sentence contains more positive words than negative words (as identified by LIWC), we classify the sentence as positive, and vice versa. If the sentence contains an equal number of positive and negative words or does not contain any positive or negative words, we classify the sentence as neutral. Table 5 reports the number of positive, neutral, and negative sentences, along with the supportive (pro) and attacking (con) sentences.

Table 5 shows that negative sentiment is more likely in a con argument. Despite many neutral sentences, positive sentiment is also more likely in a pro argument. We formally evaluate this relationship between the polarity of sentiment (i.e., positive, neutral, and negative) and the stance of argument (i.e., pro and con) with a chi-squared test and report the result in Figure 6. The $p$-value of the chi-squared test at the bottom right shows that we can reject the null hypothesis of the independence between these two measures.

The color of the bars indicates whether the observed frequencies deviate from the expected frequencies if the measures are independent. The blue bars (i.e., negative-contra sentence and positive-pro sentence) show that the observed frequencies are *larger* than expected if they were independent. The red bars indicate the opposite. Overall, the results show a positive relationship between IS-AM and sentiment analysis, supporting concurrent validity.

**Table 5.** Contingency Table of Stances of Argument and Polarity of Sentiment.

| Stances of Argument | Negative | Neutral | Positive |
|---|---|---|---|
| Attacking (con argument) | 634 (29.41%) | 1055 (48.93%) | 467 (21.66%) |
| Supportive (pro argument) | 273 (7.38%) | 1704 (46.07%) | 1722 (46.55%) |
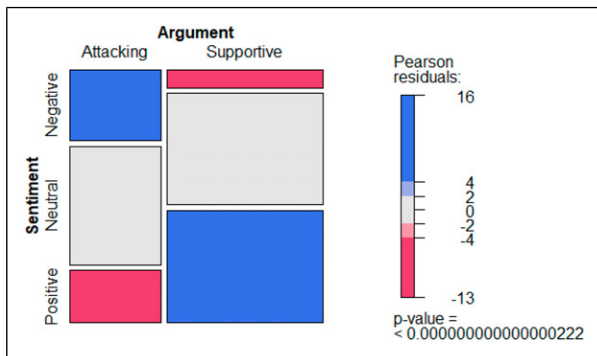


**Figure 6.** Chi-squared test of argument and sentiment.

## Comparison of News Articles with Reddit Reviews

We further demonstrate the convergent validity of IS-AM in identifying reasons using an additional source of data: Reddit reviews. We used the same query and parameters for clustering analysis, deleted unsuitable clusters and fine-tuned the cluster names. Comparing the clusters obtained from the Reddit reviews (Table 6) with those from the news articles (Table 4), we find that 87.5% of the pro clusters and 75% of the con clusters from the Reddit reviews overlap with those from the news articles. This overlapping in clusters increases our confidence in the validity of the results. The only unique pro cluster in the Reddit reviews is "Undisturbed experience," and the con clusters unique in the Reddit reviews are "Speed limits" and

**Table 6.** List of all Clusters with Pro Arguments (Left) and Con Arguments (Right) from Reddit Reviews.

| Cluster Name (Pro) | Number of Arguments | Cluster Name (con) | Number of Arguments |
|---|---|---|---|
| Accessible and stop everywhere | 142 | Dangerous | 207 |
| Saving money | 84 | Unstable investment | 90 |
| Demand in the city | 47 | Expensive | 58 |
| Battery durability | 47 | Poor security system | 33 |
| High speed and reach | 38 | Speed limits | 33 |
| Flexible for different road conditions | 29 | Ankle hurt | 32 |
| Public transportation | 19 | Dockless scooters blocking sidewalks | 23 |
| Undisturbed experience | 17 | Disturbing pedestrians on the pavement | 17 |

"Ankle hurt." These unique clusters from Reddit reviews highlight first-hand experiences from e-scooter users, which complements to news article's "observer" view. However, Reddit reviews reveal fewer clusters than news articles because user-generated content is usually shorter and contains fewer arguments. In addition, the Reddit reviews yield more unsuitable clusters than the news articles because user-generated content contains a strong bias toward weak reasons and slang.

