Vectorspace AI

vectorspace.ai

whitepaper 0.27
11/05/2017

C. Pate, M. Muldoon, R. Podowski, S. Hassan, K. Franks
info@vectorspace.ai

Introduction

The Vectorspace AI platform enables intelligent dynamically generated “token baskets” based on user-selected trends that exist in search, social media and news. The reason we’ve included the word ‘intelligent’ is based the ability for these baskets to determine for themselves whether or not to include additional cryptocurrencies or components from related baskets that may increase overall returns. Baskets that interact this way with one another will conduct these kinds of transactions between one another using the Vectorspace utility token, VEC which is also required to dynamically generate baskets.

Components within the dynamically generated baskets can be made up of cryptocurrencies along with other new innovations for stores of value in the cryptocurrency space that may arise in the future. Traders, investors and funds can optimize basket returns by applying filters such as technical analysis indicators and custom algorithms. Human researchers and curators of baskets are allocated VEC for their work in this area. The VEC utility token can be used to purchase blocks of dynamic basket generations in addition to subscriptions to higher level features offered by the platform including transacting baskets as trades on exchanges. VEC utility tokens are awarded to creators of top performing baskets which are displayed on a leader board each month.

A growing number of organizations offer the ability to trade a basket or group of cryptocurrencies with a single transaction. Similar to a traditional ETF (Exchange-Traded Fund), “token baskets” as they’re more commonly called, enable a fund or individual investor to spread their risk, diversify and lower volatility while maximizing gains with more safety and stability.

This approach is valuable but incomplete. For example, taking the top 20 best performing cryptocurrencies and placing them in a basket or fund minimizes gains compared to enabling fine-grained custom creation of token baskets.

Which groups of cryptocurrencies are beating the market? Why? What do they have in common? How strong are those relationships? Are those relationships well known and obvious or are they hidden relationships? Are the relationships numerical or non-numerical based on concepts, context and sentiment or a combination of both?

These are the questions Crypto Discover is designed to answer. Vectorspace AI’s Crypto Discover platform is an advanced cryptocurrency discovery engine that enables a user to automatically generate a token basket or “mini-index fund” made up of cryptocurrencies that are related to trends in news, global and local searches, concepts, context, keywords, hashtags, social media, URLs and other dynamic content.

The process for trading baskets of cryptocurrencies or token baskets is fairly straightforward. The more difficult process is enabling a position trader, swing trader, day trader, long or short term individual investor or fund manager to generate customized token baskets based on any criteria they choose and for long or short positions.

Automatic customized generation of token baskets, mini-index token funds or sectors based on any trend, concept, keyword, hashtag, URL or news story with optional filtering represents the future we see for the global cryptocurrency market. This can only be done with advanced Natural Language Processing (NLP), sentiment analysis, variants of deep learning combined with incentivized human curation.

Vectorspace has begun execution in this area as shown by our public facing demo of this technology located here: http://vectorspace.ai – It is our aim to advance this product for commercial grade use which includes a utility token used to reward cryptocurrency fund researchers and basket curators along with providing access to high-level token basket creation services, customized filters, indicators, algorithms and transactions.

Background and Significance

A single trend represented by a concept, keyword, hashtag, URL or news story can represent a network of cryptocurrencies based on their relationship to one another and the context or concepts that surround that trend. This relationship network of cryptocurrencies can represent a tradable token basket or closely related group of cryptocurrencies that have known and hidden symbiotic, parasitic and sympathetic relationships. They may trade in a group and move up and down together. They might all be impacted negatively or positively by the news that contains a particular set of keywords, context, concepts or trends.

Cryptocurrencies, like other trading vehicles such as stocks, at times trade in sympathy with others. Historically, traders have called these "sympathy plays", however, many of them are based on hidden or relatively unknown relationships. These relationships change over time and are described in whitepapers, cryptocurrency descriptions, company profiles, news releases, editorials, social media and other forms of public and private content and databases.

In part, the platform is based on building large, rich and ranked feature attributes located in mathematically defined vectors. Comparing these vectors for similarity is almost as important as how one constructs the vectors in vector space. Mimicking the way a human might manually construct a vector remains key.

Relying on vector similarity as opposed to direct keyword matches is an approach that's taken. Utilizing statistical and probabilistic approaches as opposed to "words and rules", the system is able to construct relationship networks between cryptocurrencies by using more advanced forms of Natural Language Processing (NLP) combined with approaches in sentiment analysis, Reinforcement Learning (RL), Recurrent Neural Networking (RNNs), a variant approach in deep learning.

Traders, investors and hedge funds can engage is quick information arbitrage using the platform. For example, if a cryptocurrency spikes up 50% in minute, you can insert its symbol, a trend, concept, keyword, hashtag, URL or news story related to the cryptocurrency that ran up and then, in seconds, have an automatically generated token basket of related cryptocurrencies, a tradable targeted token basket of cryptocurrencies that all have sympathetic, symbiotic and parasitic relationships with one another. This can be done faster than any manual research effort can uncover these connections. A rising tide lifts all boats or a lowering tide lowers them.

The algorithm and system were developed to capture events described in a paper written by Gur Huberman titled "Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar" detailing what happened when Entremed (ENMD) rose from 12/share to 85/share overnight. Stocks that had known and unknown (or hidden) relationships with Entremed also rose but a bit later. Good profits were made from those that were able to identify those relationships before others.

