Quantitative Investment Strategies
Using Lend RX Technology Tweet Sentiment Analysis for Pharmaceutical Industry

Authors : Lend-Rx Technology SAS
hello@lend-rxtech.com
https://lend-rxtech.com

2022 March

Introduction

The healthcare sector, and in particular, the pharmaceutical sector has attracted the interest of both institutional and individual investors over the recent year. The exponential expansion of internet and social media is providing investors another source of information beside traditional sources like financial reporting and analyst ratings. Information from social media allows investors to access information more widely and measure the impact of information that could link to company's future performance. Though this source of information so rich, unstructured and noised for traditional investors can collect, handle and extract useful information to consistently generate profit.

Lend RX Technology's team has built a powerful framework to collect, analyze, classify and structure social media data, from Twitter in particular, related to the pharmaceutical&biotechnology industry. From this framework, we provide tweet sentiment analysis related to drugs as well as their relationship to company.

In this paper, we demonstrate and analyze a methodology to build a smart beta portfolio in US pharmaceutical&biotechnology industry using our sentiment analysis data as the driving factor.

The methodology presented in this paper is designed to follow traditional criteria of investment portfolio construction methodology such as sizing and diversification, auditability, interpretability and consistency under changing market condition as well as evolution of internet and social media usage.

It consists of creating smart-beta long portfolio of with a factor called **time-weighted average optimism intensity** of the portfolio is maximized under certain constraints. Simply speaking, it can be interpreted as allocating capital into stocks where the intensity of positive news dominates negative one recently.

The content of this paper is as following:

- Section 1: We will present Lend RX Technology sentiment analysis data, the way is structured, standardized and combined. We will formulate some sentiment metrics at the company and portfolio level using drug sentiments as well as their interpretation. This will motivate the portfolio construction methodology.
- Section 2: We explain the portfolio construction methodology in detail
- Section 3: We present essential result of backtest for our portfolio construction methodology over different topic and scope of stock
- Section 4: Conclusions and Perspective

1 Understanding of the data and formulation of sentiment indicators at stock and portfolio level

The objective of this section is to describe the drug sentiment indicators and generalize these indicators to the stock and portfolio level considering the decaying importance of news over time.

This formulation is to make driving factors based on them the investment portfolio is built in form of an optimization problem with constraints.

Believing that with complexity and overfitting comes instability, we do not use machine learning methods for building investment signals. Instead, we formulate investment decision factors with simple mathematics models that try to reflect human observation, reasoning and decision making. We will show that these formulations, using rolling statistics, automatically evolve with time and do not require significant learning period.

1.1 Sentiment indicators at drug, company and portfolio level

1.1.1 Tweet analysis: topics and sentiment probabilities

Each tweet collected is tagged to a specific drug. Our AI analyzes the tweet content to determine context and topics of the post and the probability of sentiments corresponding to each topic. Then an aggregate sentiment will be calculated as a weighted average of detected topics sentiment if the sentiment is **non neutral**.

There are 6 topics: REGISTRATION, CLINICAL_TRIAL, EFFICACY, MILESTONE, SAFETY and NEWS and a pseudo-topic named ALL_TOPICS which is the aggregation of topics.

For example:

TWEET_ID	1	2
TIMESTAMP	2021-12-09 00:07:08 UTC	2021-12-10 00:05:36 UTC
TEXT	Text 1	Text 2
DRUG_NAME	DRUG X	DRUG X
REGISTRATION_POSITIVE	0.9	0.0
REGISTRATION_NEUTRAL	0.0	1.0
REGISTRATION_NEGATIVE	0.1	0.0
EFFICACY_POSITIVE	0.7	0.1
EFFICACY_NEUTRAL	0.0	0.0
EFFICACY_NEGATIVE	0.3	0.9
POSITIVE	0.8	0.1
NEUTRAL	0.0	0.0
NEGATIVE	0.2	0.9

1.1.2 Daily drug sentiment intensities

We sum sentiment probabilities detected for each drug over each period of 24h, from 06:00:00 UTC of the previous day to 05:59:59 UTC of the value day. The aggregated values that we call in this paper daily drug sentiment intensities are tagged to the value day.

Using the above example, we obtain the daily sentiment intensities of the drug DRUG X on the value date 2021-12-10:

VALUE_DATE	2021-12-10
DRUG_NAME	DRUG X
REGISTRATION_POSITIVE	0.9
REGISTRATION_NEUTRAL	1.0
REGISTRATION_NEGATIVE	0.1
EFFICACY_POSITIVE	0.8
EFFICACY_NEUTRAL	0.0
EFFICACY_NEGATIVE	1.2
POSITIVE	0.9
NEUTRAL	0.0
NEGATIVE	1.1

Additionally, by summing positive, negative and neutral intensities, we obtain the daily volume of tweets for each topic.

1.1.3 Company daily sentiment intensities

By combining with daily drug company relationship data that tells us the partnership role of the company regarding the drug in question, we will compute daily company sentiment intensities.

For example, on the value date 2021-12-10, the company A has 3 drugs X, Y and Z with following sentiments.

VALUE_DATE	2021-12-10	2021-12-10	2021-12-10
DRUG_NAME	DRUG X	DRUG Y	DRUG Z
REGISTRATION_POSITIVE	0.9	4.5	0.0
REGISTRATION_NEUTRAL	1.0	2.0	0.0
REGISTRATION_NEGATIVE	0.1	0.5	0.0
EFFICACY_POSITIVE	0.8	0.0	7.5
EFFICACY_NEUTRAL	0.0	0.0	1.0
EFFICACY_NEGATIVE	1.2	0.0	3.5
POSITIVE	0.9	4.5	7.5
NEUTRAL	0.0	2.0	1.0
NEGATIVE	1.1	0.5	3.5

By summing all sentiments intensities over these drugs on that day, we obtain:

VALUE_DATE	2021-12-10
COMPANY	COMPANY A
REGISTRATION_POSITIVE	5.4
REGISTRATION_NEUTRAL	3.0
REGISTRATION_NEGATIVE	1.6
EFFICACY_POSITIVE	8.3
EFFICACY_NEUTRAL	1.0
EFFICACY_NEGATIVE	4.7
POSITIVE	12.9
NEUTRAL	3.0
NEGATIVE	5.1

1.1.4 Company daily sentiment polarity and optimism intensity

To measure the polarity of daily company sentiments, we introduce an indicator called "daily sentiment polarity", of which the formula is:

$$P_{company,topic,date} = \frac{I^{+}_{company,topic,date} - I^{-}_{company,topic,date}}{I^{+}_{company,topic,date} + I^{-}_{company,topic,date} + I^{0}_{company,topic,date}}$$

In this formula:

1. $P_{company,topic,date}$ is the daily sentiment polarity for the company considering sentiments of a specific topic.