## Summary and Conclusions

A major challenge for service providers is identifying the attributes of their service that require improvement. We address this challenge by suggesting a text-mining technique in a subfield of computational linguistics, IS-AM, which identifies and classifies arguments from a large body of documents or, more broadly, natural language. The arguments contain reasons for and against using a service, which points to attributes that users find important. Therefore, IS-AM moves textual analysis toward capturing reasoning, which Berger et al. (2020) have identified as a pressing problem in business research using textual analysis.

We are the first to show and validate an application of IS-AM in service and marketing research. Our empirical study applies IS-AM to news articles and reviews about electric scooter-sharing systems. We find evidence that IS-AM is a promising technique to improve service; in our data set, it enabled us to identify 40 reasons and their importance for using or not using electric scooter-renting systems. The comparison of IS-AM with sentiment analysis also supports the validity of our results.

Furthermore, we find that using news articles as a data source for identifying service attributes is more effective than reviews because news articles are longer and contain more arguments and, thus, reasons. New articles also provide the advantage of being widely available, which means they are accessible to new service providers considering entering a market.

Our research shows that service providers can use IS-AM to extract reasons for identifying attributes from publicly available or internal textual data to develop and improve service. Researchers can use IS-AM as an additional tool for textual analysis to understand the reasoning in addition to attitudes, such as sentiment. Policy makers could use it to understand controversial topics better.

One limitation of the IS-AM method applied in this study is that it conducted argument detection on the sentence level. A single sentence is enough for many text types, including news text, to form a valid and comprehensive argument. However, more complex reasoning, as found in the scientific literature, often spans multiple sentences, making IS-AM much more difficult. Some methods detect arguments on the word level and allow for detecting arguments spanning multiple sentences. However, none of them has been adapted to detect multi-sentence arguments. Analyzing complex argumentation spanning several sentences is typically only workable with approaches from discourse analysis, which detect pro and con arguments and chains of arguments. Future work should seek to create methods combining methods from IS-AM and discourse analysis, which also apply to argument mining in scientific literature.

Service researchers can further explore how argument mining can benefit businesses. For example, most services receive customer complaints, and Van Vaerenbergh et al. (2019) recommend that service providers appropriately address those complaints. Future research could use IS-AM to identify the customer's most crucial argument and guide the customer complaint agent toward addressing it instantly. In addition, our paper uses IS-AM to cluster the arguments to identify the most important pro and con arguments for all providers. Future research could analyze subsets of providers or periods to see how competition between providers evolves.

## Author's Note

Bernd Skiera is also a Professorial Research Fellow at Deakin University (Australia).

## ORCID iDs

Bernd Skiera https://orcid.org/0000-0001-9285-2013
Shunyao Yan https://orcid.org/0000-0001-8576-6600
Marcus Dombois https://orcid.org/0000-0002-7886-0612

## References

Ajjour, Yamen, Henning Wachsmuth, Dora Kiesel, Patrick Riehmann, Fan Fan, Giuliano Castiglia, Rosemary Adejoh, Bernd Fröhlich, and Benno Stein. 2018. "Visualization of the Topic Space of Argument Search Results in Args.Me." In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 60-65. Stroudsburg, PA: Association for Computational Linguistics.

Ajjour, Yamen, Henning Wachsmuth, Johannes Kiesel, Martin Potthast, Matthias Hagen, and Benno Stein. 2019. "Data Acquisition for Argument Search: The Args.Me Corpus." In *Ki 2019: Advances in Artificial Intelligence*, edited by Christoph Benzmüller and Heiner Stuckenschmidt, 48-59. Heidelberg: Springer International Publishing.

Antons, David and Christoph F. Breidbach. 2018. "Big Data, Big Insights? Advancing Service Innovation and Design with Machine Learning." *Journal of Service Research*, 21 (1): 17-39.

