ARM that text, make it useful!

Christopher Painter[1]
[1] Meme-machines.com

Meme Machines, Gloucester, U.K.
Chris.Painter@zigzag.co.uk

Abstract

This paper recommends the extension of Association Rule Mining (ARM) to unstructured text. This unsupervised approach is logical, has minimal parameters, and is well underpinned theoretically.

The combinational challenge can be met with the parallelism of GPUs. Sherlock is an example, a parallel implementation of the A Priori algorithm coupled with an optimal rule selector; it was initially deployed to understand changes in numbers, and is now being extended to incorporate simultaneous textual commentary.

A neural network approach to learning, Word2Vec [1], which also adopts the distribution hypothesis, merits comparison. ARM output is here directly compared to the output of a Word2Vec model trained on exactly the same data, and using the same frequency threshold. This comparison uses standard Python libraries, and source is provided [8]. Word2Vec does not prevail, but we understand it better.

ARM has virtues that are not available to neural net models in general. It is grounded, so rules can be explained back to the evidence. Furthermore these rules are easy to understand and to combine, allowing filtering, even by time. ARM's fundamental data unit is the 'yes or no' answer to a question, *any* question, providing easy extension from one symbol stream to another, as here recommended with numbers and text.

Motivation

Sherlock was designed to accumulate knowledge from a variety of sources, both textual and numerical, with minimal human intervention. The knowledge so obtained had to be grounded in demonstrable fact, chaining all the way back to the data if necessary. These constraints were entailed by the domain, foreign exchange trading, in which failure is easy to quantify, but difficult to learn from.

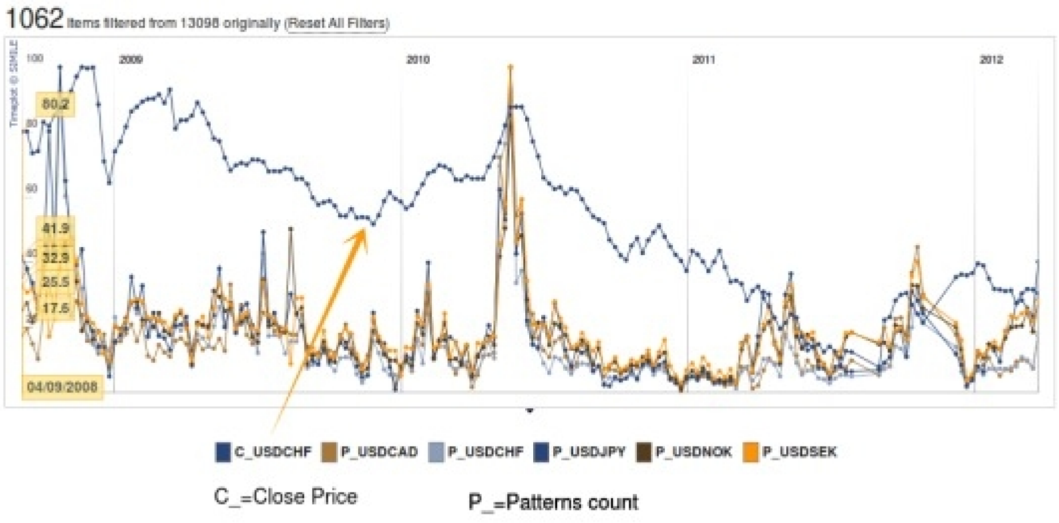
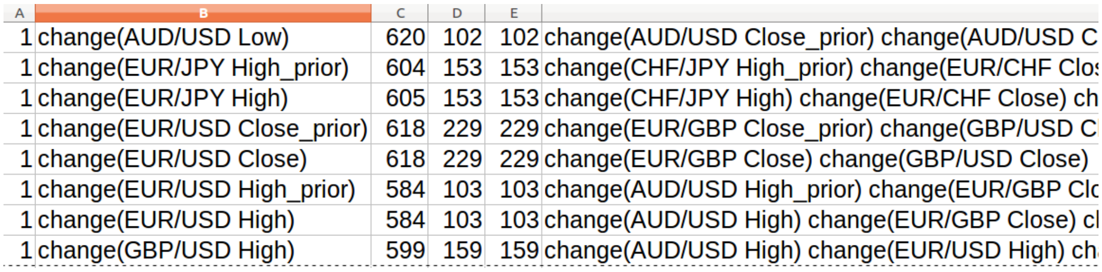
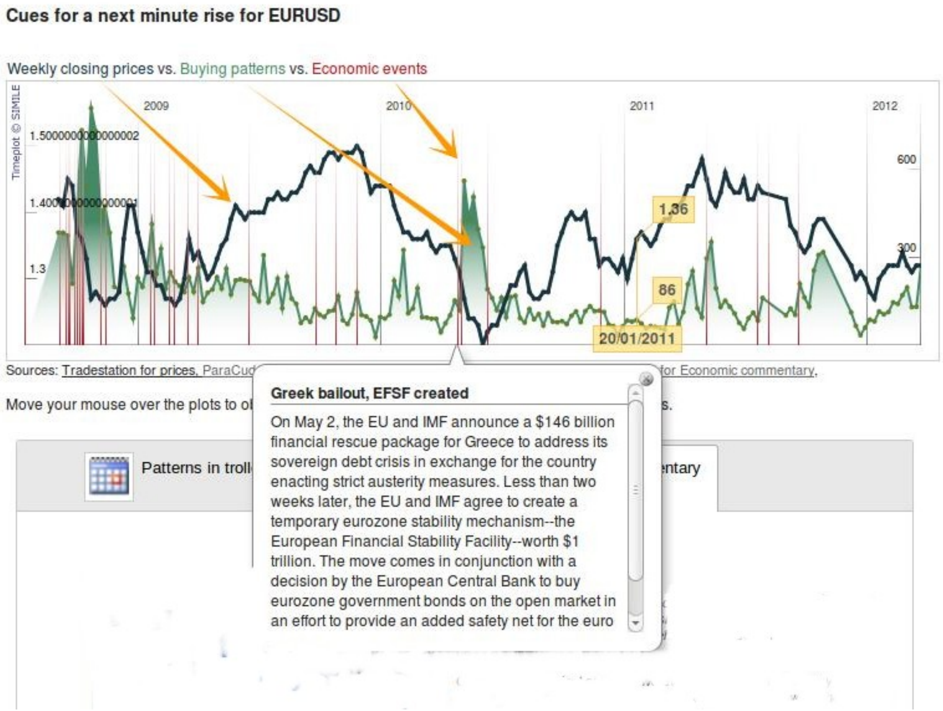