2. $I^{sentiment}_{company,topic,date}$ factors are for sentiment intensities of the corresponding topics for the company on the same day. With $+$,$-$ and 0 stand for positive, negative and neutral respectively.

Not only the sentiment polarity is important, but the volume of tweets also plays an important role, especially to the volatility of the stocks. We try to capture both polarity and volume by a measure called **optimism intensity**.

$$O_{company,topic,date} = I^{+}_{company,topic,date} - I^{-}_{company,topic,date}$$

In this formula: $O_{company,topic,date}$ is the daily optimism for the company considering sentiments of a specific topic.

VALUE_DATE	2021-12-10
COMPANY	COMPANY A
REGISTRATION_POLARITY	0.38
EFFICACY_POLARITY	0.26
ALL_TOPICS_POLARITY	0.37
REGISTRATION_OPTIMISM	3.8
EFFICACY_OPTIMISM	3.6
ALL_TOPICS_OPTIMISM	7.8

The daily sentiment polarity indicates which sentiment dominates the most on the daily basis considering the overall tweet volume related to each topic. On the other hand, the optimism intensity measure how strong that domination is.

1.1.5 Company time-weighted sentiment polarity and optimism intensity

We suppose that good and bad news may affect stock performance for a certain period, news importance decays with time.

We try to formulate modified versions of sentiment polarity and optimism that consider the effect of time and use these time-weighted indicators as key factors for stock selection and portfolio construction.

From a point of view of optimal control, if the dependency of a system on its control variables is too much complex or non-continued, its behavior will be very unstable. We try to limit the number of hyper parameters (control variable) and make the dependency of the indicators on them most smooth as possible.

For simplicity, we suppose that the importance of daily sentiments decays exponentially with time. Hence we introduce the time-weighted sentiment polarity and optimism intensity.

$$\Pi_{company,topic,date_0} = \frac{\sum_{j=0}^{T} e^{-j\pi} P_{company,topic,date_{-j}}}{\sum_{j=0}^{T} e^{-j\pi}}$$

$$\Omega_{company,topic,date_0} = \frac{\sum_{j=0}^{T} e^{-j\omega} O_{company,topic,date_{-j}}}{\sum_{j=0}^{T} e^{-j\omega}}$$

In this formula:

- $\Pi_{company,topic,date_0}$ and $\Omega_{company,topic,date_0}$ are respectively the time-weighted sentiment polarity and optimism intensity for the company on the value date $date_0$ considering each topic.
- j is the number of day that we look back into the past.
- T is the maximum period of lookback. In our choice, $T = 55$ which is equivalent to 8 weeks of data.
- π and ω are the forgetting rate (or time exponential decay ratio). In our choice of modelling, $\pi = 0.05$ and $\omega = 0.2$.
- We have then $e^{-28\pi} \approx 0.25$ and $e^{-7\omega} \approx 0.25$.That means the importance of daily sentiment polarity of 4 weeks before will be 25%. For the daily optimism intensity, this needs only 1 week for the importance to reduce to 25%.

Convention for handle days with no sentiment detected

We consider that daily polarity and optimism equals to 0.0 for days with no sentiment and take these days into the calculation. This choice will smooth the time-weighted indicators betters and significantly reduces the importance of the choice for the maximum lookback period T.

Remarks

Note that by formulation, the optimism intensity is additive whilst the sentiment polarity is not. We will use sentiment polarity as a measure for stock picking and optimism intensity as a factor to optimize during portfolio construction.

1.1.6 Portfolio time-weighted optimism intensity

Given an investment portfolio of which the allocation is characterized by the asset weights $w_i, i \in \{1, ..., N\}$ (value in percentage with respect to the portfolio total value). We define the time-weighted) optimism intensity at portfolio level:

$$\Omega_{Portfolio,topic,date} = \sum_{i=1}^{N} w_i \Omega_{company_i,topic,date}$$

In this formula:

- $\Omega_{Portfolio,topic,date}$ is the (time-weighted) optimism intensity of the whole portfolio.
- w_i is the weight (in percentage) of the company i regarding portfolio value. $w_i > 0$ when we long the stock and $w_i < 0$ if we short it.
- $\Omega_{company_i,topic,date}$ is the time weighted optimism intensity of the stock on the calculation date.

Note that the data is delivered on a weekly basis each Monday, if the rebalance date is not Monday, we will also use data available on Monday for portfolio rebalancing during our simulation. This is to simulate the real-life context.

1.2 Some key statistics

1.2.1 Drug level

We will present here some keys drug sentiment statistics from 2016 until 2021.

We observe that the number of drugs mentioned in Twitter increased steadily over years. On the other hand, the tweet volume jumped in 2020 regarding the previous period, probably due to Covid-19. Therefore, the distribution of sentiment changed significantly between pre-Covid and post-Covid. This phenomenon would affect the performance of trading model calibrated (and usually overfitted) on the pre-Covid period.

