Tony Brownlee

Jesse Sommerfeld

Kyle Hansen

*1401 6th Avenue South, Clear Lake, IA 50428, USA
Tel: 641-355-1000; E-mail: Tony.Brownlee@Kingland.com

**2420 Lincoln Way, Ames, IA 50014, USA
Tel: 515-268-7020

***1401 6th Avenue South, Clear Lake, IA50428, USA
Tel: 641-355-1000

Journal of Securities Operations & Custody
Vol. 12, No. 2, pp. 102–115

1753-1802

How AI is changing operations: From settlement optimisation to automating risk monitoring

Received: 2nd December, 2019

Tony Brownlee*
President, Kingland, USA

Jesse Sommerfeld**
Head of Data Science, Kingland, USA

Kyle Hansen***
Head of AI Engineering, Kingland, USA

Tony Brownlee *is the President of Kingland and is responsible for all business operations for the company. Tony is recognized as a creative thought leader and is a frequent speaker at executive level events as well as working with Fortune 500 leadership teams on strategies to advance how people use data and technology. He holds a Bachelor of Arts in economics from Central College in Pella, Iowa and an MBA from Iowa State University.*

Jesse Sommerfeld *is Head of Data Science at Kingland. He studied at Iowa State University's Ivy College of Business. At Kingland, he has also worked as a Data Analyst, Data Process Lead and Solution Analyst, and Data Research Analyst.*

Kyle Hansen *is Head of AI Engineering at Kingland and works on AI (artificial intelligence) technologies for new and existing solutions. Kyle graduated from Iowa State University with a degree in Computer Science and a minor in English. He has worked on data projects involving database design and management, back-end data processing, settlement and more.*

ABSTRACT

Banking executives continue to evaluate new ways to incorporate artificial intelligence (AI) into the middle and back office to improve efficiency, mitigate risk and reduce cost. Even with these efforts, a Business Insider report[1] estimates the aggregate potential cost savings from AI applications at US$248bn by 2023 in the middle and back office. In this paper, the authors discuss two uses of AI that are experiencing investments from firms: settlement optimisation (maximising daily resolved trades by using AI to optimise the order that transactions are settled) and automating risk monitoring (using natural language processing to systematically review and filter content relevant to a specific risk-monitoring purpose).

Keywords: settlement optimisation, text analytics, risk monitoring, natural language processing, text summarisation, content collection, robotic process automation, event logging

SETTLEMENT OPTIMISATION

The burden of confirming failed transactions is expensive and time consuming for back-office operations. Looking at the symptoms of rigid processing logic and limited look-ahead scenarios, a core problem that must be solved for the settlement process is to determine the optimal order in which transactions should be processed. Doing so will allow for the optimisation of asset lending activities and margin facilities. The goal is to increase the percentage of the day's trades that can be resolved.

In order to tackle settlement optimisation, firms must understand the natural relationship of buyer and seller, where if someone sells shares, someone else has

bought those shares. The legacy settlement systems in many big banks typically process these large quantities of transactions across many asset classes as one large volume of buy/sell information. As a result, failures occur daily, which lead to teams of individuals chasing down details on short sales, securities lending discrepancies, allocation issues and a multitude of other fail[2] scenarios. This process is far from optimal.

VALIDATING TRANSACTIONS

Many firms attempt to institute strict rules to prevent fails and ensure the validity of transactions as part of the settlement process, but these rules fall short of helping firms manage the full picture. Firms want to understand what they can settle and what they cannot, how to keep clients happy and how to maximise the trading or lending strategies at play. For example, oftentimes, a firm may prioritise settling transactions involving a specific firm first. Doing so can hamper the rest of the settlement process because it is challenging to understand the complete picture. Firms are unable to see other routes or possibilities to settling more transactions.

Back-office personnel will contact clients regarding failed transactions, only to discover after speaking with the client that the client's transactions should have been completed but did not due to other trades that happened throughout the trading day. The back and forth of phone calls and e-mails to verify what can and cannot close puts pressure on T+1 or T+2 timelines.

To understand this problem better, we will use an example of volume for a single security. If there are n transactions for one security, there are $n!$ different ways to execute those transactions. Suppose the example is isolated to one security with 800 transactions. There would be permutations of those transactions to consider. Because some permutations are more optimal — ie they result in a larger percentage of transactions settled — more must be done than to just consider an arbitrary settlement order or adherence to strict rules. Because settling a given transaction in a particular order influences the positions of the various clients (an algebraic problem), the settlement of that one transaction carries with it long-term dependencies that must also be considered. Therefore, it is best to not simply compare pairs of transactions in isolation as would typically be the case when sorting, because there is a lack of information needed to confidently say which of the transactions should occur first.

What this means in simple terms is determining the optimal order of processing for these transactions is exponentially more complex than some other problems. Figure 1 points out the size and complexity of just one security, not even the hundreds of thousands of securities settled each day.

Figure 1 does not take into account other factors and variables (eg member priority settings) that may also increase the number of potential computations.

Using the given example, it is not reasonably feasible to examine all possible orderings in a 24-hour period even using significantly powerful computing capabilities. Even if a machine were able to compute and evaluate millions of possibilities a second, it would take multiple years to evaluate all possible scenarios. There are simply too many permutations.

