

AN AUGMENTED INTELLIGENCE MODEL TO EXTRACT PRAGMATIC MARKERS

Vijay Perincherry¹, David White² and Staci Warden³

¹Indiggo Associates, Bethesda, Maryland, USA

²Oteemo Inc, Washington, DC, USA

³Milken Institute, Washington, DC, USA

ABSTRACT

This paper presents a novel methodology for automatically extracting pragmatic markers from large streams of texts and repositories of documents. Pragmatic markers typically are implications, innuendos, suggestions, contradictions, sarcasms or references that are difficult to define objectively, but that are subjectively evident. Our methodology uses a two-stage augmented learning model applied to a specific use case, extracting from a repository of over 1500 Article IV country reports prepared for government officials by International Monetary Fund (IMF) staff. The model uses principles of evidence theory to train a machine to decipher the textual patterns of suggested actions for government officials and to extract those suggestions from the country reports at scale. We demonstrate the effectiveness of the model with impressive precision and recall metrics that over time outperform even the human trainers..

KEYWORDS

Natural Language Understanding, Augmented Intelligence, Pragmatics, Text Processing

1. OVERVIEW

In recent years digital text processing has been used extensively to classify documents, extract specific content, and conduct semantic analysis of various kinds at scale. Examples range from the classification of content by aggregation platforms, to processing medical and diagnostic reports [1][2], to the semantic analysis used by retailers and consumer service organizations to uncover positive or negative sentiment. These approaches, for the most part, represent rule-based extractions of text patterns or word sequences drawn from custom repositories for specific purposes [3][4]. More advanced models overcome the limitations of relying only on specific terms by generating variations of wild card sequences around target phrases [5][6]. All the use cases, however, represent situations where unambiguous conclusions within specific domains need to be extracted at scale. They do not address the issue of detecting latent intent.

The following section discusses some related work published by researchers in various areas. The two sections that follow detail the architectural and analytical underpinnings of the systems that we developed. In the last section, we demonstrate the effectiveness of the model using an application to extract suggestions for Government actions implied by IMF staff in Article IV country reports. The Article IV repository we used contains more than 1,500 documents prepared over the last 18 years. While the suggestions extracted from them are used for analysing trends and variations in IMF policy recommendations, this paper focuses on the system used to extract them.

2. RELATED WORK

While the ability to process streams and explicit expressions is adequate for most traditional applications, the recent proliferation of unsubstantiated news stories and unverified claims through social media channels has revealed the limitations of these models. In such cases, it is often not the content, but the way the content is presented that influences a reader's perception and interpretation. But the volume of content and variety of sources have made it impossible for consumers to verify authenticity of information and to identify hidden intents, exposing them to deception and manipulation on a massive scale.

It is now critical to develop methods to detect hidden intents easily and at scale, and in order to accomplish this, we need approaches that can identify the terms and lexical sequences that linguists refer to as pragmatic markers [7] hidden within documents. Pragmatic markers typically include implications, innuendos, suggestions, or sarcasms that are difficult to objectively define using objective search criteria. By definition, they represent linguistic attempts to influence readers' interpretations, and therefore, do not lend themselves to such prescriptive approaches. From a linguistic data-extraction standpoint, what makes this challenging is that pragmatic markers are usually characterized by specific phraseologies and terminologies that are often ignored as "stop words." But although they are hard to define, humans are fairly good at identifying such patterns.

A very important and timely analysis of the use of pragmatics markers in manipulative political discourse was conducted by Peter Furko[11]. This study pointed out that a majority of critical analysis tended to focus on the manipulative potential of the choice of content words and syntactic choices. The study provides a detailed coverage of transcripts of political interviews aired by CNN and BBC. It highlights the role of "of course" and extenders like "stuff like that" in conveying hidden messages that are not overtly stated. Britt Erman [12] studied the use of "you know" in common language to understand whether adolescents and adults use it differently, and also to understand whether its meaning and function are changing. The study concludes that the marker does signal build up of text and as a bridge for coherence, and may signal a sense of discontinuity in the speaker's mind.

One of the major challenges that limit effective use of pragmatic markers to extract hidden intents is the lack of a generalized collection of such markers. They are usually context driven and inherit the linguistics structure of the community and its common parlance. This would require the development of custom markers for each use case. The approach suggested by Marin et al [8] to detect forum claims in Wikipedia discussions is perhaps the closest in scope to the problem of pragmatic analysis. They addressed one significant limitation that plagued most other approaches, namely the identification of negative markers. However, they also resorted to knowledge-driven word lists drawn from relevant content as targets without seeking out intent. More recently, the use of deep learning and artificial intelligence based systems in language processing is becoming increasingly popular [13][14].