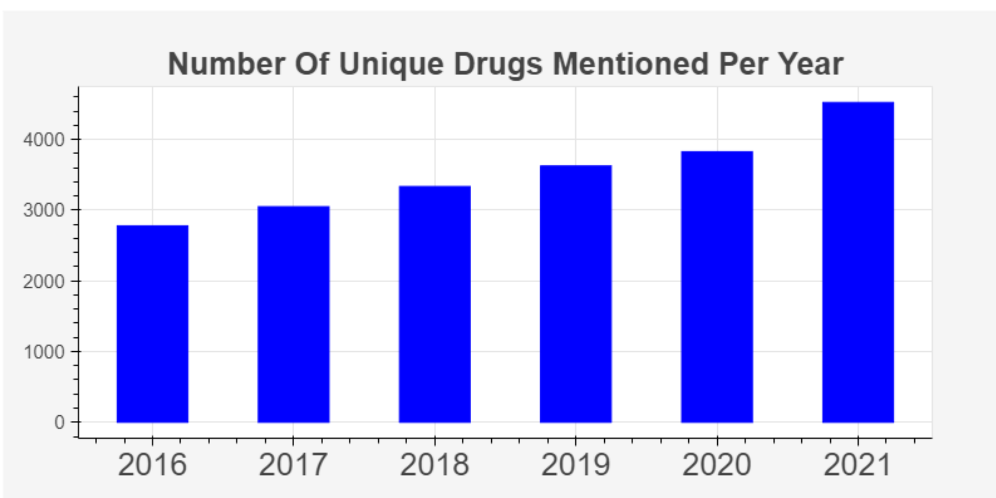

Figure. 1: Number of unique drugs that have been mentioned per year

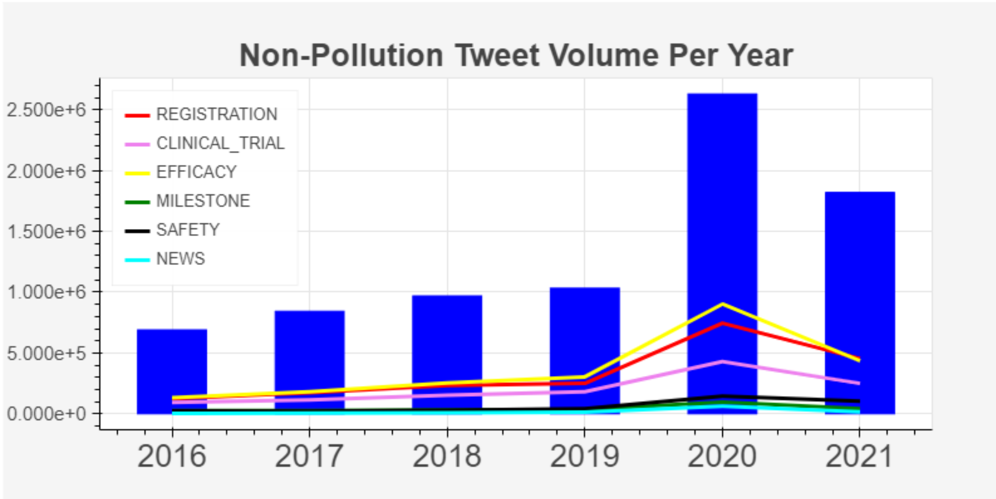

Figure. 2: Tweet volume per year and per topic

Table 1: Yearly topic frequency

FREQUENCY/YEAR	2016	2017	2018	2019	2020	2021
REGISTRATION_FREQUENCY	18.3%	20.6%	24.1%	24.3%	27.6%	28.3%
CLINICAL_TRIAL_FREQUENCY	13.6%	13.5%	15.7%	17.5%	16.1%	13.8%
EFFICACY_FREQUENCY	19.4%	21.5%	26.0%	29.4%	33.8%	25.5%
MILESTONE_FREQUENCY	3.6%	3.2%	3.2%	3.2%	3.7%	2.9%
SAFETY_FREQUENCY	3.4%	2.5%	3.1%	3.9%	5.5%	6.3%
NEWS_FREQUENCY	0.6%	0.5%	0.8%	1.2%	2.3%	1.3%

Table 2: Yearly non-neutral sentiment ratio

RATIO/YEAR	2016	2017	2018	2019	2020	2021
POSITIVE_RATIO	20.5%	20.4%	20.5%	21.1%	19.6%	16.5%
NEGATIVE_RATIO	3.8%	2.7%	3.2%	3.2%	4.8%	3.8%
REGISTRATION_POSITIVE_RATIO	49.1%	49.9%	42.7%	43.1%	35.5%	29.8%
REGISTRATION_NEGATIVE_RATIO	4.2%	2.7%	3.5%	3.5%	5.0%	6.1%
CLINICAL_TRIAL_POSITIVE_RATIO	24.1%	22.3%	22.2%	22.0%	21.3%	17.5%
CLINICAL_TRIAL_NEGATIVE_RATIO	15.5%	10.5%	9.6%	8.4%	13.3%	7.7%
EFFICACY_POSITIVE_RATIO	33.9%	27.2%	25.0%	23.8%	24.1%	21.8%
EFFICACY_NEGATIVE_RATIO	4.8%	3.1%	3.9%	3.8%	6.0%	4.7%
MILESTONE_POSITIVE_RATIO	76.3%	68.7%	59.9%	56.2%	51.4%	48.8%
MILESTONE_NEGATIVE_RATIO	0.5%	1.3%	1.1%	1.1%	1.1%	2.1%
SAFETY_POSITIVE_RATIO	6.2%	7.9%	9.0%	8.7%	5.8%	11.2%
SAFETY_NEGATIVE_RATIO	9.9%	15.8%	11.4%	8.9%	12.5%	6.0%
NEWS_POSITIVE_RATIO	37.8%	39.3%	40.3%	32.9%	32.4%	24.6%
NEWS_NEGATIVE_RATIO	7.2%	3.4%	6.2%	4.5%	8.8%	4.1%

Table 3: Yearly sentiment polarity (positive ratio – negative ratio)