Excerpt from the paper:

"That news about a breakthrough in cancer research affects not only the stock of a firm that has direct commercialization rights to the development is not surprising; the market may recognize potential spillover effects and surmise that other firms may benefit from the innovation. Moreover, the market may interpret the news as good for other firms because it may suggest that the research and development conducted by these other firms is closer to commercial fruition. However, the news did not break on May 4, 1998, but on November 27, 1997. And the people with the expertise to evaluate the spillover effects closely follow the news within the scientific community, probably read Nature, and pay attention to the coverage of biotechnology in the Times even when the relevant material appears well inside the newspaper.

The motivation and identity of the people who traded the seven stocks so aggressively on May 4 is puzzling. If they are experts on the fundamental aspects of biotechnology, they could and should have traded five months earlier. If they are stock market experts with no special understanding of biotechnology, it is unclear how they picked these particular seven stocks. Perhaps they speculated on noise trader behavior, but why with these stocks?"

This kind phenomenon happens on the long and short side. Identifying "sympathetic" non-obvious relationships before most others is a form of information arbitrage. Having a system that automatically identifies hidden or non-obvious relationships is also key.

Other examples include Merck (MRK) dropping from 44.07 per share to 33 per share or 26.7% September 30th 2004 (https://finance.yahoo.com/quote/MRK/history?period1=1096095600&period2=1096959600&interval=1d&filter=history&frequency=1d) due to the Vioxx debacle (http://www.nytimes.com/2011/11/23/business/merck-agrees-to-pay-950-million-in-vioxx-case.html). Vioxx was a drug Merck produced which caused heart attack, stroke and death. Pfizer (PFE) had a hidden relationship with Vioxx at the time due to development of COX2 and Coxib inhibitors (Vioxx related) in their pharmaceutical pipeline. Sure enough, PFE drops from a day high of 29.10 on Dec. 16th to a day low of 21.99 per share or 24.4% Dec. 17th (https://finance.yahoo.com/quote/PFE/history?period1=1096441200&period2=1104480000&interval=1d&filter=history&frequency=1d) when the research analysts and the market manually figured out what symbiotic, parasitic and sympathetic hidden relationship existed between Merck and Pfizer (http://shareholdersfoundation.com/case/pfizer-inc-nyse-pfe-investor-securities-class-action-lawsuit-12152004). The Vectorspace Crypto Fund Creation platform can inform one of those hidden relationships in seconds and instantly providing an opportunity for unique, powerful and profitable information arbitrage strategies.

It's well known that trading vehicles can have symbiotic, sympathetic and parasitic relationships that sometimes traders, investors and fund managers are not completely aware of. This is where advanced discovery algorithms and technology that operate on "concepts" and "context" can add real value and opportunity.

Price correlations or Betas are important. However, these cannot be relied on alone. Time and again we can observe a single trading vehicle move up significantly based on news, a new blockchain protocol, cryptocurrency application, revenue, high-level partnership, winning a contract, an outright buy-out or just plain irrational exuberance and then witness minutes and sometimes days later, that a basket of related cryptocurrencies will also begin moving up. The delay obviously is where the money is made if you can position before others.

Optimizing Basket Clusters

How you fine tune your token baskets can also be customized. Once a token basket is generated additional filters can be applied to fine tune returns. For example, filters can be in the form of technical analysis indicators such as Moving Average Convergence Divergence (MACD) where only oversold or overbought cryptocurrencies within a basket are selected to trade. Other indicator filters might include:

- Acceleration Bands (ABANDS)

- Accumulation/Distribution (AD)
- Average Directional Movement (ADX)
- Absolute Price Oscillator (APO)
- Aroon (AR)
- Aroon Oscillator (ARO)
- Average True Range (ATR)
- Volume on the Ask (AVOL)
- Volume on the Bid and Ask (BAVOL)
- Bollinger Band (BBANDS)
- Band Width (BW)
- Bar Value Area (BVA)
- Bid Volume (BVOL)
- Commodity Channel Index (CCI)
- Chande Momentum Oscillator (CMO)
- Double Exponential Moving Average (DEMA)
- Plus DI (DI+)
- Directional Movement Indicators (DMI)
- Ichimoku (ICH)
- Fill Indicator (FILL)
- Keltner Channel (KC)
- Linear Regression (LR)
- Linear Regression Angle (LRA)
- Linear Regression Intercept (LRI)
- Linear Regression Slope (LRM)
- Max (MAX)
- Money Flow Index (MFI)
- Midpoint (MIDPNT)
- Midprice (MIDPRI)
- Min (MIN)
- MinMax (MINMAX)
- Momentum (MOM)
- Adaptive Moving Average (AMA)
- Exponential (EMA)
- Moving Average Convergence Divergence (MACD)
- Simple Moving Average (SMA)
- T3 (T3)