Figure 1: Prices and patterns: - USDCHF and Dollar patterns



A	B	C	D	E	
1	change(AUD/USD Low)	620	102	102	change(AUD/USD Close_prior) change(AUD/USD C
1	change(EUR/JPY High_prior)	604	153	153	change(CHF/JPY High_prior) change(EUR/CHF Clos
1	change(EUR/JPY High)	605	153	153	change(CHF/JPY High) change(EUR/CHF Close) ch
1	change(EUR/USD Close_prior)	618	229	229	change(EUR/GBP Close_prior) change(GBP/USD C
1	change(EUR/USD Close)	618	229	229	change(EUR/GBP Close) change(GBP/USD Close)
1	change(EUR/USD High_prior)	584	103	103	change(AUD/USD High_prior) change(EUR/GBP Clo
1	change(EUR/USD High)	584	103	103	change(AUD/USD High) change(EUR/GBP Close) cl
1	change(GBP/USD High)	599	159	159	change(AUD/USD High) change(EUR/USD High) ch

Figure 2: Patterns and rules: - Spotting the obvious. line 5 ~ "The Euro will be up against the Dollar whenever the Euro is up against Pound, and the Pound is up against the Dollar – other rules apply"



Cues for a next minute rise for EURUSD

Weekly closing prices vs. Buying patterns vs. Economic events

Sources: Tradestation for prices, ParaCud

for Economic commentary,

Move your mouse over the plots to ol

s.

Greek bailout, EFSF created

On May 2, the EU and IMF announce a $146 billion financial rescue package for Greece to address its sovereign debt crisis in exchange for the country enacting strict austerity measures. Less than two weeks later, the EU and IMF agree to create a temporary eurozone stability mechanism--the European Financial Stability Facility--worth $1 trillion. The move comes in conjunction with a decision by the European Central Bank to buy eurozone government bonds on the open market in an effort to provide an added safety net for the euro

Figure 3: Motivation: -The synchronized march of prices, patterns, and text?

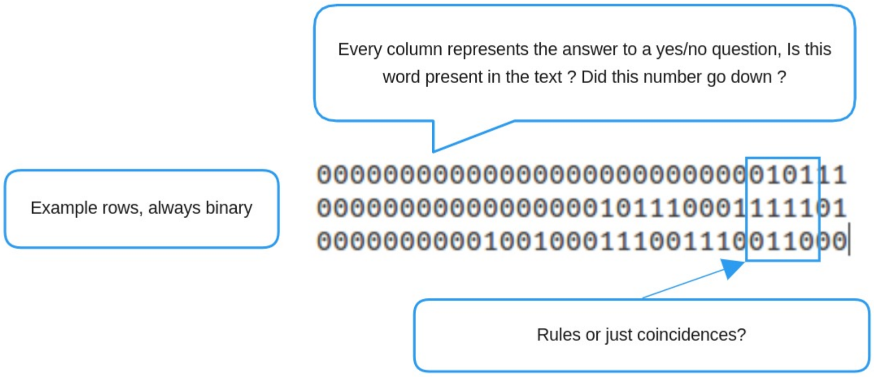
Each red spike in the figure above represents a commentary from the U.S. Council on Foreign Relations . This overlay against the price to pattern count *seems* to relate events to significant changes in price direction, prompting the following observations.

- More words than numbers that trade
- More number changers than word generators

Motivation? Perhaps it is possible to understand market prices as reactions to text and patterns, perhaps even predictable reactions.

Method

Rule discovery obliges unsupervised learning, and only one approach to such learning also avoids numerical assumptions and the parameter tuning they bring. That is to extract rules from sets, and to use logic and counting to identify the sets. All can be built on the property that every non-empty subset of a frequent set is also frequent.