In addition to the complexity of the problem and the speed, there are other variables. Even if an optimal settlement order can be achieved in a reasonable amount of time, the start-of-day (SOD) positions of some clients provide additional complexity to the settlement of their own transactions as well as others. Additionally, there are likely a multitude of rules and variables that impact the processing and complexity of the problem, including a limit on the amount of

A Chessboard	The Known Universe	One CUSIP with 800 transactions
7.7×10^{45} Configurations	3.28×10^{80} Particles	7.71×10^{1976} Settlement Orders
		

Figure 1 Optimising for size and complexity
Note: CUSIP, Committee on Uniform Securities Indentification Procedures.

cash an entity is allowed to spend or collateral settlement.

Asset lending and margin pledging provide additional strategies to resolve transactions and reduce fails, but they require additional calculation and processing. These fails create additional overhead not just in the processing but also in support time and costs and should be considered in the overall optimisation process. Given these factors, reinforcement learning[3] offers the best solution. Reinforcement learning algorithms can learn from interactions and improve with time, learning optimal settlement transactions. According to Sutton and Barto, 'the key idea of reinforcement learning generally, is the use of value functions to organize and structure the search' for choosing the next action.[4]

OPTIMISING THE SETTLEMENT PROCESS

Our strategies to optimise the settlement process use multiple AI approaches and algorithms that optimise a firm's throughput by identifying a specific client with the sufficient shares. The algorithms can simply optimise the settlement process or understanding thereof by using network flow.[5] The algorithm finds pipes for shares to move through that allow for more

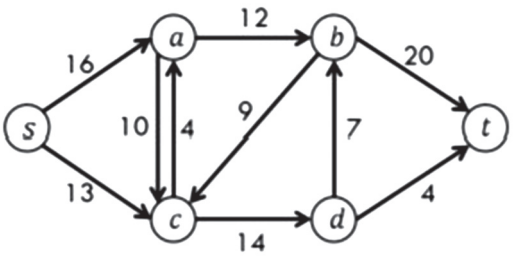

Figure 2 Network flow diagram[6]

trades to settle. It looks at the entire problem as one, finding multiple pipes of transactions instead of using strict rules that incorporate finding transactions from one pipe.

Network flow is a directed graph where each edge has a capacity and each edge receives a flow (Figure 2). In this example, imagine an image of nodes that are connected by pipes. The nodes would represent shareholders. The pipes would represent transactions made throughout the day. By modelling participants, the graph would reveal the quantity of shares and collateral moving through the pipes and use an algorithm to maximise the flow. By maximising the flow of transactions through the pipes, the transactions that will settle are maximised. If a failed transaction is discovered, it can be removed, and the

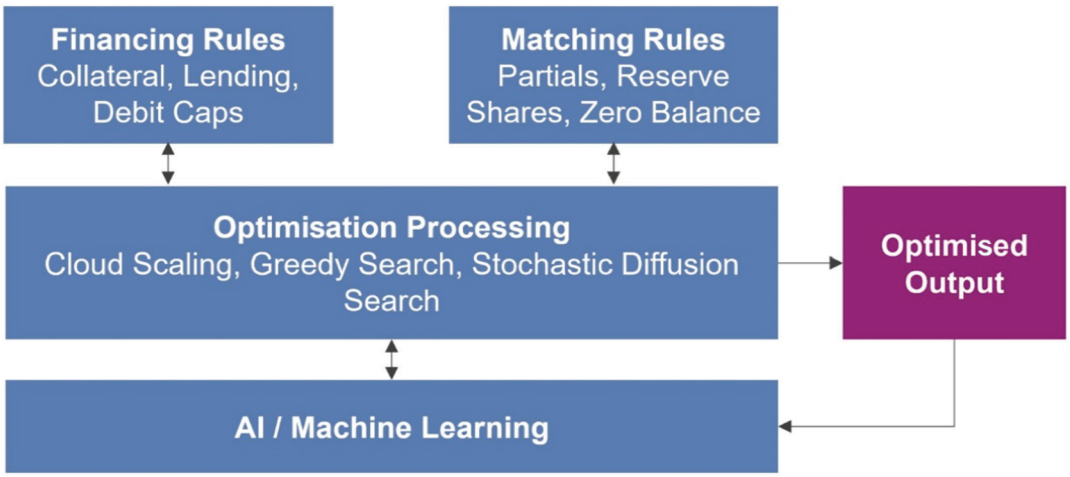

Figure 3 Allocation and asset lending processing
Note: AI, artificial intelligence.

algorithm can run again in order to find an optimum level of settlement success.

Before starting the settlement optimisation algorithm, each client's optimised position is determined, which is the sum of a given client's SOD balance and their receives, less their deliveries. This final balance represents the client's predicted position, assuming all transactions settle. If this number is below the client's tolerable position, the client needs to borrow from a lender in order to meet its obligations. The algorithm could inspect the available lenders in an effort to create loan transactions to bring every client's optimised position, cash balance and collateral balance into bounds (subject to constraints imposed by the particular simulation the system is running). This entire process is motivated by the idea that settling transactions is more important than avoiding negative client positions; there may be, however, conditions where the system must consider the needs of a particular set of clients above the need of others to settle all transactions. Figure 3 provides a high-level visualisation of the flow of logic.