POLARITY/YEAR	2016	2017	2018	2019	2020	2021
POLARITY	16.7%	17.7%	17.3%	17.9%	14.8%	12.7%
REGISTRATION_POLARITY	44.9%	47.2%	39.2%	39.6%	30.5%	23.7%
CLINICAL_TRIAL_POLARITY	8.6%	11.7%	12.6%	13.6%	8.0%	9.8%
EFFICACY_POLARITY	29.1%	24.1%	21.1%	20.0%	18.1%	17.1%
MILESTONE_POLARITY	75.7%	67.4%	58.7%	55.0%	50.3%	46.8%
SAFETY_POLARITY	-3.7%	-7.9%	-2.4%	-0.2%	-6.7%	5.3%
NEWS_POLARITY	30.6%	35.9%	34.2%	28.4%	23.5%	20.5%

1.3 Remarks and criteria for building a portfolio using sentiment data

We have seen that the positive sentiment dominates the negative sentiment statistically. In addition, the actual delivery schedule (on a weekly basis) does not allow us to react to bad news enough quickly to downside move of stocks, especially when pharmaceutical industry stocks are very volatile.

Hence we will present a long only portfolio methodology that, in expectation, would consistently outperform the pharmaceutical industry. Investors who prefer a long-short equity strategy can obtain the objective by combing the strategy presented with offsetting, for example, an appropriate market index to obtain a long-short equity strategy.

2 A portfolio construction framework using sentiment features as driving factor

2.1 Overview

From the point of view of this paper, using any fixed threshold value that do not evolve with time is considered as overfitting and causing potential instability in the outcome. Hence, we try to make the less usage of such factor as possible when creating the portfolio methodology.

The idea is to beat a market index by a long only strategy using **sentiment indicators as a stock picking method**, in addition criteria of investment universe and company maturity. After picking stocks that are considered as potentially good for being long. We will try to create a portfolio that invests in stocks where good news dominates bad news the most. In other words, we invest in stocks of which the sentiment detected are considered optimist. The portfolio optimism intensity which is a simple candidate for being a measure of optimism will be choose as the score to optimize.

We have first run individual topic to produce a unique investment signal as we suppose that each topic has its own impact on the investment decision. A first choice, we have preferred to build a meta-portfolio (portfolio of topic portfolio) instead of building a unique investment signal combining the different topics for example using machine learning technique because results are more interpretable for the purpose of a white paper.

Hence our methodology consists of following steps:

1. **FILTERING**: Filtering of drug and company that satisfy a set of criteria of maturity, concretely we have selected U.S companies with at least 3 drugs in at least phase 1 point in time historically. Drugs that failed or have been discontinued over time are no longer counted. Market capitalization could be an alternative to this filter.

2. **TOPIC**: Choosing a topic to test.

3. **SIGNAL**: Compute sentiment indicators for the chosen topic.

4. **UNIVERSE SELECTION**: Selection of a stock universe defined by an index. The reason is that stocks inside the same index will belong to the same range of market cap which evolve over time and share a common average dynamic. This will also help us to observe if our strategy could outperform the same market. The stock index constituents are obtained by a proxy which the holdings of iShares EFT that tracks the index. This information is publicly published and updated in our simulation every Saturday following the last open trading day of each month. This step is called universe selection.

5. **STOCK PICKING**: Selection of stocks based on the available distribution of sentiment polarity of all company after maturity filtering on historical rebalance dates up to the actual rebalance date. Concretely, we define a thresholding quantile for potentially being long on a stock. On the rebalance date, we select stocks with time-weighted sentiment polarity strictly greater than that quantile value. In our case, the quantile value is 50%. In other words, we will only consider being long on the stock if its sentiment polarity is strictly better than 50% of all historically observed values until the rebalance date. This step is called stock picking.

6. **PORTFOLIO OPTIMIZATION**: Build a portfolio objective function to optimize, a set of portfolio constraints and resolve the optimization problem. Concretely, the objective function is the time-weighted portfolio optimism intensity. In human language, we try to allocate our portfolio in stocks where the amount of good news recently dominates the amount of bad news in term of volume. The portfolio constraints are traditional ones: position sizing and capital usage.

7. **PORTFOLIO TUNING**: Eventual portfolio composition retuning: this is an optional step which is resizing the computed weights of stocking using their historical volatility while preserving the same capital usage computed in the step iv.

8. **STRATEGY COMBINATION**: As we have various topics and different stock indices, we will combine different strategies using a methodology that dynamically weighting the capital to considered strategies. This consists of create a portfolio of portfolio

(that we call meta-portfolio) considering historical volatility to build a risk adjusted allocation. If there is not enough data for calculating volatility, the component strategies are equally weighted. The objective is to know if we can obtain a strategy that have both better risk-return profile and interesting returns by combining various strategies.

2.2 Formulation of elementary investment strategies

Step 1. Drug and company filtering

Case	Drug maturity	Number of drugs	Description
1	>=Phase 1	>=3	Consider only U.S. company with 3+ drugs at least phase 1, discontinued/failed drugs are not considered

Step 2. Topic selection

Case	Topic	Description
1	ALL_TOPICS	Weighted average of all topics: Registration, Clinical Trial, Efficacy, Safety, Milestone and News
2	REGISTRATION	Post covering registration topic
3	CLINICAL_TRIAL	Post covering clinical trial results topic
4	EFFICACY	Post covering efficacy related topic
5	SAFETY	Post covering safety related topic

Step 3. Sentiment indicators calculation

Case	Rolling period	Sentiment polarity forgetting rate π	Optimism intensity forgetting rate ω
1	56	0.05	0.2

Step 4. Stock index selection

Case	Index	ETF	Description
1	NASDAQ Biotechnology	IBB	
2	S&P 500	CSPX	
3	NASDAQ Biotechnology Excluding S&P 500		Obtained by excluding S&P 500 constituents from NASDAQ Biotechnology