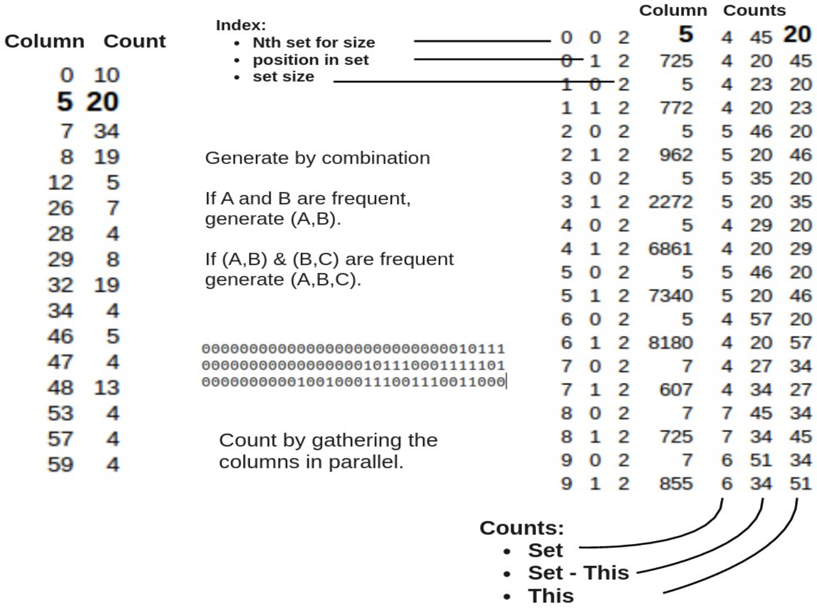


Every column represents the answer to a yes/no question, Is this word present in the text ? Did this number go down ?
Example rows, always binary
Rules or just coincidences?

```
00000000000000000000000000010111
00000000000000000101110001111101
00000000001001000111001110011000
```

Figure 4: Definitions

Defining our world in binary we start from an occurrence count for short sets, and use that 'A Priori' principle to generate longer valid candidates for counting.



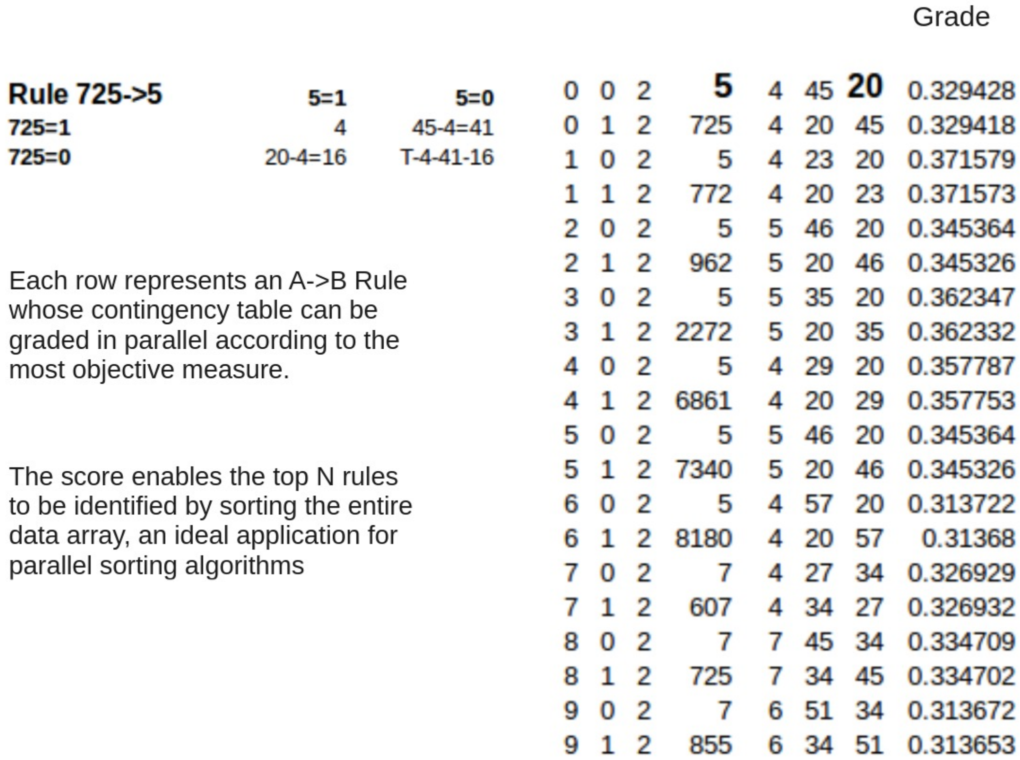
Index:
Nth set for size
position in set
set size
Generate by combination
If A and B are frequent, generate (A,B).
If (A,B) & (B,C) are frequent generate (A,B,C).
Count by gathering the columns in parallel.
Counts:
Set
Set - This
This

Column	Count
0	10
5	**20**
7	34
8	19
12	5
26	7
28	4
29	8
32	19
34	4
46	5
47	4
48	13
53	4
57	4
59	4

```
00000000000000000000000000010111
00000000000000000101110001111101
00000000001001000111001110011000
```

			Column	Counts		
0	0	2	**5**	4	45	**20**
0	1	2	725	4	20	45
1	0	2	5	4	23	20
1	1	2	772	4	20	23
2	0	2	5	5	46	20
2	1	2	962	5	20	46
3	0	2	5	5	35	20
3	1	2	2272	5	20	35
4	0	2	5	4	29	20
4	1	2	6861	4	20	29
5	0	2	5	5	46	20
5	1	2	7340	5	20	46
6	0	2	5	4	57	20
6	1	2	8180	4	20	57
7	0	2	7	4	27	34
7	1	2	607	4	34	27
8	0	2	7	7	45	34
8	1	2	725	7	34	45
9	0	2	7	6	51	34
9	1	2	855	6	34	51

Figure 5: The A Priori process



Rule 725->5	5=1	5=0
725=1	4	45-4=41
725=0	20-4=16	T-4-41-16

Each row represents an A->B Rule whose contingency table can be graded in parallel according to the most objective measure.