A high-level explanation of the processing follows:

- Cloud scaling allows for limitless compute power on all the transactions that need to be processed.
- Greedy search algorithm[7] picks the best solution at the moment without regard for consequences. It picks the best immediate output but does not consider the big picture; hence, it is considered greedy.
- Stochastic diffusion search[8] is a well-characterised robust swarm intelligence global metaheuristic that can efficiently solve search and optimisation problems with compositional structure.

Optimised positions assume all transactions will clear, and subsequent allocation calculations assume all able clearing members are willing to lend in order to bring all optimised positions into an acceptable range. As lending situations may be more complicated than that, simulations can calculate solutions when lending constraints apply, either in the form of future obligations or

another requirement. These solutions may not clear all transactions, but they are a more accurate representation of real-life circumstances.

After calculating optimised positions, computing the ideal settlement order for a security's transactions can begin. Because the brute force approach (also known as exhaustive search) to allocation is not tenable due to the sheer number of calculations required, firms may need to perform this computation using some other approach. One alternative is greedy search, which iterates through settlement orders in a similar way to exhaustive search but abandons lines of inquiry when they fail to show promise. This works well for securities with smaller transaction sets but runs into the same dimensionality problems as exhaustive search above about ten transactions.

For securities with large volumes of transactions, different AI families of algorithms are considered (eg stochastic diffusion search) to represent the best approach for this settlement optimisation (speed and settlement rate) problem. Again, the goal is to settle transactions optimally and improve with repeated iterations, so the more iterations one allows, often the better one obtains (up to and including the 'optimal' solution, where here 'optimal' means all transactions have cleared).

Once a settlement order has been determined, transactions can be applied in order to update each clearing member's cash and collateral balances. When all securities are finished processing, each clearing member's balance will represent its end-of-cycle values for cash and collateral.

EXTENDING OPTIMISATION WITH MACHINE LEARNING

This provides the ability to automatically tune variable configuration points based on complex patterns derived from data. The configuration points used by the search algorithm are called 'metaparameters', or parameters that influence the success of the algorithm but cannot be derived from the input data in an obvious way. Manually choosing values for such metaparameters is difficult and subject to data idiosyncrasies; thus one should employ machine learning to tie configuration decisions directly to past data.

Such a system learns from its own successes and failures to build a statistical model that represents the real-life implications of selecting a particular set of configuration parameters for a given scenario. One parameter could be rules that apply across securities. How does one resolve participant rules against the security level rules during settlement? One would think securities with more transactions need additional compute power, but with different machine learning techniques, that is not necessarily true. One approach to learning would be the Bayes family of classifiers,[9] which reconstruct conditional probability distributions by analysing sample information. Such a classifier can learn to answer questions like, 'Given the settlement history of this security and these particular clearing members, what is the probability the security will require extra computing power?'

Another learning example is the random forest,[10] a collection of decision trees the learning algorithm builds by analysing sample data. Each tree in the collection votes with its opinion of the correct answer, and the majority vote wins. Random forests can, among other tasks, answer questions like, 'Based on past knowledge, if we apply Strategy X, will this calculation succeed in a reasonable amount of time?'

Beyond analysing historical data, running hypothetical simulations can help train machine learning models how to suggest runtime strategies. Runtime strategies are motivated by a settlement scenario's unique requirements, and they serve to map high-level settlement requirements to low-level

configuration levers. These simulations allow parties to examine the effects of changing isolated variables, and they can inform future automated decisions even if the hypothetical scenarios from the simulations never previously arose. Using certain unsupervised learning strategies can help users understand the patterns that can then be used by the core processing.

This approach is founded by the idea that settling transactions is more important than avoiding negative client positions. By using a learning algorithm that can settle trades in a specific order, firms can settle trades in an optimal fashion, lowering processing time and increasing the speed and accuracy of settlements. Back-office operations could use this approach to appropriately validate and take necessary actions on failed trades. Reinforcement learning can help firms conduct more lending, for example, due to having an accurate account of the balances and transactions during the final life cycle stage of trades, thereby reducing the overestimation of failed transactions and optimising the settlement process.

AUTOMATING RISK MONITORING

While daily middle- and back-office processes are running, working to clear and settle transactions, the total volume and exposure fluctuate, and risk departments monitor thousands of clients and counterparties to assess risk. The risk-monitoring process is another critical function that can be streamlined using AI. To accurately assess a variety of risks, firms need to understand the activities conducted by their counterparties, such as customers, suppliers/vendors, employees and prospects. In an effort to identify activities that could be impactful to a risk position, analysts will review an aggregation of sources from news articles and public filings to social media and arrest records. These sources exist across many mediums, geographies and languages. When expanding this monitoring activity across hundreds of thousands or more counterparties of a firm, it quickly becomes a high-cost, time-intensive and unsustainable exercise.

As a result, it is common for organisations to prioritise (eg 'high-profile' list) who they are going to monitor while employing a rotation and/or sampling strategy for the remainder. Even with this strategy, there is a limit to how much content a risk analyst can consume and review over any given time period; these monitoring activities are likely the least exciting function of their job, and they may not even consider it to be their core function. This is a poor mix of variables when it comes to monitoring an organisation's risks, but it can also undermine the efficacy of an organisation's growth strategy.