Aristotle. 1984. *The Complete Works of Aristotle*. Princeton, NJ: Princeton University.

Bacon, Donald R. 2012. "Understanding Priorities for Service Attribute Improvement." *Journal of Service Research* 15 (2): 199-214.

Baltas, George, Stelios Tsafarakis, Charalampos Saridakis, and Nikolaos Matsatsinis. 2013. "Biologically Inspired Approaches to Strategic Service Design: Optimal Service Diversification through Evolutionary and Swarm Intelligence Models." *Journal of Service Research* 16 (2): 186-201.

Bar-Haim, Roy, Lilach Eden, Roni Friedman, Yoav Kantor, Dan Lahav, and Noam Slonim. (2020). "From Arguments to Key Points: Towards Automatic Argument Summarization." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4029-4039. Stroudsburg, PA: Association for Computational Linguistics.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel. (2020). "Uniting the Tribes: Using Text for Marketing Insight." *Journal of Marketing* 84 (1): 1-25.

Biemans, Wim G., Abbie Griffin, and Rudy K. Moenaert. (2016). "Perspective: New Service Development: How the Field Developed, its Current Status and Recommendations for Moving the Field Forward." *Journal of Product Innovation Management* 33 (4): 382-397.

Blaseg, Daniel, Christian Schulze, and Bernd Skiera. 2020. "Consumer Protection on Kickstarter." *Marketing Science* 39 (1): 211-233.

Chakraborty, Ishita, Minkyung Kim, and K. Sudhir. (2022). "Attribute Sentiment Scoring with Online Text Reviews: Accounting for Language Structure and Missing Attributes." *Journal of Marketing Research* 59 (3): 600-622.

Chernodub, Artem, Oleksiy Oliynyk, Philipp Heidenreich, Bondarenko Alexander, Matthias Hagen, Chris Biemann, and Alexander Panchenko. (2019). "Targer: Neural Argument Mining at Your Fingertips." In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 195-200. Stroudsburg, PA: Association for Computational Linguistics.

Danaher, Peter J. (1997). "Using Conjoint Analysis to Determine the Relative Importance of Service Attributes Measured in Customer Satisfaction Surveys." *Journal of Retailing* 73 (2): 235-260.

Daxenberger, Johannes, Benjamin Schiller, Chris Stahlhut, Erik Kaiser, and Iryna Gurevych. 2020. "ArgumenText: Argument Classification and Clustering in a Generalized Search Scenario." *Datenbank-Spektrum* 20 (2): 115-121.

Dhillon, Paramveer and Sinan Aral. 2021. "Modeling Dynamic User Interests: A Neural Matrix Factorization Approach." *Management Science*, forthcoming.

Dotzel, Thomas, Venkatesh Shankar, and Leonard L. Berry. 2013. "Service Innovativeness and Firm Value." *Journal of Marketing Research* 50 (2): 259-276.

Edvardsson, Bo and Jan Olsson. 1996. "Key Concepts for New Service Development." *Service Industries Journal* 16 (2): 140-164.

Ein-Dor, Liat, Eyal Shnarch, Lena Dankin, Halfon Alon, Benjamin Sznajder, Ariel Gera, Carlos Alzate, Martin Gleize, Leshem Choshen, Yufang Hou, Yonatan Bilu, Ranit Aharonov, and Noam Slonim. (2020). "Corpus Wide Argument Mining—a Working Solution." In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 7683-7691. Palo Alto, CA: AAAI.

Grégoire, Yany and Anna S. Mattila. 2021. "Service Failure and Recovery at the Crossroads: Recommendations to Revitalize the Field and Its Influence." *Journal of Service Research* 24 (3): 323-328.

Gustafsson, Anders, Hannah Snyder, and Lars Witell. 2020. "Service Innovation: A New Conceptualization and Path Forward." *Journal of Service Research* 23 (2): 111-115.