The score enables the top N rules to be identified by sorting the entire data array, an ideal application for parallel sorting algorithms

							Grade
0	0	2	**5**	4	45	**20**	0.329428
0	1	2	725	4	20	45	0.329418
1	0	2	5	4	23	20	0.371579
1	1	2	772	4	20	23	0.371573
2	0	2	5	5	46	20	0.345364
2	1	2	962	5	20	46	0.345326
3	0	2	5	5	35	20	0.362347
3	1	2	2272	5	20	35	0.362332
4	0	2	5	4	29	20	0.357787
4	1	2	6861	4	20	29	0.357753
5	0	2	5	5	46	20	0.345364
5	1	2	7340	5	20	46	0.345326
6	0	2	5	4	57	20	0.313722
6	1	2	8180	4	20	57	0.31368
7	0	2	7	4	27	34	0.326929
7	1	2	607	4	34	27	0.326932
8	0	2	7	7	45	34	0.334709
8	1	2	725	7	34	45	0.334702
9	0	2	7	6	51	34	0.313672
9	1	2	855	6	34	51	0.313653

Figure 6: Grading the rules

All that remains is to use what is termed in the literature a “measure of interestingness” for each line, which represents the rule that would be formed by taking this item from this set. Note that indexes start at 0, rather than 1, so the last line reads.

- Set is the tenth of length 2
- Conclusion of rule is second member of set, premise is rest of set
- rule is 7 → 855, which occurs 6 times, Premises {7} occurs 34 times, Conclusion {855} occurs 51 times and so has a comparative score of 0.313653
- The score will depend on the measure adopted see references [4,5,6].

Experiments

In the following three sections are laid out the results to probing three questions about the approach.

1. How does is it stack up against statistical methods ?
2. How could you test the results ?
3. How could you accumulate the results ?

For the first two items the textual data is stored in elasticsearch, and available to the public, while third takes RSS feeds, and stores the time-stamped rule observations into a Redis database.

ARM vs Word2Vec

What happens if the same data is passed through each algorithm for training, and the task is to compare the response of each to a stimulus vector? The output below is from a Python notebook that is available for download [8].

The data comes from a copy of the Simple English Wikipedia [9] is available for view, using the coordinates below.

```
curl -XPOST "http://public:pmb7wdkcmidejm41ai@19f7aca6869af2eb2e42e72a371ff653.us-east-1.aws.found.io:9200/wikismall/document/_search?" -d'
{"query": {"match": {
  "words": "logic"
}}}'
```

The response should appear as below, each matching document is there in text, and with a **words** field, representing the sorted set of the documents vocabulary, after subtraction of a standard stop word list.

```
"hits": {
      "total": 514,
      "max_score": 2.033338,
      "hits": [
         {
            "_index": "wikismall",
            "_type": "document",
            "_id": "AVk2_O0j4j8wYtSBp7u3",
                "_score": 2.033338,
                "_source": {
                   "author": "wikipedia",
                   "scope": "public",
                   "words": [
                      "Century",
                      "Developed",
                      "Easily",
                      "Field",
                      "Formalize",
                      "Logic",
                      "Mathematical",
                      "Mathematics",
                      "Reasoning",
                      "Symbols"
                   ],
                   "source": "Mathematical_logic.html",
                      "text": "Mathematical logic \r\n        Mathematical logic  is a field of
mathematics  , that tries to formalize  logic   so that it can be used for mathematics more
easily. Logic is about reasoning, mathematical logic tries to use symbols. Most of mathematical
logic was,developed in the 19th and 20th century. \r\n\r\n\r\n\r\n \r\n This article is issued
from  Wikipedia  - version of the Thursday, March 03, 2016. The text is available under the
Creative Commons Attribution/Share Alike  but additional terms may apply for the media files.",
                   "sherlock": "",
                   "timestamp": 1482686262189
                }
             },....
```

Taking the top ten replies, from the words field we build the input symbol set of sets, thus.

```
   sentences =
[['Century', 'Developed', 'Easily', 'Field', 'Formalize', 'Logic', 'Mathematical',
'Mathematics', 'Reasoning', 'Symbols'],
['Dilemma', 'Disambiguation', 'Good', 'Group', 'Logic', 'Mathematics', 'Meanings', 'Problems',
'Solutions', 'Song'],
['Called', 'Diagram', 'Disjunction', 'Displaystyle', 'Exclusive', 'False', 'Inclusive',
'Inputs', 'Logic', 'Lor', 'Operation', 'Scriptstyle', 'Takes', 'True', 'Venn'],
['ASCII', 'ATL', 'Cote', 'False', 'Flips', 'Input', 'Logic', 'Logical', 'Negation', 'Operation',
'Output', 'Returns', 'Takes', 'True'],
['Block', 'Called', 'Component', 'Configurable', 'Configured', 'FPGA', 'Gates', 'Hold',
'Interconnected', 'Logic', 'Number', 'Simplified'],
['Constructed', 'Engineered', 'Experiment', 'Experimental', 'Language', 'Languages',
'Linguistics', 'Logic', 'Logical', 'Philosophical', 'Philosophy', 'Types'],
['Attempts', 'Deducted', 'Deduction', 'Deductive', 'Expenditure', 'Give', 'Logic', 'Logical',
'Model', 'Natural', 'Naturally', 'Occurs', 'Profits', 'Reasoning', 'Taxation'],
['ASCII', 'ATL', 'Cote', 'False', 'Flips', 'Input', 'Logic', 'Logical', 'Negation', 'Operation',
'Output', 'Returns', 'Takes', 'True'],
['Commonly', 'Discourse', 'Element', 'Existence', 'Logic', 'Mirrored', 'Proposition',
'Quantifier', 'True', 'Universe', 'Written'],
['Aristotle', 'Grammar', 'Implication', 'Logic', 'Man', 'Men', 'Mortal', 'Suggest', 'Syllogism',
'True', 'Wednesday']]
```