This is where applying technology can help. Much like alternative data[11] can provide insight for investment, the conceptual application to monitoring for a specific use case applies. Table 1 shows some high-level functions of monitoring customers, suppliers/vendors, employees and prospects.

Most have heard of web crawling and robotic process automation given their publicity over the last ten years, but natural language processing (NLP), a key component of text analytics, is where significant power is isolated when dealing with unstructured content. Utilising parts of speech, dependency parsing, named entity recognition and phrase/pattern matching allows one to systematically review and filter content relevant to a specific risk-monitoring purpose.

As a use case, we will evaluate the Securities and Exchange Commission (SEC) administrative proceeding releases.[12] One can notice in Figure 4, the volume spiked at the end of Q3 (September 2019), so we will focus on that month.

An initial evaluation of the documents can help one start to understand the nature

Table 1: High-level functions of monitoring

Action	*Description*	*Technological Application Example*
Content collection	Literal action of navigating to news sites, service portals and reports that may or may not reveal clients engaging in events/actions that are being monitored	Web crawling
Content review	Reading the content, looking for clients and events/action being monitored	Natural language processing
Content filtering	Reducing content to those that include clients and events/action being monitored	Natural language processing and robotic process automation
Event/action logging	Tie the events/actions to something (eg client record) in the system for immediate or later use	Robotic process automation

of the text being discussed. The documents average 5.1 characters per word and 35.4 words per sentence. Comparing this to the general English average, the words are 1.07× times longer, and sentences have 2.02× more words in them. The general feeling is that these documents[13] would be considered 'wordy' and likely have language that is more difficult to understand (see Figure 5).

Leveraging named entity recognition[14] (NER), the documents average 172 named entity mentions, with organisations being the most common entity type. Figure 6[15] illustrates the distribution per document.

At this point, there is a baseline understanding of the document makeup, but now it is time to learn what the documents are about. To do so, it is advised to use text summarisation[16] and topic modelling[17] techniques coupled with document clustering.[18]

There are two main ways to approach text summarisation. One way is to use

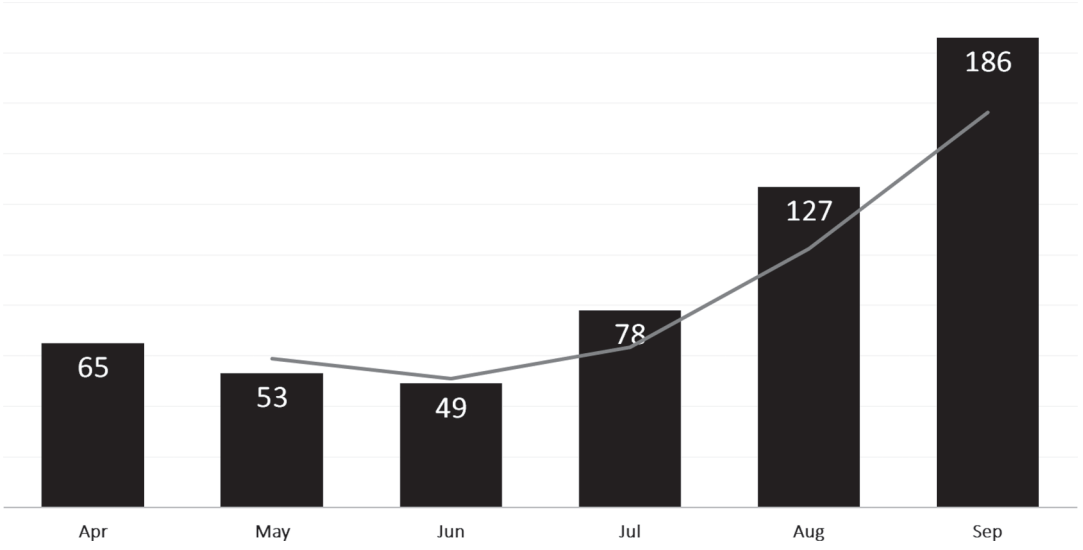

Figure 4 SEC administrative proceeding volume
Note: SEC, Securities and Exchange Commission.