- Triple Exponential Moving Average (TEMA)
- Triangular Moving Average (TRIMA)
- Triple Exponential Moving Average Oscillator (TRIX)
- Weighted Moving Average (WMA)
- Normalized Average True Range (NATR)
- On Balance Volume (OBV)
- Price Channel (PC)
- PLOT (PLT)
- Percent Price Oscillator (PPO)
- Price Volume Trend (PVT)
- Rate of Change (ROC)
- Rate of Change (ROC100)
- Rate of Change (ROCP)
- Rate of Change (ROCR)
- Relative Strength Indicator (RSI)
- Parabolic Sar (SAR)
- Session Cumulative Ask (SAVOL)
- Session Cumulative Bid (SBVOL)
- Standard Deviation (STDDEV)
- Stochastic (STOCH)
- Stochastic Fast (StochF)
- Session Volume (S_VOL)
- Time Series Forecast (TSF)
- TT Cumulative Vol Delta (TT CVD)
- Ultimate Oscillator (ULTOSC)
- Volume At Price (VAP)
- Volume Delta (Vol Δ)
- Volume (VOLUME)
- Volume Weighted Average Price (VWAP)
- Williams % R (WillR)
- Welles Wilder's Smoothing Average (WWS)

Custom algorithms can be applied as well. Once you’re satisfied with the structure of a refined token basket and have set filters and scoring thresholds to control whether certain cryptocurrencies qualify you can then extract the cryptocurrencies and trade them as baskets or use the baskets as signals or indicators for other trades. The list of things you can do with this kind of approach is largely based on your creativity and how well you might be able to optimize using filters and multiple long and short baskets. Optimizing the returns of each basket can be based on which concepts to use or data points such as the size of a cryptocurrencies short interest or float (available and tradable on the open market also known as circulating supply).

You might notice a few interesting traits in the way these related clusters of cryptocurrencies trade as baskets. Sometimes there’s divergence from a broader index or general market. This can offer opportunities for arbitrage. Just as single cryptocurrencies such as Bitcoin can correlate to global search trends for the term “bitcoin”, entire baskets can correlate to a single popular search trend with the search trends serving as leading indicators for positive price activity as shown in the figure below:

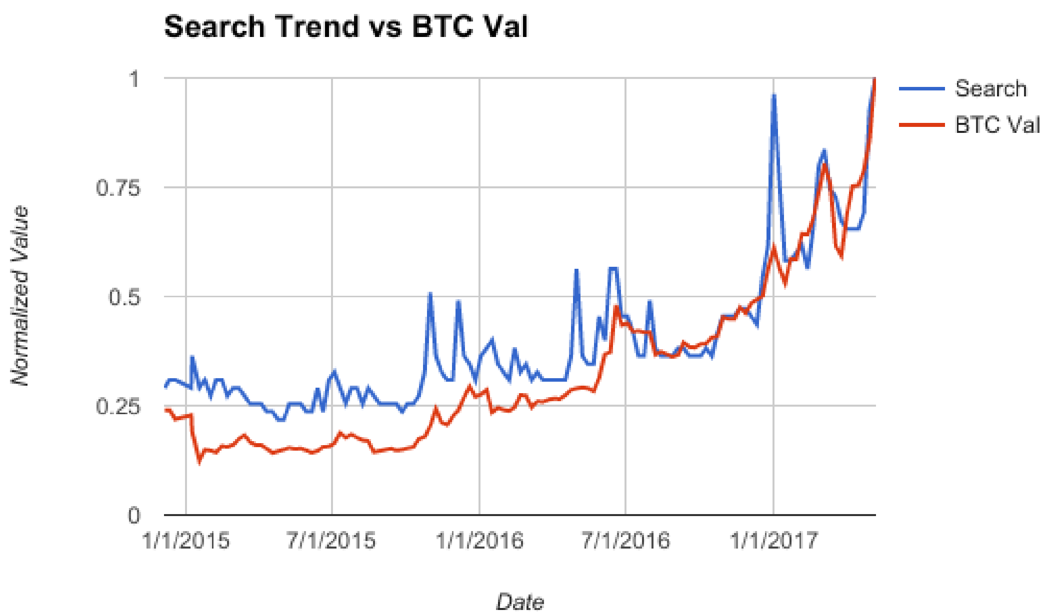

Figure 1. A single global search trend for “bitcoin” and its correlation to the price of BTC alone. (source: Google Trends)

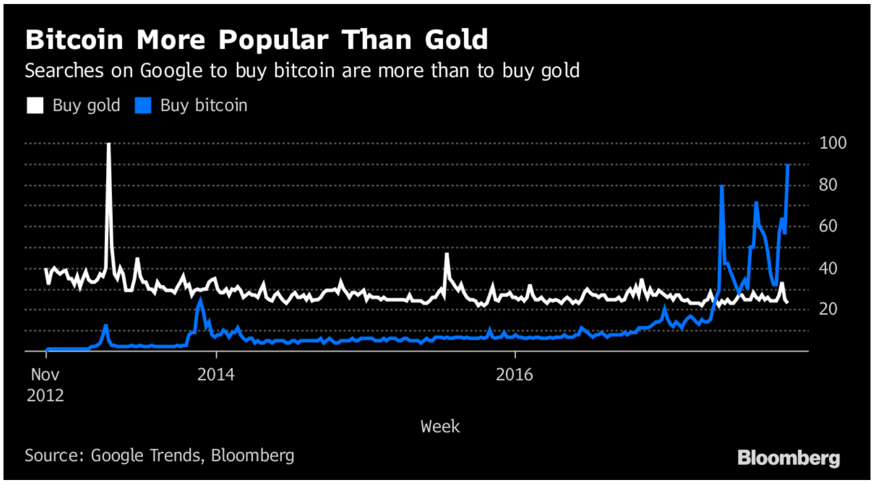

Figure 2. Global search trends for "buy gold" compared to "buy bitcoin". (source: Google Trends, Bloomberg)