A small amount of dictionary and reverse index work lines up the two approaches so that we can compare the output of an ARM rule [2] to the output of the word2vec model [3], as below

```
['Operation', 'False', 'True'] AR-> Logic <-W2V ['Logic', 'Logical']
['Operation', 'False', 'Logic'] AR-> True <-W2V ['Logical', 'True']
['Operation', 'True', 'Logic'] AR-> False <-W2V ['False', 'Takes']
['False', 'True', 'Logic'] AR-> Operation <-W2V ['Operation', 'Logical']
['Takes', 'False', 'True'] AR-> Logic <-W2V ['Logic', 'Operation']
['Takes', 'False', 'Logic'] AR-> True <-W2V ['Operation', 'True']
['Takes', 'True', 'Logic'] AR-> False <-W2V ['Operation', 'False']
['False', 'True', 'Logic'] AR-> Takes <-W2V ['Operation', 'Logical']
['Takes', 'Operation', 'True'] AR-> Logic <-W2V ['Logic', 'Logical']
['Takes', 'Operation', 'Logic'] AR-> True <-W2V ['True', 'False']
['Takes', 'True', 'Logic'] AR-> Operation <-W2V ['Operation', 'False']
['Operation', 'True', 'Logic'] AR-> Takes <-W2V ['False', 'Takes']
['Takes', 'Operation', 'False'] AR-> Logic <-W2V ['Logic', 'Logical']
['Takes', 'Operation', 'Logic'] AR-> False <-W2V ['True', 'False']
['Takes', 'False', 'Logic'] AR-> Operation <-W2V ['Operation', 'True']
['Operation', 'False', 'Logic'] AR-> Takes <-W2V ['Logical', 'True']
['Takes', 'Operation', 'False', 'True'] AR-> Logic <-W2V ['Logic', 'Logical']
['Takes', 'Operation', 'False', 'Logic'] AR-> True <-W2V ['True', 'Logical']
['Takes', 'Operation', 'True', 'Logic'] AR-> False <-W2V ['False', 'Logical']
['Takes', 'False', 'True', 'Logic'] AR-> Operation <-W2V ['Operation', 'Logical']
['Operation', 'False', 'True', 'Logic'] AR-> Takes <-W2V ['Logical', 'Takes']
['Takes', 'Operation', 'False'] AR-> True <-W2V ['Logic', 'Logical']
['Takes', 'Operation', 'True'] AR-> False <-W2V ['Logic', 'Logical']
['Takes', 'False', 'True'] AR-> Operation <-W2V ['Logic', 'Operation']
['Operation', 'False', 'True'] AR-> Takes <-W2V ['Logic', 'Logical']
```

The last line reads that the cues ['Operation', 'False', 'True'] would trigger an AR rule whose conclusion is 'Takes', whereas the Word2Vec Model would be triggered to respond 'Logic', with 'Logical' as a second choice.

Due to the logical definition, all the sets quoted by ARM actually exist, and there are no others; equally there are 91 different words in the data, so Word2Vec's close to optimal performance may suggest something akin to frequent item set mining as its underlying mechanism, and the explanation of its performance.