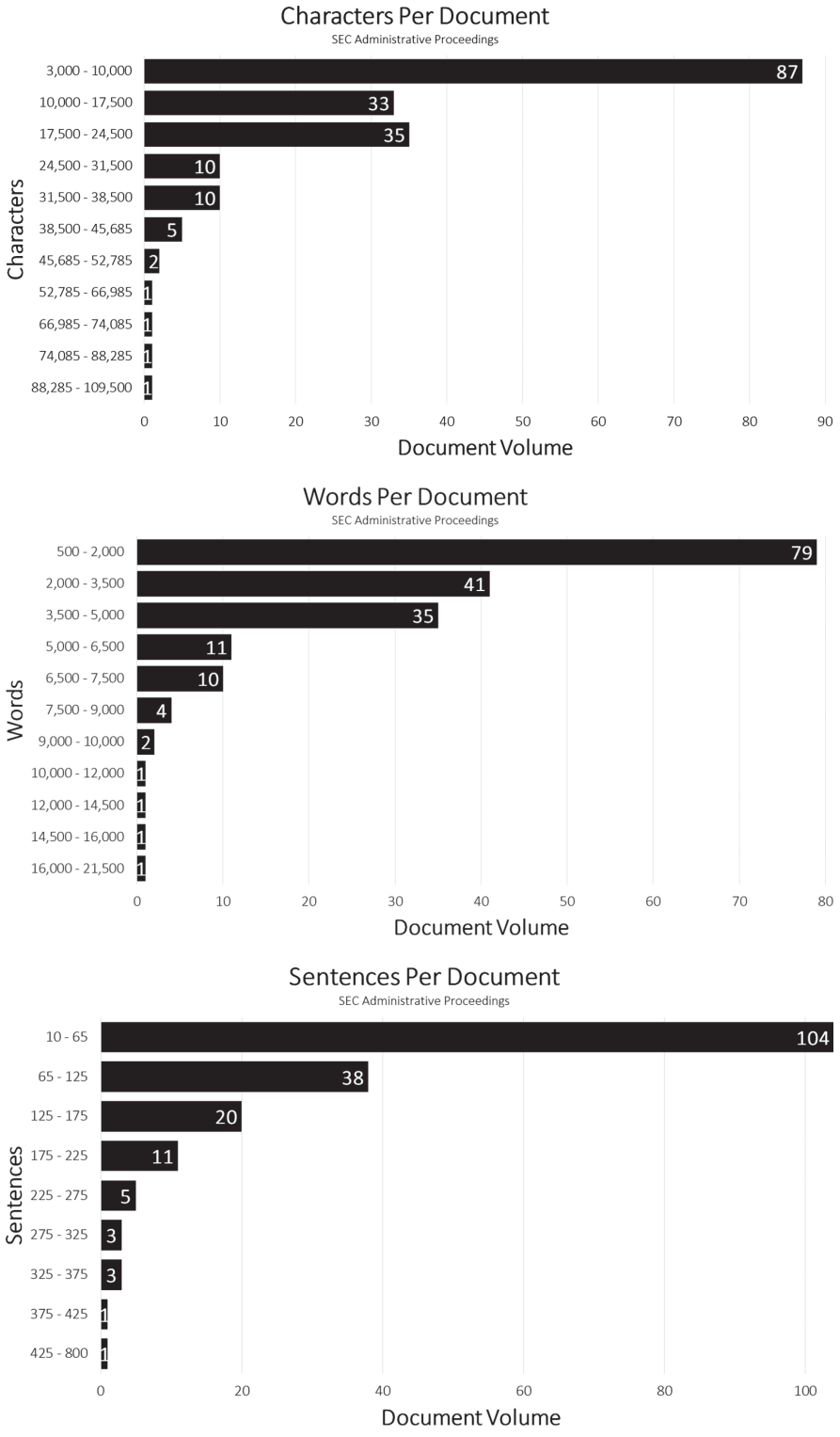

Figure 5 Data obtained from administrative proceedings with the US SEC between July and August 2019

Note: SEC, Securities and Exchange Commission.

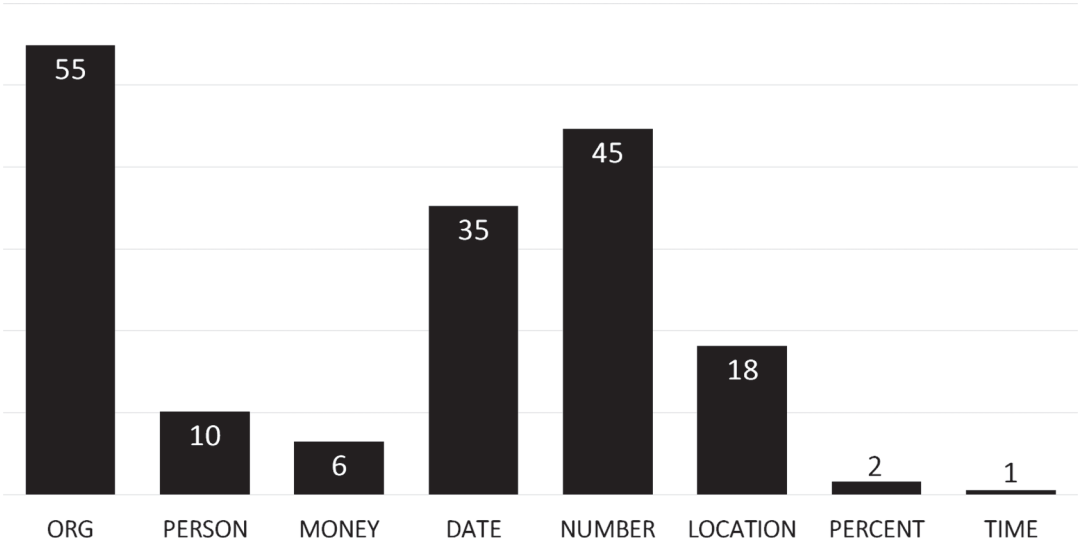

Figure 6 Named entity mentions per document

an extractive method, which attempts to identify the most important phrasing (eg sentences) in content. Another way is to use an abstractive method, which attempts to generate new text to represent its original input. It is important to understand that there is not always a 'right way' of doing things in this space but that some methods are simply better suited for specific use cases. For the following use case, we will use extractive method focusing on sentences; in order to do this, a mechanism is needed to rank each sentence by its importance.