Habernal, Ivan and Iryna Gurevych. 2017. "Argumentation Mining in User-Generated Web Discourse." *Computational Linguistics* 43 (1): 125-179.

Hartmann, Jochen, Juliana Huppertz, Christina Schamp, and Mark Heitmann. 2019. "Comparing Automated Text Classification Methods." *International Journal of Research in Marketing* 36 (1): 20-38.

Huang, Ming-Hui and Roland T. Rust. 2021. "Engaged to a Robot? The Role of Ai in Service." *Journal of Service Research* 24 (1): 30-41.

Jones, K. Sparck, Steve Walker, and Stephen E. Robertson. 2000a. "A Probabilistic Model of Information Retrieval: Development and Comparative Experiments: Part 1." *Information Processing & Management* 36 (6): 779-808.

Jones, K. Sparck, Steve Walker, and Stephen E. Robertson. 2000b. "A Probabilistic Model of Information Retrieval: Development and Comparative Experiments: Part 2." *Information Processing & Management* 36 (6): 809-840.

Kübler, Raoul V., Anatoli Colicev, and Koen H. Pauwels. 2020, "Social Media's Impact on the Consumer Mindset: When to Use Which Sentiment Extraction Tool?" *Journal of Interactive Marketing* 50 (May): 136-155.

Kurtmollaiev, Seidali, Annita Fjuk, Per Egil Pedersen, Simon Clatworthy, and Knut Kvale. 2018. "Organizational Transformation through Service Design: The Institutional Logics Perspective." *Journal of Service Research* 21 (1): 59-74.

Lawrence, John and Chris Reed. 2020. "Argument Mining: A Survey." *Computational Linguistics* 45 (4): 765-818.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. "Roberta: A Robustly Optimized Bert Pretraining Approach." arXiv preprint arXiv:1907.11692.

Ma, Xuezhe and Eduard Hovy. 2016. "End-to-End Sequence Labeling via Bi-Directional Lstm-Cnns-Crf." In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Vol. 1: Long Paper, 1064-1074. Stroudsburg, PA: Association for Computational Linguistics.

McColl-Kennedy, Janet R., Zaki Mohamed, Katherine N. Lemon, Florian Urmetzer, and Andy Neely. 2019. "Gaining Customer Experience Insights that Matter." *Journal of Service Research* 22 (1): 8-26.

Ordanini, Andrea and A. Parasuraman. 2011. "Service Innovation Viewed through a Service-Dominant Logic Lens: A Conceptual

Framework and Empirical Analysis." *Journal of Service Research* 14 (1): 3-23.

Ostrom, Amy L., A. Parasuraman, David E. Bowen, Lia Patrício, and Christopher A. Voss. 2015. "Service Research Priorities in a Rapidly Changing Context." *Journal of Service Research* 18 (2): 127-159.

Rao, Vithala R. (2014). *Applied Conjoint Analysis*. Heidelberg: Springer.

Reimers, Nils, Benjamin Schiller, Tilman Beck, Johannes Daxenberger, Christian Stab, and Iryna Gurevych 2019. "Classification and Clustering of Arguments with Contextualized Word Embeddings." In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 567-578. Stroudsburg, PA: Association for Computational Linguistics.

Rust, Roland T., William Rand, Ming-Hui Huang, Andrew T. Stephen, Gillian Brooks, and Timur Chabuk. 2021. "Real-Time Brand Reputation Tracking Using Social Media." *Journal of Marketing* 85 (4): 21-43.

Schiller, Benjamin, Johannes Daxenberger, and Iryna Gurevych. 2021. "Aspect-Controlled Neural Argument Generation." In Annual Conference of the North American Chapter of the Association for Computational Linguistics, 380-396. Stroudsburg, PA: Association for Computational Linguistics.

Schlereth, Christian, Bernd Skiera, and Agnieszka Wolk. 2011. "Measuring Consumers' Preferences for Metered Pricing of Services." *Journal of Service Research* 14 (4): 443-459.