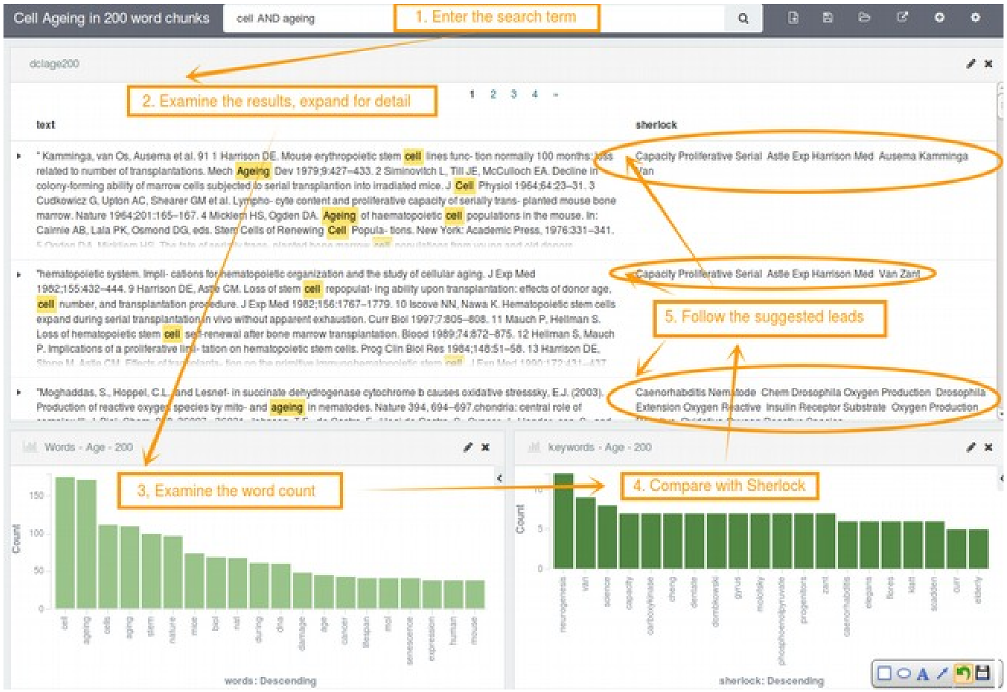
Discovering non-obvious connections

Handling new or unknown terms presents more challenge for the user than the algorithm, for which the coincidence is just some factual bitset lining up with congruity. This point was the key finding of explorations into small corpora, 200 to 500 pages each, of research papers on the related subjects of Cell Ageing, Cell Repair, and Exosomes. Each corpus has been split into consecutive chunks which have been indexed on an ElasticSearch cluster, along with a list of the uncommon words within each chunk,.

Sherlock has thereafter tagged chunks that contain themes of interest, and those results are now available for inspection via a Kibana dashboard at meme-machines.com . The dashboard compares the text and its word count in the left column to Sherlock's sorted suggestions, and their word count, in the right-hand column.

Figure 7 below shows such a dashboard for the contents of the Cell Ageing folder, with a numbered walk-through superimposed. At the top of the screen is a grey toolbar for loading and saving dashboards, just to the right of the search field where the walk-through starts.

The purpose of the walk-through is to compare Sherlock's output to user input, in order to see whether Sherlock "gets it", in this case the notion of Cell Ageing. The left hand column shows the texts and their counts, which are markedly different from Sherlock's selection. Figure 8 shows the results of Googling on Sherlock's suggestion.



1. Enter the search term

2. Examine the results, expand for detail

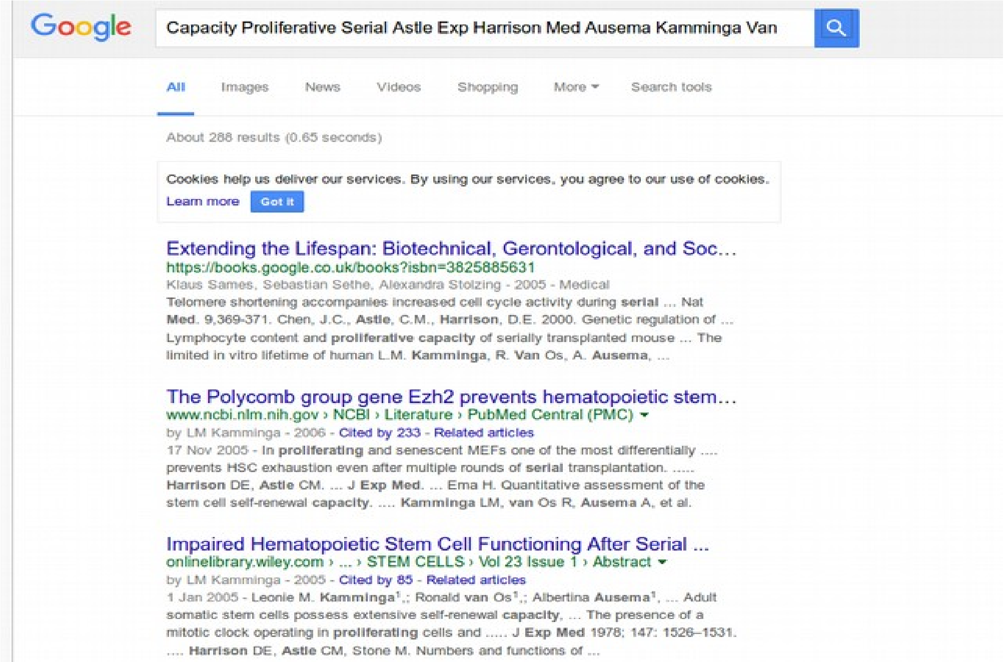
text	sherlock
" Kamminga, van Os, Ausema et al. 91 1 Harrison DE. Mouse erythropoietic stem cell lines func- tion normally 100 months: loss related to number of transplantations. Mech Ageing Dev 1979;9:427–433. 2 Siminovitch L, Till JE, McCulloch EA. Decline in colony-forming ability of marrow cells subjected to serial transplantion into irradiated mice. J Cell Physiol 1964;64:23–31. 3 Cudkowicz G, Upton AC, Shearer GM et al. Lympho- cyte content and proliferative capacity of serially trans- planted mouse bone marrow. Nature 1964;201:165–167. 4 Micklem HS, Ogden DA. Ageing of haematopoietic cell populations in the mouse. In: Cairnie AB, Lala PK, Osmond DG, eds. Stem Cells of Renewing Cell Popula- tions. New York: Academic Press, 1976:331–341.	Capacity Proliferative Serial Astle Exp Harrison Med Ausema Kamminga Van
"hematopoietic system. Impli- cations for hematopoietic organization and the study of cellular aging. J Exp Med 1982;155:432–444. 9 Harrison DE, Astle CM. Loss of stem cell repopulat- ing ability upon transplantation: effects of donor age, cell number, and transplantation procedure. J Exp Med 1982;156:1767–1779. 10 Iscove NN, Nawa K. Hematopoietic stem cells expand during serial transplantation in vivo without apparent exhaustion. Curr Biol 1997;7:805–808. 11 Mauch P, Hellman S. Loss of hematopoietic stem cell self-renewal after bone marrow transplantation. Blood 1989;74:872–875. 12 Hellman S, Mauch P. Implications of a proliferative limi- tation on hematopoietic stem cells. Prog Clin Biol Res 1984;148:51–58. 13 Harrison DE,	Capacity Proliferative Serial Astle Exp Harrison Med Van Zant
"Moghaddas, S., Hoppel, C.L., and Lesnef- in succinate dehydrogenase cytochrome b causes oxidative stresssky, E.J. (2003). Production of reactive oxygen species by mito- and ageing in nematodes. Nature 394, 694–697.chondria: central role of	Caenorhabditis Nematode Chem Drosophila Oxygen Production Drosophila Extension Oxygen Reactive Insulin Receptor Substrate Oxygen Production