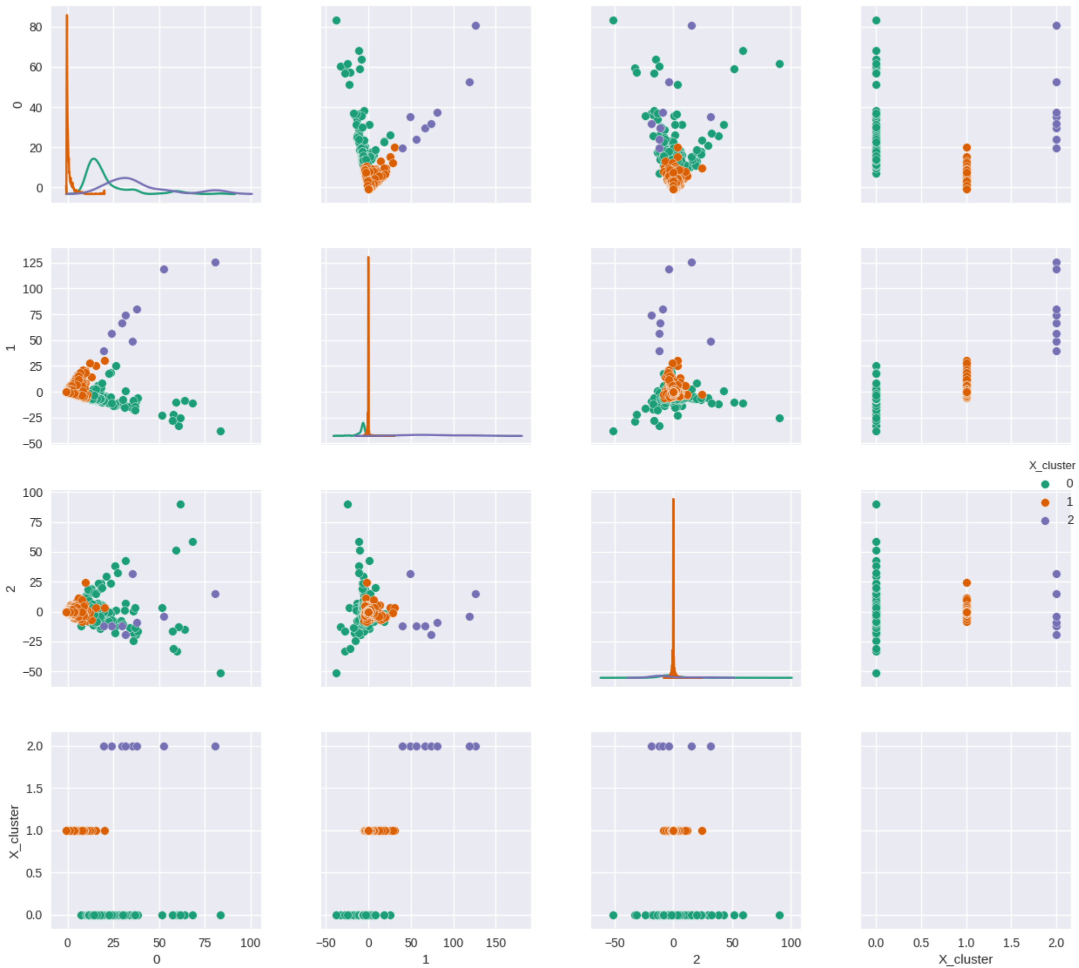

Figure 3. Cluster analysis using basic machine learning algorithms such as K-means clustering. (In production we use a more robust and accurate clustering algorithm which can be likened to K-means). Clustering serves as one of many building blocks in dynamic construction of baskets.

Detailed Algorithm and Technology Overview

Discovering Relationships

What two words come to mind when you hear the word “sky?”

For the vast majority of people, these words are “blue” and “cloud.”

These words are not synonyms. They are not semantically related. There is no direct way to relate “blue” and “cloud” to “sky.” The relationships can only be found by examining context and concepts.

Our system uses a biomimetic approach to discover:

- How cryptocurrencies are related
- How strongly they are related
- In what contexts they are related

A Biomimetic Approach

The human brain is the best pattern matcher on Earth. The best way to find relationships is to use approaches from the cognitive sciences and emulate how humans organize, search, and recall conceptually related information.

Our system auto-associates attributes related to cryptocurrencies like a human research analyst might and scores their relationships based on proximity, frequency, and uniqueness.

The Process of Auto-Association

System overview

A large body of publicly available data is required as input. Meaningful relationships between cryptocurrencies are found when statistically scored associations are generated over thousands or up to millions of attributes and objects. Smaller collections do not contain enough associations to produce meaningful results.

The system partitions the input into frames of context (Pinker, S., 1997). An auto-association strategy (Xijin Ge, Shuichi Iwata, 2002) is applied to the objects that comprise each frame. Using the resulting data, the system scores how strongly related each two unique cryptocurrencies are. These relationship scores are stored in a collection of vectors, which the system saves as an associative memory module (AMM).

AMM data is represented by a vector space model (Raghavan, V. V., and Wong, S. K. M., 1986). Object similarity is estimated using relationship consensus scores, which are based on correlation coefficients of relevant AMM vector numbers.