Humans, especially subject matter experts, are very good at recognising important information, but this is quite challenging for a machine. Taking a page out of Google's PageRank,[19] one can apply this conceptually to words in sentences referencing words in other sentences (ie 'SentenceRank'). This assumes that if a word is mentioned quite often, it is likely important; likewise, if the focus sentence has that word, it is more likely to be important as well. There are some inherent issues with this assumption, most of which can be overcome with pre-processing the content and/or leveraging an inverse document frequency[20] to understand how common the language is across documents. For this case, performing some preprocessing will do the trick, so it is not necessary to employ the resource-intensive inverse document frequency exercise. Following are the preprocessing steps we will leverage:

- Get the data split out into sentence form. Here, it is key to retain the original sentence, chronological order in the document and a way to get back to it when finished. We will be applying some preprocessing to the text, and this will allow to retrieve/read the original sentences in lieu of transformed sentences.
- Lemmatise all words in the set of sentences to reduce inflectional forms of words (eg fine/fines/fined → fine). By transforming words to a common base form, we are afforded some resiliency against the

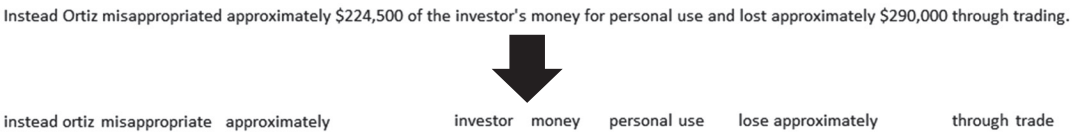

Figure 7 Before and after preprocessing example

inflectional nature of the English language.

- Perform some standard punctuation, numeric, short length and stop word removal from the document. This will help in removing common language and words that do not typically have major impact on the topic discussed in the content. At this point, Figure 7 shows an example sentence before and after preprocessing.
- Calculate the occurrence of each remaining word in the text. Some may refer to this as term frequency; one can interpret these term frequencies as 'popularity of words' within a document.
- Assign each term with points based on its popularity in the text ('term points'). Typically, the previous steps are applied at each individual document; this exercise provides us an aggregated view of the term frequencies. Table 2 shows a visual example.
- Scan back through each sentence and aggregate their term points to produce a sentence score ('Sentence Score'). Figure 8 shows an example of a pre-processed sentence score.
- Choose how many sentences to return. We will use the sentence score as a rank of top *n* sentences but also retain the order in which they came in the original text. Remember, we kept the original sentence and chronological order in the first step to enable this.

Executing these steps over the set of the SEC administrative proceeding documents, an example output for the five most important sentences from a single document looks like as shown in Table 3.

Table 2

Term	*Frequency*
fine	10
order	7
misappropriate	7
security	6
license	5
investor	5
invest	4
trade	3

This is conceptually helpful and, in most cases, a very good assistant to a human in reducing reading time. The monitoring activities of counterparties can consume countless hours. By reducing the amount of time needed to review content, a risk analyst can spend more time on core functions.

Sometimes, there are too many summaries to read, and from document-to-document, they are talking about different things. One option we have is to group (aka 'cluster') documents based on their content similarity. A couple of use cases for clustering documents is here:

- One simply wants to classify documents into groups for organisation purposes.
- One would like to have different subject matter experts (SME) review documents based on pairing the SME's knowledge with the content in the document (eg one SME reviews content related to the

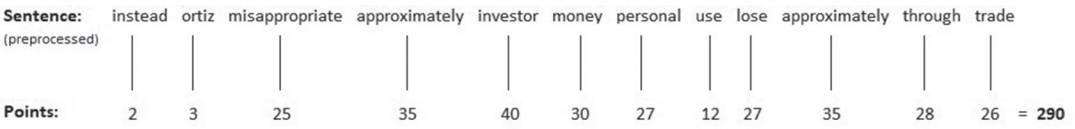

Figure 8 Preprocessed sentence score

Table 3: Text summary example from an SEC document

Item	*Value*
Prominent person	Gonzalo Ortiz
Text summary	'He has no securities licenses and has never been registered with a broker-dealer or associated with an investment adviser.' 'In May 2015, Ortiz solicited an investor with an offer to manage the investor's money.' 'The Commission's Complaint alleged that, inter alia, from May 2015 to May 2017 Ortiz convinced an acquaintance to give him control over nearly $570,000 of the investor's retirement savings based on materially false statements, including promises of a 50 per cent annual return and claims that Ortiz had previously been successful investing in stocks.' 'Ortiz proceeded to invest the funds and gave the investor a false account statement purporting to show that Ortiz had generated a return of over 50 per cent on the investor's money'. 'Instead Ortiz misappropriated approximately $224,500 of the investor's money for personal use and lost approximately $290,000 through trading.'

Note: SEC, Securities and Exchange Commission.

SEC admin proceedings, while another SME reviews content from other regulatory filings).

Leveraging the text summary output from each document, in lieu of using the entire document, one can be more specific about what content (ie the summaries) is used from each document during a clustering exercise. This is especially useful when there is text that occurs across documents but has minimal importance (ie noise).