Step 5. Optimization constraints and objective function

Parameter	Value	Description
Time weighted sentiment polarity threshold	50%	Stocks with polarity on rebalance date greater than historical median (over all stocks) are potential long opportunities
Minimum number of historical prices datapoints for a stock	252	Exclude all stocks that have less than 1 year of market data as of the rebalance date
Minimum weight of long positions	0.0	Stocks that have negative or not enough good contribution of portfolio optimism intensity will be ignored
Maximum weight of long positions	Min(0.1,Max(0.05,1.5/ Number of picked stocks))	If there are more than 20 picked stocks, the max position weight is 0.05.If there are less than 15 picked stocks, max weight is 0.1
Maximum total weights of all long positions	100%	
Trading fee	0.05% of transaction value	Slippage, bid ask spread and commission are simulated by a small trading fee
Dividend tax	0%	We do not take tax into account

The optimal target portfolio composition is the solution of the following optimization problem:

$$(w_i^{optimal})_{i\,in\,I} = Arg(Max\left\{\sum_{i\,in\,I} w_i \Omega_{company_i,topic,latest\,release\,date}\right\})$$

With I being the set of picked stocks for potential long after all filtering steps and $(w_i^l)_{i\,in\,I}$ satisfying the above constraints.

Step 6. Eventually tuning position weights using inverse volatility

Case	Tuning strategy	Description
1	None	Directly use the weights computed by Step 5
2	Using inverse volatility	

$$w_i^{tuned} = \frac{w_i^{optimal}}{\sigma_i^{adjusted}} \times \frac{\sum_{j\,in\,I} w_j^{optimal}}{\sum_{j\,in\,I} \frac{w_j^{optimal}}{\sigma_j^{adjusted}}}$$

In this formula:

- w_i^{tuned} is the tuned weight for the stock i.
- $w_j^{optimal}$ is the computed weight for the stock j obtained by step 5.
- $\sigma_j^{adjusted}$ is the volatility of the stock j over recent 1 year adjusted by the following rule:

$$\sigma_j^{adjusted} = Max(\sigma_j, 0.2, Mean(\sigma_k, k\,in\,I, w_k > 0) - Std(\sigma_k, k\,in\,I, w_k > 0))$$

2.3 Formulation of meta strategies for combining different strategies

1. First we select a subset of strategy based on topics, stock index and optimization method to combine.
2. Secondly we combine the elementary strategies holdings using the weights of elementary portfolios with respect to the meta-portfolio. The weighting methodology is analogous to the step 6 of strategy formulation. We use inverse historical volatility of elementary strategies for weighting.

The step 2 requires at least 1 year of data for computing annualized strategy volatility (and not the volatility of the holdings portfolio, which may be another methodology to try). Therefore before having enough performance data, we give equal weights for elementary strategies.

$$W_p = \frac{1}{S_p^{adjusted}} \times \frac{1}{\sum_{q\,in\,P} \frac{1}{S_q^{adjusted}}}$$

Or $W_p = \dfrac{1}{Card(P)}$ if there is not enough historical data for computing strategy rolling volatility (i.e. 1 year).

In this formula:

- P represents the set of chosen elementary strategies.
- W_p is the weight of the strategy p on the rebalance date.
- $S_q^{adjusted}$ is the volatility of the strategy q over recent 1 year adjusted by the following rule:

$$S_q^{adjusted} = Max(S_q, 0.2, Mean(S_r, r\,in\,P) - Std(S_r, r\,in\,P))$$

2.4 Solver

We use Python 3 and the package SciPy to formulate and solve the optimization problems.

3 Backtest key results

3.1 Backtest conditions

BACKTEST PERIOD: The backtest starts on 2016 March 1st and ends on 2021 October 29th (as a result, performance statistics for these years are calculated on incomplete duration). The initial capital of all strategies is set at $10 million. There is no calibration period for hyper parameters. We started the live test with sentiment data available on the first rebalance date. After that, the sentiment history is incrementally enriched on each release date. The sentiment dataset starts on 2016-01-01, so we have 2 months for computing initial sentiment indicators statistics before the live backtest starts.

For the meta-strategy, we start weighting elementary strategies using strategy inverse volatility after 2017-03-01. Between 2016-03-01 and 2017-03-01, the elementary strategies are given equal weights.

REBALANCING: Note that the data is release on a weekly basis, on Monday. The rebalance date is first trading of the week. If the rebalance date is not Monday, we only use data available on the latest Monday to simulate the real production pipeline.

We supposed that the rebalancing process is realized just before the close of rebalance date.

The Adjusted Close is used for handling dividends and evaluate performance in the backtest.

FEES: We set the trading fee at 0.05% (apply to all transactions buy and sell) to approximate slippage and transaction fee during rebalancing. If a same stock is found in the portfolio before and after the rebalancing, the fee is applied on the necessary transaction to obtain the new target holding (i.e. the position delta).

3.2 Key results

3.2.1 Strategy level

Note that there is no learning period for to calibrate portfolio hyper parameters for these strategies and the same set of hyper parameters is applied over all topics and stock universes. Researchers have the possibility to calibrate parameters optimized for each topic and stock universe to obtain better strategies.

We found that:

1. Every topics except CLINICAL TRIAL topic outperform the benchmark IBB in general. More in details considering stock universe, NASDAQ BIOTECH and NASDAQ

BIOTECH MINUS biotech stocks also member of the SP500 shows the best returns. Large Capitalization ie stocks members of the S&P500 index have smallest but more consistent returns.

2. For all stock universes, REGISTRATION, ALL_TOPICS and EFFICACY topics give most return. However, REGISTRATION driven strategies are a little more volatile than EFFICACY and ALL_TOPICS. CLINICAL driven strategies are very volatile. SAFETY driven strategies have moderate returns as there are not many signals, however, the risk-return profile of SAFETY driven strategies are very good.