3, Examine the word count

4. Compare with Sherlock

5. Follow the suggested leads

Figure 7: The walk through dashboard



Capacity Proliferative Serial Astle Exp Harrison Med Ausema Kamminga Van

About 288 results (0.65 seconds)

Cookies help us deliver our services. By using our services, you agree to our use of cookies.

Extending the Lifespan: Biotechnical, Gerontological, and Soc...
https://books.google.co.uk/books?isbn=3825885631
Klaus Sames, Sebastian Sethe, Alexandra Stolzing - 2005 - Medical
Telomere shortening accompanies increased cell cycle activity during **serial** ... Nat **Med**. 9,369-371. Chen, J.C., **Astle**, C.M., **Harrison**, D.E. 2000. Genetic regulation of ... Lymphocyte content and **proliferative capacity** of serially transplanted mouse ... The limited in vitro lifetime of human L.M. **Kamminga**, R. **Van** Os, A. **Ausema**, ...

The Polycomb group gene Ezh2 prevents hematopoietic stem...
www.ncbi.nlm.nih.gov › NCBI › Literature › PubMed Central (PMC)
by LM Kamminga - 2006 - Cited by 233 - Related articles
17 Nov 2005 - In **proliferating** and senescent MEFs one of the most differentially .... prevents HSC exhaustion even after multiple rounds of **serial** transplantation. ..... **Harrison** DE, **Astle** CM. ... J **Exp Med**. ... Ema H. Quantitative assessment of the stem cell self-renewal **capacity**. .... **Kamminga** LM, **van** Os R, **Ausema** A, et al.

Impaired Hematopoietic Stem Cell Functioning After Serial ...
onlinelibrary.wiley.com › ... › STEM CELLS › Vol 23 Issue 1 › Abstract
by LM Kamminga - 2005 - Cited by 85 - Related articles
1 Jan 2005 - Leonie M. **Kamminga**^{1},; Ronald **van** Os^{1},; Albertina **Ausema**^{1}, ... Adult somatic stem cells possess extensive self-renewal **capacity**, ... The presence of a mitotic clock operating in **proliferating** cells and ..... J **Exp Med** 1978; 147: 1526–1531. .... **Harrison** DE, **Astle** CM, Stone M. Numbers and functions of ...

Figure 8: Would you have thought those keywords would matter?

Sherlock speaks Google: and neither speak English! The suggestions are sorted sets of clues that matter to Sherlock, sometimes names get pulled into the association, as in Figure 7. It's not unreasonable, consider the Black Scholes equation, or Einstein's Theory of Relativity, but it doesn't make easy reading for humans.

Performance: In an ideal world every chunk would have a helpful suggestion, but several factors militate against this, namely the size and homogeneity of the chunks, and the setting of stopwords and the stopping condition.

Chunk size: The process progressively reduces the text, firstly into chunks, then into just the uncommon words, and finally into just the uncommon words that combine with other uncommon words in interesting ways. Smaller chunks mean fewer uncommon words, and therefore less chance of interesting combinations, as the following coverage table shows.

Subject	Chunk size*	Count	Tagged	Tagged %
Cell Ageing	100	3454	1030	30
Cell Ageing	150	2348	1375	59
Cell Ageing	200	1795	1512	**84**
Exosomes	150	3010	1261	41
Cell Repair	150	3020	1070	35

* The folder for each subject was indexed as a sequence of 150 word chunks. Furthermore 100 and 200 word chunk indices for the Cell Ageing folder were prepared in order to indicate sensitivity to chunk size. Memory limits prevented 200 word indices on Exosomes and Cell Repair.

Inconsistencies were noted in word totals, probably due to server conflicts as to whether a document existed already or not (<1%).