System implementation

Context frames

Context frames are created by selecting a sequence of natural language attributes from the data. One frame is anchored on each attribute. The number of attributes included in the frame is determined by how many lie within the distance threshold from the anchor. Each object must anchor exactly one frame, and may appear as a non-anchor in many frames.

Relationship strength

Relationship scores between attributes and cryptocurrency pairs are based on a distance decay function utilizing the Fibonacci sequence (RA Dunlap, 1997). Decay functions based on exponential ($e^{-\Delta x}$) and inverse distance (Δx^{-1}) sequences were also considered (Fig. 4).

Given strength, defined as:

$$\mathbf{S_{ij} = \phi^{\Delta x}}$$

Where:

S_{ij} = relationship score between cryptocurrency attributes *i* and *j*,

ϕ = 0.618034 is the Golden Ratio component "phi"[†], and

$\Delta x = |x_i - x_j|$ is the relative position between cryptocurrency attributes *i* and *j*.

[†] ϕ is the decimal component of the Golden Ratio Φ = 1.618034.

The following decay curves result:

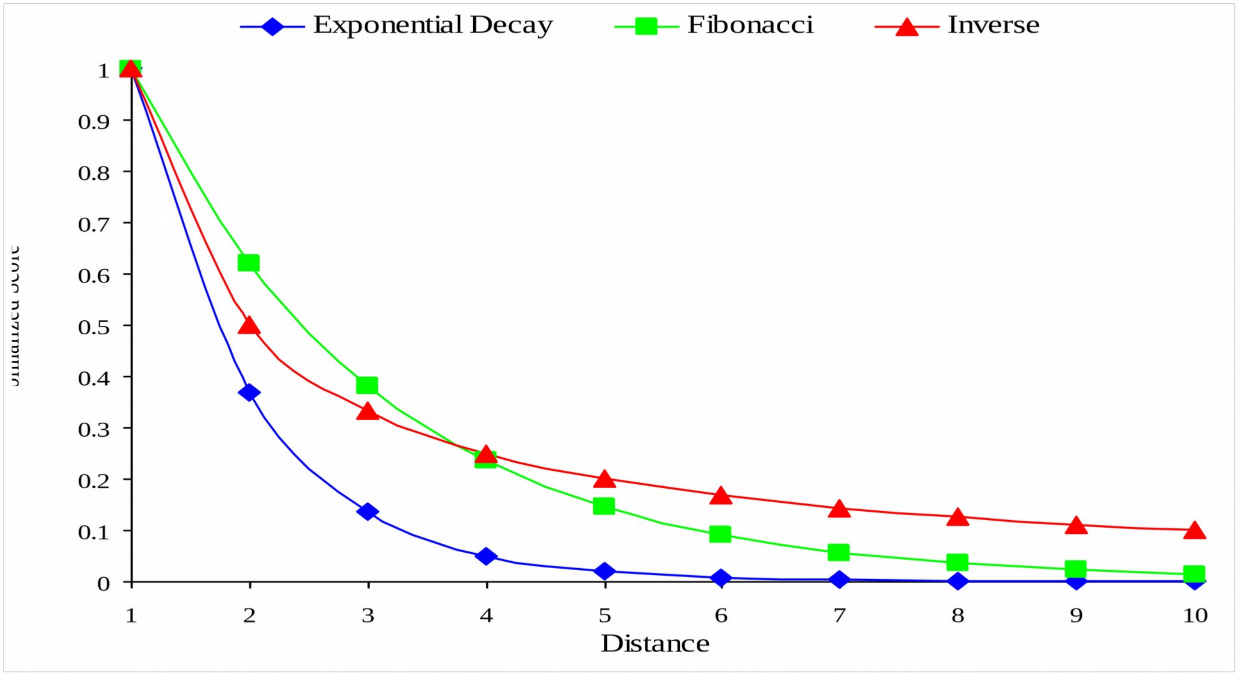

Figure 4. Distance Decay Curves. A comparison of three distance decay functions.

Exponential and inverse decay curves flatten out too quickly, resulting in poor differentiation among objects in the 6-10 distance range. The inverse decay curve does not approach zero, which also makes differentiation more difficult.

The decay curve using the Fibonacci sequence yields the most desirable result: it approaches zero at a rate that enables good separation of scores within the distance range of interest.

Cryptocurrency Attributes

For each unique cryptocurrency attribute, a vector is used to contain its collection of attributes, where each attribute is a related object. Attributes have scores associated with them, which are the result of combining distance with other metrics. Vectors may contain tens of thousands attributes per cryptocurrency.

Significant relationships between cryptocurrencies and other attributes e.g. global search trends, are defined by both strong and unique connections between shared attributes.

Associative memory modules

These object vectors are used to create AMMs. Relationships are inferred by considering the intersections of scored associative attributes of two or more cryptocurrency attribute vectors within an AMM.

Subsets of AMM data can also be used as filters.

Hierarchy-independent Analysis

The basic elements of analysis are cyrptocurrency attributes, which can be discrete elements of data, objects or pointers to collections of elements. This enables the system to operate in a hierarchy-independent manner.

Data collection identifiers are embedded in the framed context of each attribute, so queries can be traced to related groups of attributes, objects or sets of data sources. To generalize we can refer to all cryptocurrency attributes as objects to further illustrate.