3. REGISTRATION, ALL_TOPICS and EFFICACY topics within NASDAQ BIOTECH or NASDAQ BIOTECH excluding SP500 outperform the benchmark in general with positive alpha. However, these strategies are more volatile than the benchmark due to concentration (the algorithm consists of picking about 20 stocks while IBB has about 200 stocks).

4. For some stock universe such as SP500, as there are not many stocks to select, the capital is not always used at 100% under common position sizing constraint. However, to avoid overfit, we do not try to optimize the position sizing constraint.

We will present the results of each topic for the cases where NASDAQ BIOTECH and NASDAQ BIOTECH excluding SP500 here.

NASDAQ BIOTECH INDEX, IBB and TOPICS cumulative performance

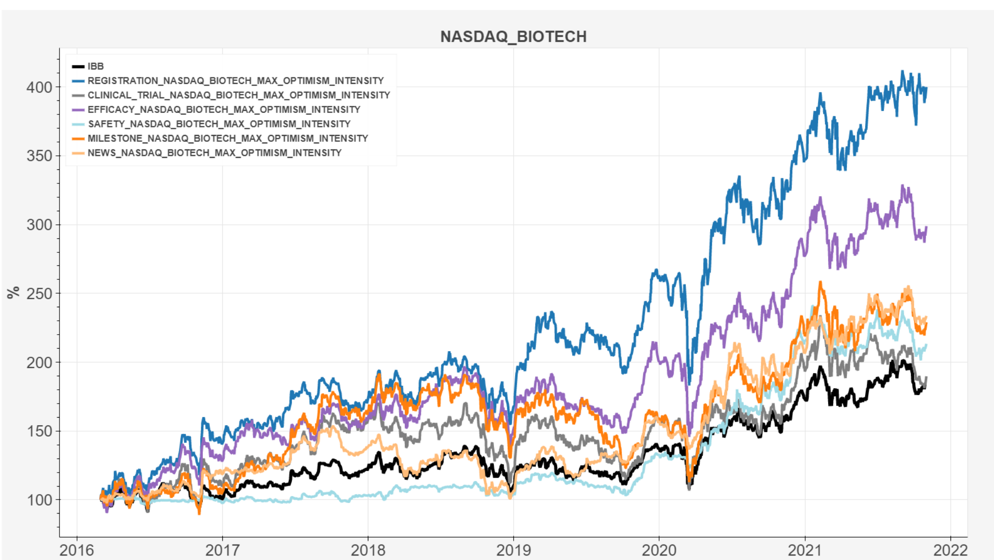

Figure. 3: Evolution of portfolio value of different topics with NASDAQ_BIOTECH (IBB) as stock universe

Table 4: Performance metrics of the different topics over the backtest period

STATS	**REGIS-TRATION**	**CLINICAL TRIAL**	**EFFICACY**	**SAFETY**	**MILES-TONE**	**NEWS**	**IBB**
RETURN_2016	40.40%	9.74%	27.89%	-2.22%	4.82%	18.46%	0.23%
RETURN_2017	19.01%	39.11%	20.81%	5.08%	50.78%	14.90%	18.90%
RETURN_2018	-4.73%	-23.47%	-6.40%	0.85%	-13.17%	-23.60%	-11.66%
RETURN_2019	56.29%	28.25%	30.96%	27.78%	9.65%	47.89%	23.84%
RETURN_2020	32.52%	22.84%	38.44%	63.72%	38.14%	32.32%	26.67%
RETURN_2021	14.92%	-3.19%	6.92%	-1.32%	3.36%	11.20%	6.24%
SHARPE	0.952	0.5103	0.7938	0.8724	0.6167	0.7913	0.5587
SORTINO	1.5011	0.8295	1.2631	1.259	1.0143	1.1781	0.909
MEAN_RETURN	29.13%	16.58%	23.80%	14.80%	19.74%	17.36%	13.62%
VOLATILITY	30.6%	32.48%	29.98%	16.96%	32.01%	21.94%	24.37%
DOWNSIDE_VOLATILITY	19.41%	19.98%	18.84%	11.75%	19.46%	14.73%	14.98%
BETA	1.1454	1.1819	1.1273	0.464	1.1776	0.7152	1.0
ALPHA	0.1318	-0.0013	0.0805	0.0819	0.0327	0.0742	0.0
MAX_DRAWDOWN	-31.59%	-37.48%	-32.26%	-16.49%	-41.16%	-34.39%	-26.64%
EXPECTED_SHORTFALL	-4.30%	-4.48%	-4.20%	-2.47%	-4.28%	-2.97%	-3.43%
CALMAR	0.8762	0.319	0.6597	0.8665	0.382	0.469	0.4217
CONSECUTIVE_LOSSES	8.0	7.0	7.0	7.0	8.0	8.0	8.0
CONSECUTIVE_WINS	8.0	10.0	7.0	10.0	8.0	11.0	9.0

NASDAQ BIOTECH EXCLUDING SP500 listed stocks (focus on small to mid-size biotech stocks)

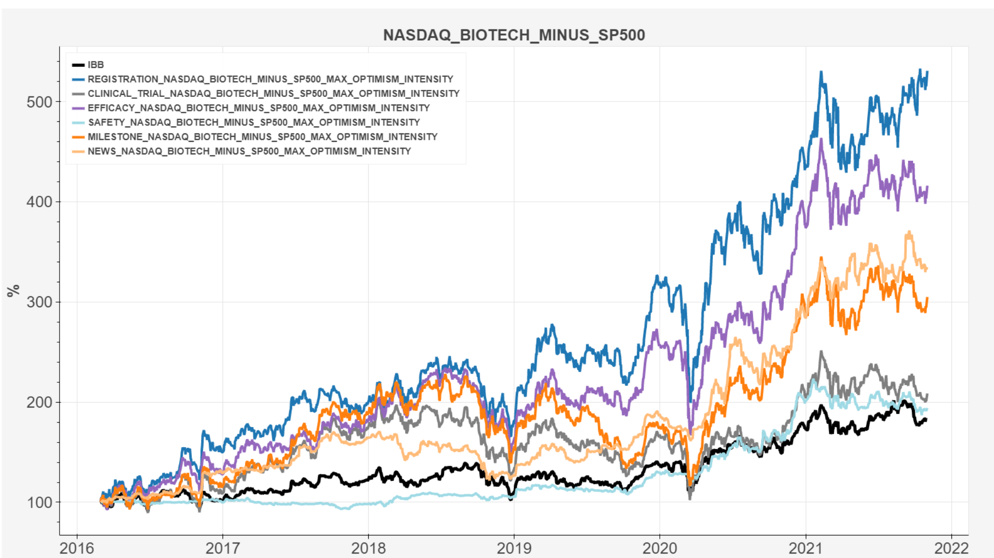

Figure. 4: Evolution of portfolio value of different topics with NASDAQ_BIOTECH_MINUS_SP500 as stock universe