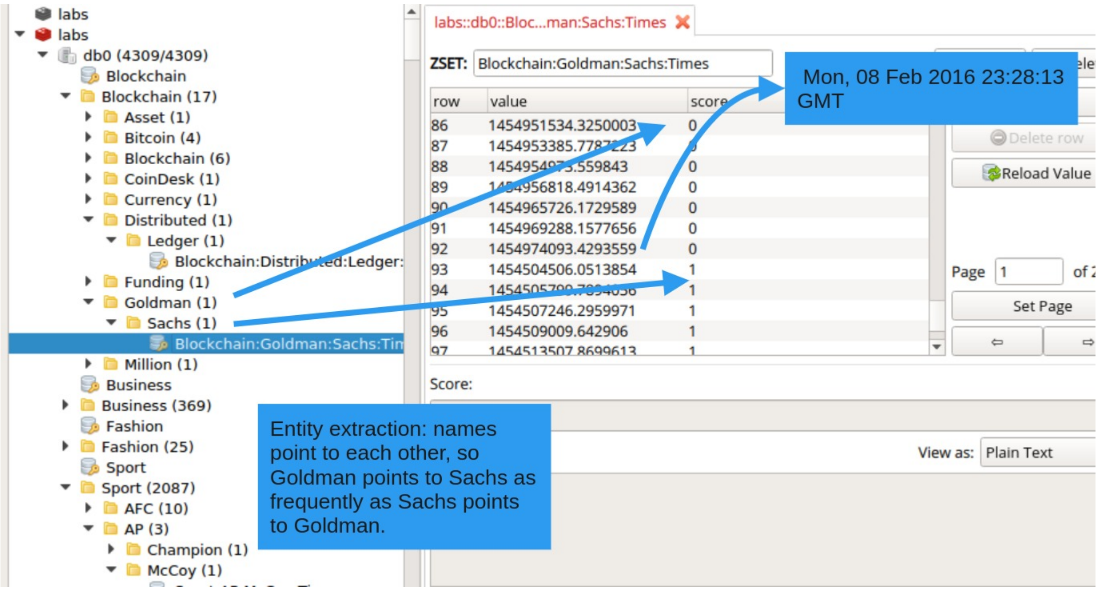
Chunk homogeneity: Mechanical extraction can be fooled by repetitive patterns that are not related to the underlying content, such as copyright notices and journal titles. Restricting the input texts to content would therefore improve results.

Stopwords : For this examination the standard Google stopwords for English, as reported by Wikimedia, were ignored; that list can be augmented to improve quality.

Stopping condition : Sherlock mines progressively deeper until the results contain at least a given number of unique words, 100 in these tests. Increasing that threshold would reveal more obscure correlations, and increase coverage, but might overwhelm memory resources by generating too many candidate solutions.

Accumulating the results

The ordered set structure of ARM rules make accumulation easy, in effect hanging branches off a tree. In figure 9 below are accumulated sequences gathered from news feeds on Blockchain, Business, Fashion, and Sport; from each is suspended an ordered tree, grounded by the timings and rule observations. The set {Goldman,Sachs} produces Goldman → Sachs as often as it does Sachs → Goldman, because it's Goldman **&** Sachs.



ZSET: Blockchain:Goldman:Sachs:Times

row	value	score
86	1454951534.3250003	0
87	1454953385.7787223	0
88	1454954973.559843	0
89	1454956818.4914362	0
90	1454965726.1729589	0
91	1454969288.1577656	0
92	1454974093.4293559	0
93	1454504506.0513854	1
94	[illegible]	1
95	1454507246.2959971	1
96	1454509009.642906	1
97	1454513507.8699613	1

Page 1 of 2

Score:

View as: Plain Text

Mon, 08 Feb 2016 23:28:13 GMT

Entity extraction: names point to each other, so Goldman points to Sachs as frequently as Sachs points to Goldman.

Figure 9: Redis ZSets making a timestamped knowledge base

Conclusions

The same GPU parallel acceleration that powered the resurrection of interest in neural net learning has also renewed interest in areas that were previously avoided because of combination challenges. Frequent Item Set and the Association Rule Mining fall into that zone. Applying such techniques to text looks even more challenging, due to the size of the vocabulary of an average corpus (the Simple English Wikipedia has more than 180,000 distinct words).

However, neither logic nor counting have to be done in sequence. Partial parallelism on 1000+ core hardware and within 8 GB memory puts the entire Simple English Wikipedia firmly on the list of possibles.

ARM and Word2Vec confirm the attraction of looking for bags of words; Sherlock has been tried on many languages, right to left readers like Arabic included. You'll need only a stoplist to get going, conveniently like a teacher [7] who can point to combinations *not* to search, *' forget all connectives' etc.*.

Perhaps it is disappointing that brute force wins, and that the encoded bias of the numerical approach is exposed by the data; on the other hand, humans often exhibit conflicting signals, and unfounded distribution assumptions have often produced catastrophic results. Nowhere was this more true than in the financial sector. ARM the data instead ~ it's more understandable..

“In the end, people are persuaded not by what you said, but by what they understand”.

John C. Maxwell

References

[1.] Mikolov. T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR (2013)

[2] Christian Borgelt's Web Pages, http://www.borgelt.net/fpm.html

[3] Radim Rehurek: topic modelling for humans http://radimrehurek.com/gensim/models/

[4] Michael Hahsler, A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules, http://michael.hahsler.net/research/association_rules/measures.html

[5] PangNing, T., Vipin, K., Jaideep, S.: Selecting the Right Objective Measure for Association Analysis, Information Systems. Vol. 29, no. 4, pp293313 (2004)

[6] Mosteller, F.: Association and Estimation in Contingency Tables, J. Am.Stat.Assoc. Vol. 63, pp 128 (1968)

[7] Mikolov, T., Joulin, A., & Baroni, M. (2015). A Roadmap towards Machine Intelligence. https://arxiv.org/abs/1511.08130

[8] Painter,C.,: ARM vs Word2Vec Python notebook https:meme-machines.com/ARM_Word2Vec.ipynb

[9] Wikipedia.org : Simple English Wikipedia https://simple.wikipedia.org/wiki/Main_Page