Process flow

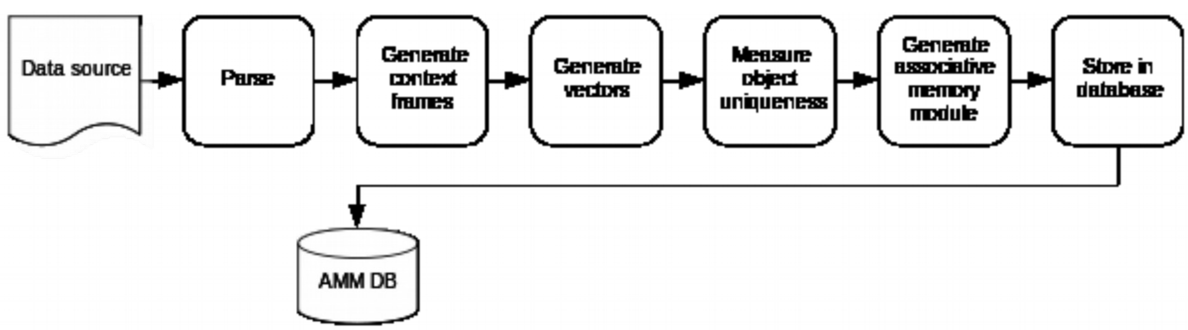

Figure 5. AMM generation.

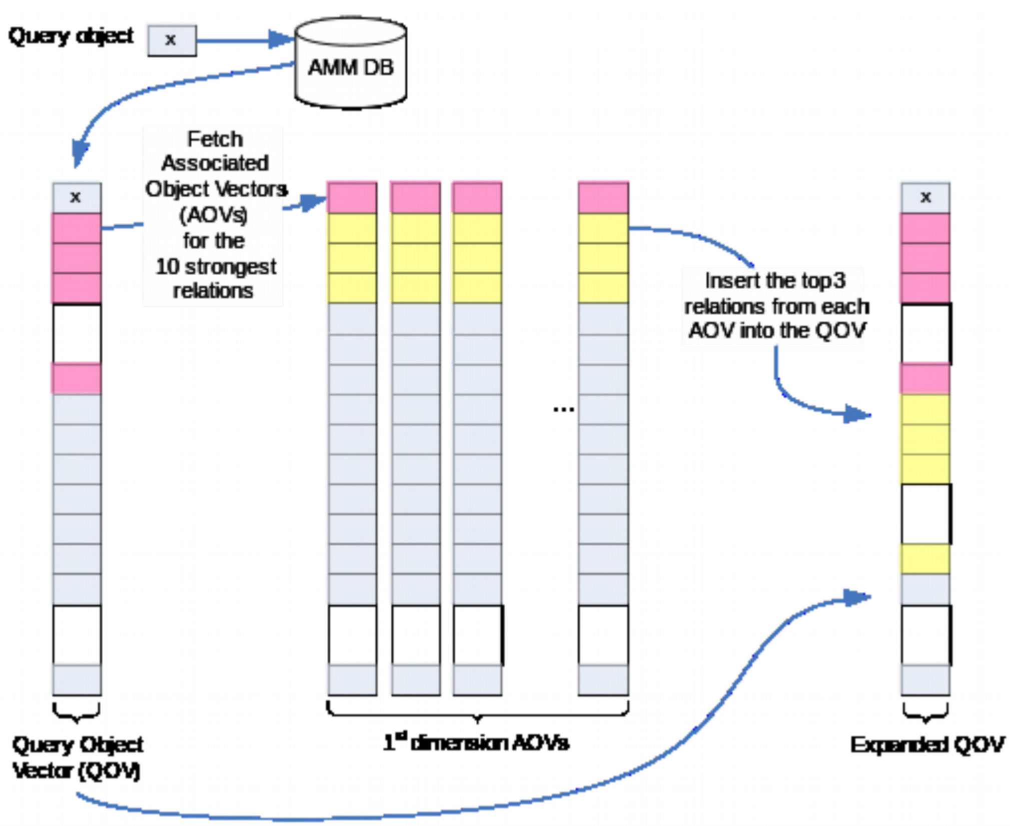

Figure 6. Query Object Vector (QOV) expansion.