Table 5: Performance metrics of the different topics over the backtest period

STATS	REGISTRATION	CLINICAL TRIAL	EFFICACY	SAFETY	MILESTONE	NEWS	IBB
RETURN_2016	49.61%	8.48%	27.61%	-0.16%	15.47%	28.95%	0.23%
RETURN_2017	30.47%	56.63%	39.30%	-1.98%	58.18%	26.70%	18.90%
RETURN_2018	-10.45%	-25.26%	-11.92%	7.38%	-19.32%	-22.68%	-11.66%
RETURN_2019	76.09%	23.96%	55.84%	23.07%	12.50%	47.96%	23.84%
RETURN_2020	36.80%	23.04%	45.83%	56.12%	62.74%	52.21%	26.67%
RETURN_2021	21.92%	0.55%	10.53%	-4.32%	5.67%	15.33%	6.24%
SHARPE	1.015	0.5435	0.9045	0.8572	0.7316	1.1563	0.5587
SORTINO	1.5946	0.8887	1.4304	1.2478	1.2058	1.6393	0.909
MEAN_RETURN	35.59%	19.24%	31.03%	12.81%	25.97%	23.33%	13.62%
VOLATILITY	35.06%	35.40%	34.30%	14.94%	35.50%	20.17%	24.37%
DOWNSIDE_VOLATILITY	22.32%	21.65%	21.69%	10.27%	21.54%	14.23%	14.98%
BETA	1.2547	1.246	1.2271	0.3795	1.256	0.55	1.0
ALPHA	0.1799	0.016	0.1377	0.0738	0.0826	0.1573	0.0
MAX_DRAWDOWN	-39.03%	-49.15%	-38.94%	-16.27%	-49.05%	-28.59%	-26.64%
EXPECTED_SHORTFALL	-5.06%	-4.81%	-4.80%	-2.13%	-4.78%	-2.65%	-3.43%
CALMAR	0.8764	0.2815	0.7336	0.7623	0.4427	0.8299	0.4217
CONSECUTIVE_LOSSES	8.0	7.0	7.0	7.0	8.0	8.0	8.0
CONSECUTIVE_WINS	9.0	10.0	8.0	9.0	9.0	10.0	9.0

3.2.2 Meta-strategy: Combining Registration and Efficacy topics

The objective of this section is not to find the best combination but to demonstrate that a smart combination of different strategies can generate a new strategy that suits specific investment criteria better. We will demonstrate that combining different topics and different stock universe can help obtain more consistent performance.

In this example, we intend to build a meta strategy that is more diversified then less volatile but still give a decent performance.

This meta strategy example consists of combining the 6 strategies with Registration and Efficacy as topics, Nasdaq Biotech excluding SP500 and SP500 as stock universe (filter of market capitalization). For each combination topic – stock index, we use 2 optimization methods: maximum optimism intensity and volatility adjusted maximum optimism intensity. The second method will allocate more capital on stocks that are less volatile, however, we found that the performance is more important and volatile than the first one. That means the news change the volatility of the stock (jump of volatility). As SP500 based strategies do not use 100% of capital, we use only one optimization method.

The backtest is run under the same period and trading fee hypothesis.

Table 6: Elementary strategies to build the meta strategy

Number	Stock universe	Topic	Optimization method
1	Nasdaq Biotech minus (i.e excluding) SP500	Registration	Maximum Optimism Intensity
2	Nasdaq Biotech minus (i.e excluding) SP500	Registration	Maximum Optimism Intensity with Volatility-adjusted weights
3	Nasdaq Biotech minus (i.e excluding) SP500	Efficacy	Maximum Optimism Intensity
4	Nasdaq Biotech minus (i.e excluding) SP500	Efficacy	Maximum Optimism Intensity with Volatility-adjusted weights
5	SP500	Registration	Maximum Optimism Intensity with Volatility-adjusted weights
6	SP500	Efficacy	Maximum Optimism Intensity with Volatility-adjusted weights

As a result, we obtain a strategy with decent return, low beta (0.85), reasonable alpha (0.16) and more limited downside risk as shown below.