We will walk through some steps to cluster our documents based on their text summaries are discussed here. One can think of this process as grouping documents into neighbourhoods based on their similarity. With the method used here, one can refer to documents as being neighbours, with their similarity being a measure of distance between them (must be calculated). Figure 9 is a visual representation of what we will be doing.

One can perform the following to start clustering the documents based on their respective text summaries:

- Convert each summary into term bit vectors.[21] This means that all words from all summaries are given a score of 1 or 0 as to whether they exist in the individual summary. As it has already been determined important, term frequencies are of a lesser concern.
- Calculate the matrix distance[22] using cosine as our distance function.[23] One could use a different function such as Euclidean, but cosine was used for this exercise.
- Apply k-medoids[24] clustering algorithm to minimise the average dissimilarity to

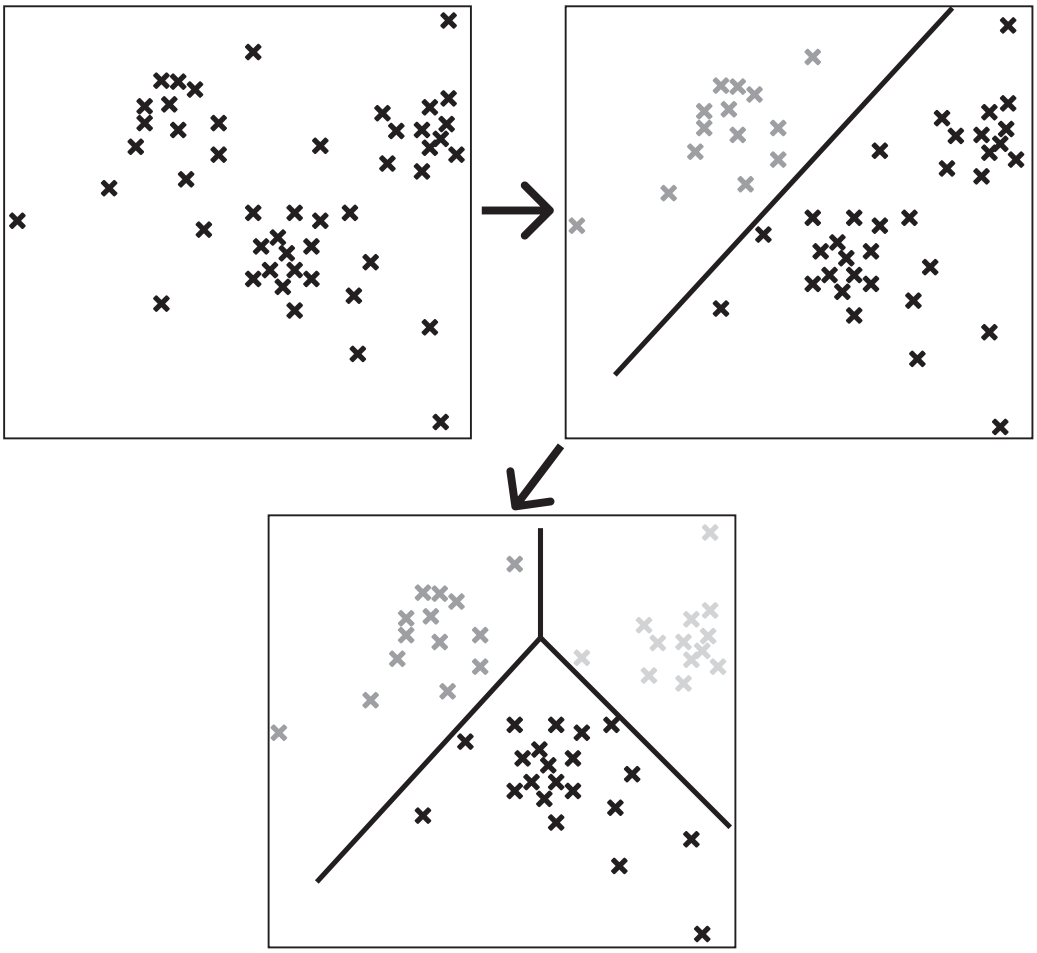

Figure 9 Text clustering example

all other text summaries in the respective cluster.

With the SEC administrative proceeding documents, Table 4 shows an example clustered neighbour (ie similar document) to the previous text summary example regarding Gonzalo Ortiz.

Determining who reviews or what class they are provided is ultimately up to the client/consumer, but these methods have shown to make a much better use of an analyst's time, allowing them to review considerably more content related to their risk-monitoring functions. For example, a risk analyst could have an over-abundance of content to review based on a firm's clients and yet, using text analytics, the analyst will have the ability to drastically reduce time spent monitoring an organisation's risk.