Schütze, Hinrich, Christopher D. Manning, and Prabhakar Raghavan. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Schweidel, David A. and Wendy W. Moe. 2014. "Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice." *Journal of Marketing Research* 51 (4): 387-402.

Shnarch, Eyal, Carlos Alzate, Lena Dankin, Martin Gleize, Yufang Hou, Leshem Choshen, Ranit Aharonov, and Noam Slonim. 2018. "Will It Blend? Blending Weak and Strong Labeled Data in a Neural Network for Argumentation Mining." In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Vol. 2: Short Paper, 599-605. Stroudsburg, PA: Association for Computational Linguistics.

Slonim, Noam, Yonatan Bilu, Carlos Alzate, Roy Bar-Haim, Bogin Ben, Francesca Bonin, Leshem Choshen, Edo Cohen-Karlik, Lena Dankin, and Lilach Edelstein, 2021. "An Autonomous Debating System." *Nature* 591 (7850): 379-384.

Stab, Christian. 2017. "Argumentative Writing Support by Means of Natural Language Processing." Doctoral Dissertation, Technische Universität Darmstadt, Darmstadt, Germany.

Stab, Christian, Johannes Daxenberger, Chris Stahlhut, Tristan Miller, Benjamin Schiller, Christopher Tauchmann, Steffen Eger, and Iryna Gurevych. 2018a. "ArgumenText: Searching for Arguments in Heterogeneous Sources." In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, 21-25. Stroudsburg, PA: Association for Computational Linguistics.

Stab, Christian, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. 2018b. "Cross-Topic Argument Mining from Heterogeneous Sources." In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Vol. 1: Long Paper, 3664-3674. Stroudsburg, PA: Association for Computational Linguistics.

Stede, Manfred. 2020. "Automatic Argumentation Mining and the Role of Stance and Sentiment." *Journal of Argumentation in Context* 9 (1): 19-41.

Stefanowitsch, Anatol. 2020. *Corpus Linguistics: A Guide to the Methodology.* Berlin: Language Science Press.

Sudbury-Riley, Lynn, Philippa Al-Abdin, Ahmed Hunter-Jones, Ahmed Al-Abdin, Daniel Lewin, and Mohabir Vic Naraine. 2020. "The Trajectory Touchpoint Technique: A Deep Dive Methodology for Service Innovation." *Journal of Service Research* 23 (2): 229-251.

Toubia, Olivier, Garud Iyengar, Renée Bunnell, and Alain Lemaire. 2019. "Extracting Features of Entertainment Products: A Guided Latent Dirichlet allocation Approach Informed by the Psychology of Media Consumption." *Journal of Marketing Research* 56 (1): 18-36.

Trautmann, Dietrich, Johannes Daxenberger, Christian Stab, Hinrich Schütze, and Iryna Gurevych. 2020. "Fine-Grained Argument Unit Recognition and Classification." In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 9048-9056. Palo Alto, CA: AAAI.

Van Vaerenbergh, Yves, Dorottya Varga, Arne De Keyser, and Chiara Orsingher (2019). The Service Recovery Journey: Conceptualization, Integration, and Directions for Future Research. *Journal of Service Research*, 22(2), 103-119.

Vink, Josina and Kaisa Koskela-Huotari. 2021, "Building Reflexivity Using Service Design Methods." *Journal of Service Research*, forthcoming. https://doi.org/10.1177/10946705211035004

Wachsmuth, Henning, Martin Potthast, Khalid Al Khatib, Yamen Ajjour, Puschmann Jana, Jiani Qu, Jonas Dorsch, Viorel Morari, Bevendorff Janek, and Benno Stein. 2017. "Building an Argument Search Engine for the Web." In Proceedings of the 4th Workshop on Argument Mining, 49-59. Stroudsburg, PA: Association for Computational Linguistics.

Wertenbroch, Klaus and Bernd Skiera. 2002. "Measuring Consumers' Willingness to Pay at the Point of Purchase." *Journal of Marketing Research* 39 (2): 228-241.