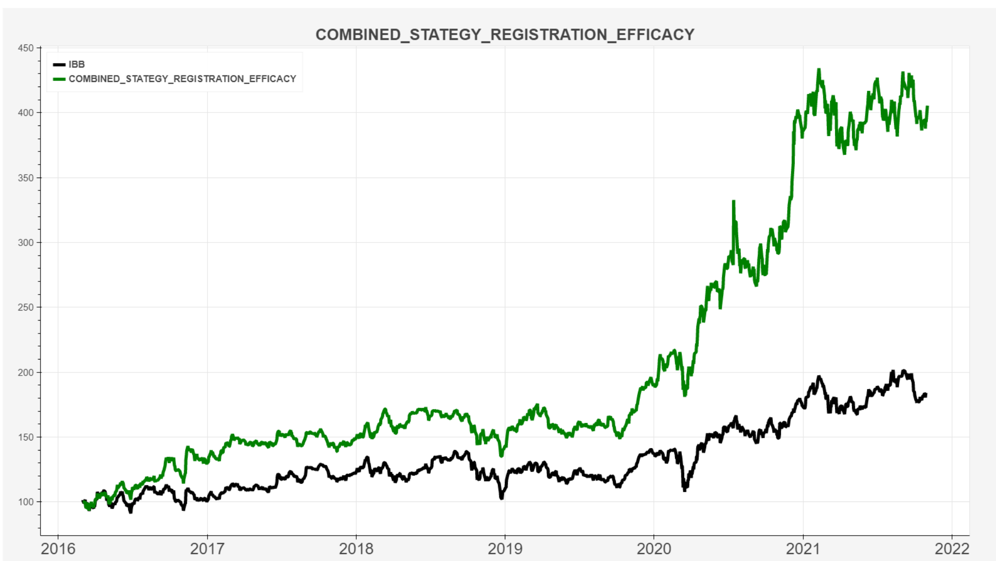

Figure. 5: Evolution of portfolio value of the meta strategy (Initial portfolio value: 10$M)

Table 7: Detailed performance statistics of the meta-strategy

STATS	COMBINED STRATEGY REGISTRATION EFFICACY	IBB
RETURN_2016	29.41%	0.23%
RETURN_2017	12.95%	18.90%
RETURN_2018	-5.16%	-11.66%
RETURN_2019	33.31%	23.84%
RETURN_2020	103.70%	26.67%
RETURN_2021	4.48%	6.24%
MEAN_RETURN	27.81%	13.62%
VOLATILITY	25.13%	24.37%
DOWNSIDE_VOLATILITY	17.01%	14.98%
SHARPE	1.1067	0.5587
SORTINO	1.6354	0.909
BETA	0.8527	1.0
ALPHA	0.1566	0.0
MAX_DRAWDOWN	-21.87%	-26.64%
EXPECTED_SHORTFALL	-3.29%	-3.43%
CALMAR	1.2794	0.4217
CONSECUTIVE_LOSSES	8.0	8.0
CONSECUTIVE_WINS	8.0	9.0

4 Conclusion and perspectives

We have seen that a simple portfolio construction method combining traditional risk management constraints and investment decision based on sentiments produced by our NLP can give returns outperforming the market index.

The nature of this investment decision making rule is following the social media sentiments. Adding topic segmentation to semantic analysis to deliver sentiment by topic have been developed to mimic fundamental investors bottom-up approach.

Some categories of sentiment like topics around **registration, efficacy or milestones** have effects that last longer as they are directly related to the business revenue. Therefore, strategies based on those categories have more consistent outperformance to the benchmark index.

Strategy around **clinical trials results** involve more anticipation from investors, therefore the effect of those binary event last shorter and price adjustment can't be capture by our approach. A weekly rebalance schedule is not quick enough to capture the sentiment trend in time, this explains why strategies based on this topic is more volatile and overall doesn't bring any alpha.

As we did not combine the news with the stock prices to optimize the decision making nor build a complex model to fine tuning the performance, there are possibility to improve the investment strategies.

Terms

- **Tweet sentiment and topics**: The analysis result for each tweet corresponding to a specific drug containing topics detection, sentiment probability and an aggregate sentiment over all detected topics.
- **Daily drug sentiment intensities**: The sum of sentiment probabilities of all tweets corresponding to each drug over a period of 24h.
- **Daily company sentiment intensities**: The sum of daily drug sentiments of all considered drugs belonging to the company.
- **Daily company sentiment optimism intensity**: The difference between the polarity of daily company sentiments and the volume of tweets.
- **Value date**: The timestamp to daily drug sentiment. All sentiment probabilities of tweets published from 06:00:00 UTC of the previous date to 05:59:59 UTC of the value date will be aggregated to obtain the daily drug sentiment intensities.
- **Drug owner and phase**: The daily metadata that show the relationship between a drug and a company as well as the current phase of the drug as of each value date.
- **Release date**: The time stamp of weekly data delivery. It is systematically fixed at 12:00:PM UTC each Monday.
- **Rebalance date**: The date when investment portfolio weights are recomputed. For this paper, the rebalance trade are realized just before the market close of the rebalance date. Rebalance dates are fixed on the first trading day of each week. On rebalance day, only drug sentiment data delivered on the latest release date is available.
- **Strategy**: A method of build portfolio which is characterized by sentiment factors specified by topic, and period of observation, an information forgetting rate, a stock universe, a set of portfolio constraints and an objective function.
- **Meta-portfolio**: An investment portfolio weighted combination of portfolio (which paper the holdings governed by each strategy) (i.e., portfolio of portfolios).
- **Meta-Strategy**: A method of combining portfolios governed by different strategies to obtain a final portfolio that supposed to be better in term of risk-return profile (i.e., strategy of strategies).
- **Annualized return**: The average daily percentage return annualized by multiplying by 252.
- **Annualized volatility**: The volatility of daily percentage return annualized by multiplying by $\sqrt{252}$.
- **Annualized downside volatility**: The daily volatility of negative percentage return annualized by multiplying by $\sqrt{252}$.

5 Addendum

Authors

- **Olivier LEHERLE**, Chief Executive Officer, Founder.
- **Khoi Nguyen NGUYEN**, Chief Scientific Officer.
- **Anaëlle PIERREDON**, NLP Engineer.

Reaching Lend-Rx Technology

Lend-Rx Technology SAS

124 rue Saint Dominique

75007 Paris France

Phone: 00 336 5089 90 73

Mail : hello@lend-rxtech.com

https://lend-rxtech.com

Disclaimer

This White Paper is not intended for the purposes of making a decision to carry out a transaction or trade. Nor does it provide any form of investment advice, or make any recommendations regarding particular financial instruments, investments or products. Lend-Rx Technology may discontinue or change the White Paper content at any time, without notice. Lend-Rx Technology does not guarantee or warrant the accuracy completeness or timeliness of the White Paper. For more detailed disclaimer, please refer to the Terms of Use present on Lend-Rx Technology website https://lend-rxtech.com