CONCLUSION

AI will reshape the middle and back office by automating settlement and risk-monitoring activities. Using AI techniques to scale, optimise, augment and automate these processes, firms will be able to exploit new opportunities. AI is not one size fits all, though; with a wide variety of algorithms and techniques available, those that are purposefully aligned with specific use

Table 4: Text summary example from an SEC document

Item	Value
Prominent person	Carol Ann Pedersen
Text summary	'Pedersen has never been registered with the Commission and has never held any securities licenses.' 'The Commission's complaint alleged that in connection with the offer or sale of securities and while acting as an investment adviser, Pedersen made materially false statements to advisory clients and investors, orchestrated a Ponzi scheme, misappropriated investor funds for her personal use, made misrepresentations to advisory clients and investors concerning how their money would be used and the risks associated with their investments, prepared and sent fake account statements indicating that investor funds were fully invested and earning returns, and otherwise engaged in a variety of conduct that operated as a fraud and deceit on advisory clients and investors.' 'On 20th March, 2019, Pedersen pled guilty to one count of wire fraud in violation of Title 18 United States Code, Section 1343 before the United States District Court for the Central District of California, in *United States v. Carol Ann Pedersen*.' 'She was sentenced to a prison term of 97 months followed by 3 years of supervised release and ordered to make restitution in the amount of $27,547,839.70.' 'The count of the criminal information to which Pedersen pled guilty alleged, inter alia, that Pedersen defrauded investors and obtained money and property by means of materially false representations, and that she used interstate wires to transfer investor funds.'

Note: SEC, Securities and Exchange Commission.

cases will be the most effective. With a focus on the settlement process, the authors believe firms can reduce errors and enable value-added opportunities by addressing fails better and faster, extending fails management across all asset classes and increasing additional balance sheet opportunities by maximising their use of collateral across the enterprise. With risk monitoring, firms can quickly act on a myriad of corporate actions and other regulatory or compliance events as they are announced by domestic and international entities, regulatory bodies and more.

References

(1) 'The impact of artificial intelligence in the banking sector & how AI is being used in 2020', BI, New York: BI, June 2019; 'The impact of artificial intelligence in the banking sector & how AI is being used in 2020', available at: https://www.businessinsider.com/the-ai-in-banking-report-2019-6 (accessed June 24, 2019).
(2) 'Fail', Investopedia, July 2019, available at: https://www.investopedia.com/terms/f/fail.asp (accessed 13th August, 2019).
(3) 'Reinforcement learning', Wikipedia, October 2019, available at: https://en.wikipedia.org/wiki/Reinforcement_learning (accessed 1st October, 2019).
(4) Sutton, R. S. and Barto, A. G. (2014–2015) Cambridge, MA. 'Reinforcement Learning: An Introduction', The MIT Press, p. 89.
(5) Park, J. (2015) 'Network flow problems', Stanford University, available at: https://web.stanford.edu/class/cs97si/08-network-flow-problems.pdf (22nd August, 2019).
(6) *Ibid*.
(7) 'Greedy algorithm', Wikipedia, October 2019, available at: https://en.wikipedia.org/wiki/Greedy_algorithm (accessed 3rd October, 2019).
(8) 'Stochastic diffusion search', Wikipedia, September 2019, available at: https://en.wikipedia.org/wiki/Stochastic_diffusion_search (accessed 3rd October, 2019).

(9) Stecanella, B. (2017) 'A practical explanation of a Naive Bayes classifier', available at: https://monkeylearn.com/blog/practical-explanation-naive-bayes-classifier/ (accessed 10th September, 2019).
(10) Liaw, A. and Wiener, M. (2001) 'Classification and regression by randomForest', *Forest*, Vol. 2/3, p. 22.
(11) 'Alternative data (finance)', Wikipedia, October 2019, available at: https://en.wikipedia.org/wiki/Alternative_data_(finance) (accessed 13th September, 2019).
(12) 'Administrative proceedings', U.S. Securities and Exchange, July–August 2019, available at: https://www.sec.gov/litigation/admin.shtml (accessed 13th September, 2019).
(13) *Ibid*.
(14) 'Named-entity recognition', Wikipedia, November 2019, available at: https://en.wikipedia.org/wiki/Named-entity_recognition (accessed 9th November, 2019).
(15) 'Administrative proceedings', ref. 12 above.
(16) 'Automatic summarization', Wikipedia, October 2019, available at: https://en.wikipedia.org/wiki/Automatic_summarization (accessed 3rd October, 2019).
(17) 'Topic model', Wikipedia, September 2019, available at: https://en.wikipedia.org/wiki/Topic_model (accessed 3rd October, 2019).
(18) 'Cluster analysis', Wikipedia, October 2019, available at: https://en.wikipedia.org/wiki/Cluster_analysis (accessed 3rd October, 2019).
(19) 'PageRank', Wikipedia, September 2019, available at: https://en.wikipedia.org/wiki/PageRank (accessed 3rd October, 2019).
(20) 'tf-idf', Wikipedia, September 2019, available at: https://en.wikipedia.org/wiki/Tf%E2%80%93idf (accessed 3rd October, 2019).
(21) 'Bit array', Wikipedia, May 2019, available at: https://en.wikipedia.org/wiki/Bit_array (accessed 3rd October, 2019).
(22) 'Distance matrix', Wikipedia, December 2018, available at: https://en.wikipedia.org/wiki/Distance_matrix (accessed 3rd October, 2019).
(23) 'Cosine similarity', Wikipedia, August 2019, available at: https://en.wikipedia.org/wiki/Cosine_similarity (accessed 3rd October, 2019).
(24) '*k*-medoids', Wikipedia, October 2019, available at: https://en.wikipedia.org/wiki/K-medoids (accessed 3rd October, 2019).