In this paper, we present an Augmented Intelligence model that automatically develops a graph-base representation of pragmatic markers based on statements identified by humans. It is a two-stage learning model that has the ability to continually improve and adapt to evolving structures and notions. Over time, this knowledge base can then be used to detect pragmatic markers in a content stream and/or a large repository.

3. TWO-STAGE LEARNING MODEL

In this application, we attempted to extract suggested actions from IMF Article IV documents. The model proceeded in two stages. First, the model attempted to detect and identify positive patterns that indicate that a given statement might contain a target marker of a suggested action. Second, the model filtered that subset of possible suggested-actions statements for negative pattern that would disqualify a given statement as a suggested action.

More generally, we refer to the statements of interest within a document as target statements, and the others as normal statements. In the generalized version of our model, it first filters all statements for positive markers indicating that they might be target statements. Then, in a second stage, it develops a set of negative filters to indicate the possible normal statements among them, in order to improve precision.

The negative filter is critical because the flexibility and richness of linguistic expressions means that simply seeking positive patterns will cause too many false-positive indicators. We proceed in two stages because in most practical cases, normal statements far outnumber target statements. Therefore, if we attempt to detect both positive and negative markers simultaneously, the number of possible terms and sequences of terms will be dimensionally prohibitive, even with documents of moderate size.

Intuitively, then, the model follows a cognitive approach of over-learning followed by error-correction. The two-stage learning model is schematically illustrated in Figure 1.

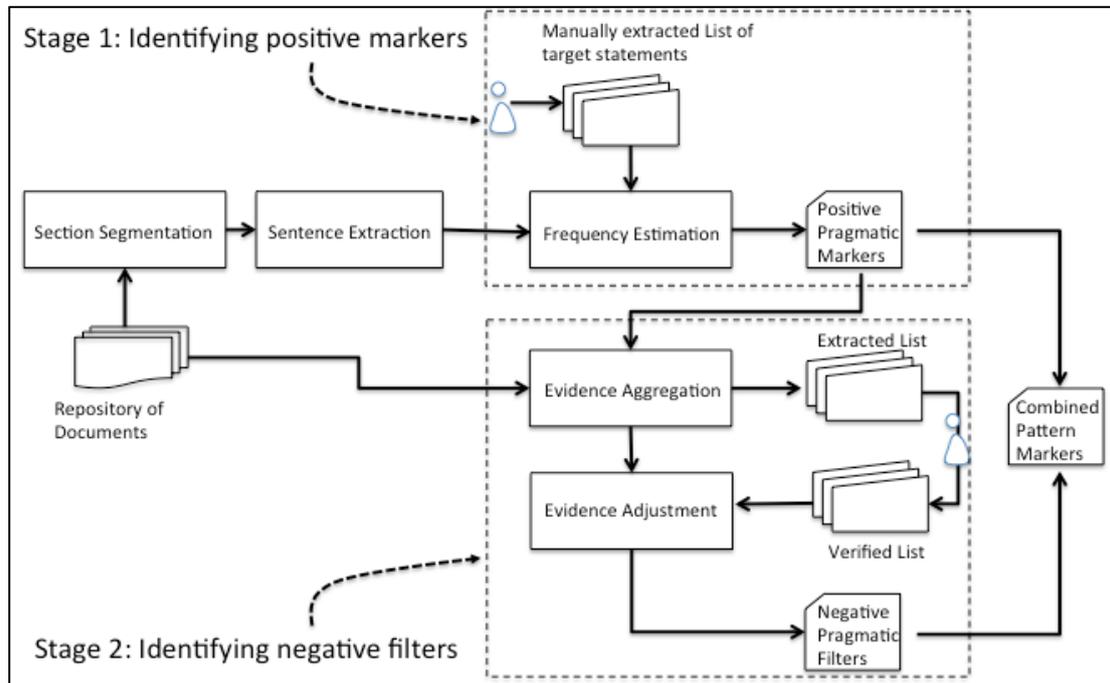


Figure 1. Schematic for the Two-Stage Learning Model

The data assembly for learning consists of two principal components: (i) a Section Segmentation Module that divides a given document into sections of contiguous blocks; and (ii) a Sentence Extraction Module that extracts sentences from text blocks using standard Regex patterns.

The two stages of learning identified in the figure are:

Stage 1: Identification of positive markers; and

Stage 2: Identification of negative filters.

The analytical framework for each stage is discussed below.

3.1. Stage 1: Identifying Positive Markers

The first step in extracting positive markers is to extract all possible combinations of words within each of the target statements. Then an N-gram frequency estimator is used to determine the relative frequencies of terms and combinations of terms in the document as a whole. These frequencies are then converted into certainty factors using the following formulation:

$$C(S_T|w) = \frac{N_{wT}}{N_{wD}} \alpha \text{len}(w) \beta(D) \quad (1)$$

where

$C(S_T|w)$ represents the level of certainty that a statement S drawn from a document D is a target statement if it contains phrase w . This is analogous to the notion of Bayesian conditional probability.