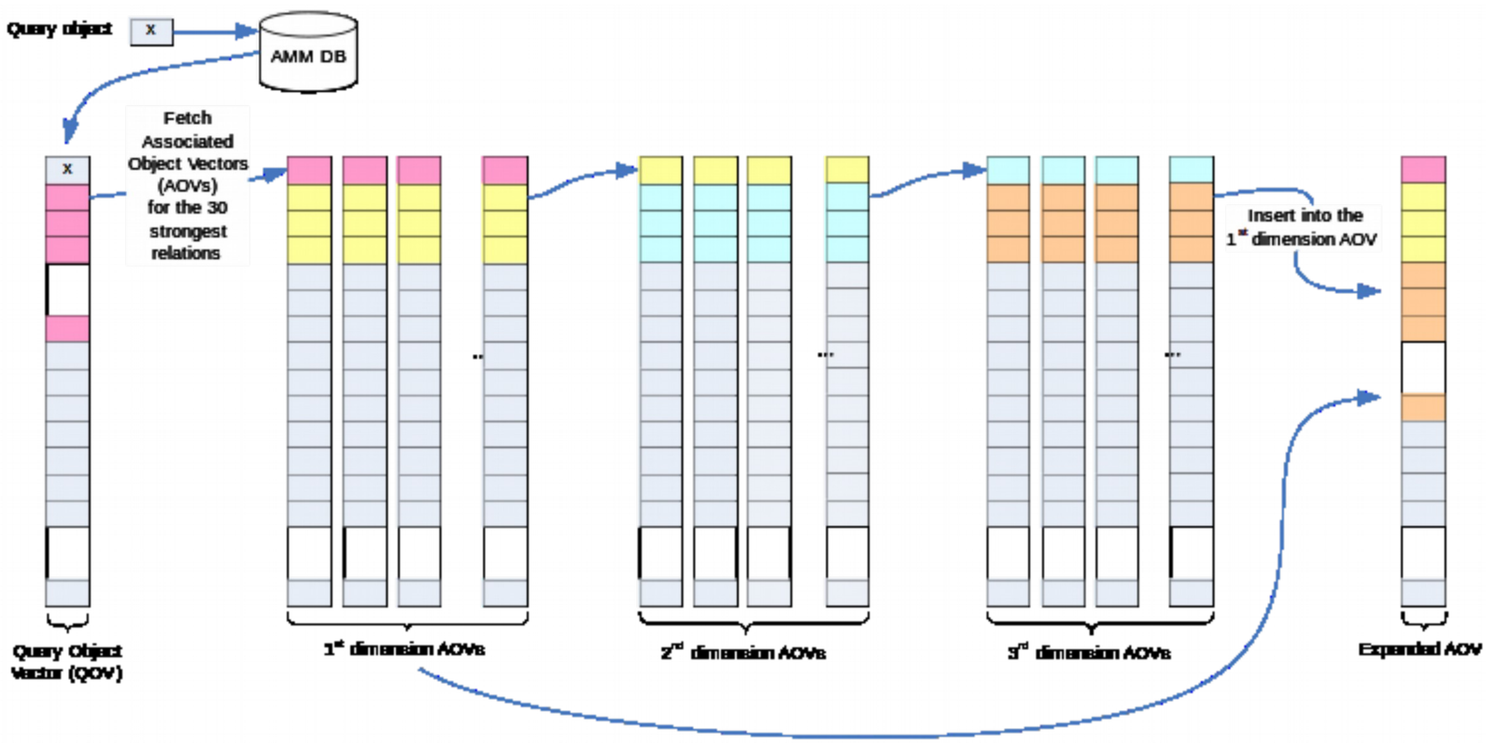

Figure 7. Associated Object Vector (AOV) expansion.

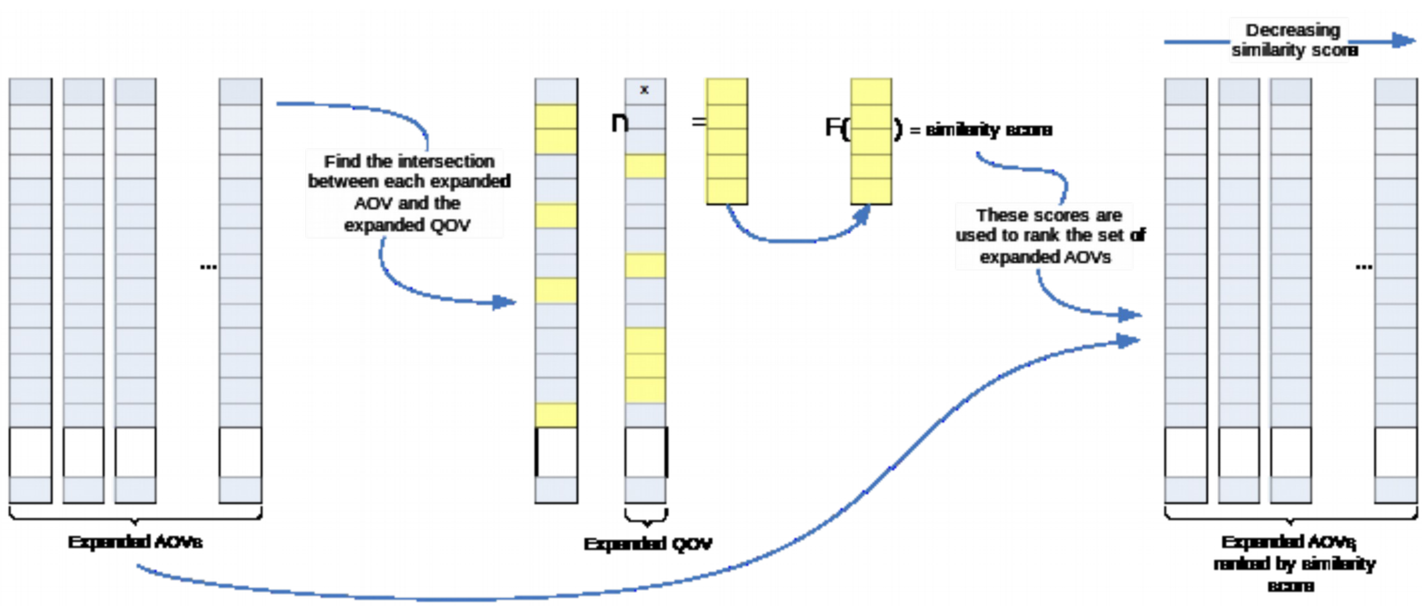

Figure 8. Ranking expanded AOVs by similarity to the expanded QOV.

Figure 5 shows the first phase of querying the AMM database for a given query object. Figure 6 shows how objects related at the 3rd dimension are pulled into the 1st dimension set of AOVs. Figure 8 shows how each expanded AOV is measured for similarity to the expanded QOV and how that score is used to rank the set of expanded AOVs.

The set of ranked, expanded AOVs represents the most common contexts for the query object. The top-ranked AOV can be used to infer the likely context of a particular query.

Conclusion

The statistical scoring of cryptocurrency attributes can be compared to how humans discover relationshpis via reinforced learning (Wenhuan, X., Nandi, A. K., Zhang, J., Evans, K. G., 2005) and the process of auto-association. Important relationships are strengthened when reinforced.

Our system uses these principles to create AMMs, which can be used to provide the following data:

- How cryptocurrencies are related.
- In what contexts they are related.
- How strong those relationships are.

In silico emulation of biomimetic object-association strategies has proven to be very effective in relationship discovery and data science, leading to new findings in numerous industries including life sciences, finance, search, social and contextual advertising.

Token Utility

The Vectorspace (VEC) tokens can be utilized in five primary ways initially which include:

- ✓ Baskets are intelligent in that they have the ability to determine for themselves whether or not to include additional cryptocurrencies or components from related baskets that may increase overall returns. Baskets that interact this way with one another will conduct these kinds of transactions between one another using the Vectorspace utility token, VEC which is also required to dynamically generate baskets.

- ✓ Tokens enable incentivized processes such as manual human curation, algorithm development and unique basket constructions.

- ✓ To secure ownership and auction-based relevant advertising rights within certain baskets which can be resold.

- ✓ The VEC utility token can be used to purchase blocks of dynamic basket generations in addition to subscriptions to higher level features offered by the platform including transacting baskets as trades on exchanges.

- ✓ VEC utility tokens are awarded to creators of top performing baskets which are displayed on a leader board each month.

In short, cryptocurrency research analysts and curators get awarded Vectorspace (VEC) tokens for mining relationships between worldwide events, global trends, local trends, news, URLs, keywords, hashtags and cryptocurrencies. In turn, Vectorspace (VEC) can then be used to purchase subscriptions to the platform, purchase API calls, queries or transact token baskets in the future. Other utility oriented actions we intend to implement include ongoing competitons, rewards and bounties for assembling optimized baskets which provide returns that outperform all other baskets.

Use of Funds

Funding will be applied to reaching the following milestones:

- 5 additional engineering hires and 2 devops hires
- Expanded office space
- Acquiring additional spot instances at Amazon AWS and Google cloud services

Milestones

- ➔ **Q1 2018**: Functionality enabling baskets to include additional cryptocurrencies or components from related baskets that may increase overall returns while also conducting these transactions using VEC utility tokens.
- ➔ **Q1 2018**: Optimization & customization of basket returns by applying filters such as technical analysis indicators and custom algorithms
- ➔ **Q2 2018**: Subscriptions to higher level features offered by the platform including transacting baskets as trades on exchanges.
- ➔ **Q3 2018**: Implementation of an auction-based relevant advertising system driven by transactions using VEC utility tokens.

Rewards & Bounties:

- 10,000 VEC go to the best performing token basket each month with 7,500 VEC going to 2nd place and 5,000 VEC going to 3rd place.
- We are offering 100,000 VEC to any VEC participant that suggests a product feature that we decide to implement and roll out to production.
- We are offering 1,000 VEC to any VEC participants that report a valid bug within our system.

Bonuses:

- 25% first 10 days of the crowdsale declining by 5% every 10 days after that

Pricing:

- 10,000 VEC tokens = 1 ETH
- 50 million total VEC tokens
- 50% to be shared

Citations

Huberman, G., Regev, T. - Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.199.4865&rep=rep1&type=pdf

Supun Kamburugamuve, Saliya Ekanayake, Pulasthi Wickramasinghe, Milinda Pathirage and Geoffrey Fox School Of Informatics and Computing, Indiana University (2015) Dimension Reduction and Visualization of the Structure of Financial Systems http://dsc.soic.indiana.edu/publications/stocks_visualization.pdf

Huang, E., Paraphrase Detection Using Recursive Autoencoder (2011) https://nlp.stanford.edu/courses/cs224n/2011/reports/ehhuang.pdf

Dunlap, Richard A. 1997. *The Golden Ratio and Fibonacci Numbers*. River Edge, NJ: World Scientific.

Pinker, Steven. How the mind works. New York: W. W. Norton & Company, 1997.

Raghavan, V. V. and Wong, S. K. M. *A critical analysis of vector space model for information retrieval.* Journal of the American Society for Information Science, Vol.37 (5), p. 279-87, 1986.

Wenhuan, X., Nandi, A. K., Zhang, J., Evans, K. G. (2005) *Novel vector quantiser design using reinforced learning as a pre-process*, Source Signal Processing, Volume 85, Issue 7.

Xijin Ge, Shuichi Iwata, *Learning the parts of objects by auto-association*, Neural Networks, v.15 n.2, p.285-295, March 2002.