N_{wT} represents the number of occurrences of phrase w in the set of target statements

N_{wD} represents the number of occurrences of phrase w in the document D

$\text{len}(w)$ represents the length of the phrase w

α represents a constant parameter for the weight of the length (=0.4)

$\beta(D)$ indicates how representative the document D is of the repository (=0.02)

Although the probabilities for each phrase are computed independently, a phrase and its substrings are considered as one N-gram class for evidence aggregation during pattern matching.

3.2. Stage 2: Identifying Negative Filters

During Stage 2, the positive markers are applied to all statements in document D using an Evidence Aggregation approach to extract an estimated set of target statements. This method aggregates the probabilities of all phrase groups using the evidence aggregation model proposed by Klir and Wierman[9]. Thus, if two independent phrase groups are present in a statement S , each with certainty levels C_1 and C_2 respectively, the total certainty for the sentence C_S is calculated as:

$$C_S = C_1 + (1 - C_1) C_2 \quad (2)$$

The total certainty for each sentence is aggregated cumulatively, and the sentence is classified as a suggested action if the total score exceeds a specified threshold. For our purposes, we used a threshold of 0.5, essentially indicating that the likelihood of the sentence being a recommendation exceeds the likelihood that it is not.

A trainer verifies this list of extracted targets and identifies normal statements that have been erroneously extracted as targets. Then a set of n-grams identified for each statement by removing the positive markers from them is used as negative filters. The certainty level associated with each of the negative markers is computed as:

$$C(S_R|w) = \frac{N_{wR}}{N_{wD}} \quad (3)$$

$C(S_R|w)$ represents the level of certainty that a statement S drawn from a document D is not a target statement if it contains phrase w .

N_{wR} represents the number of occurrences of phrase w in the set of normal statements

N_{wD} represents the number of occurrences of phrase w in the document D

4. APPLICATION MODEL

The pattern-matching pipeline is shown in Figure 2. As shown in the figure, the pipeline includes (1) a text extraction module that extracts text from specific sections of the target document, (2) a pattern repository consisting of all positive markers and negative filters, and (3) a Binary Classifier that aggregates the evidences of target and normal statements, as indicated by the markers and filters.

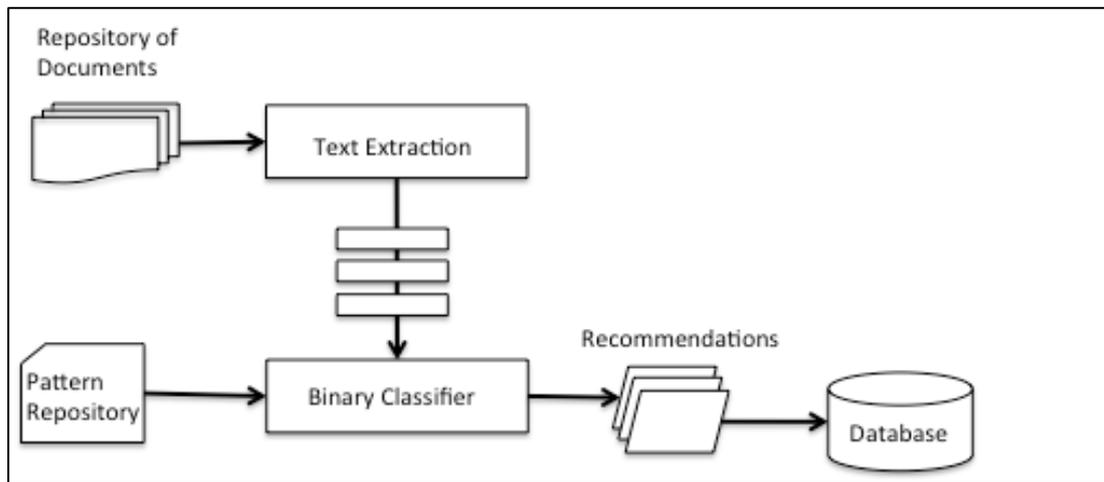


Figure 2. Pattern Matching Pipeline

The aggregated certainty level for each statement is computed as:

$$C(S_R) = (+)C_P(1 - (+)C_N) \quad (4)$$

Where

$(+)C_P$ represents aggregation of certainty of positive markers, and $(+)C_N$ represents that of all negative markers.

The aggregation model is as shown in Equation (2) above.

5. EXPERIMENTAL RESULTS

We used a repository of publicly available PDF versions of more than 1,500 Article IV country reports for the period spanning 1999 to 2017. First, we configured a text extraction module to extract text blocks and constituent sentences from the PDF documents. A standard regex definition was specified to extract sentences from the text.

5.1. Study Approach

Our approach consisted of the following steps:

- Analysts in the study team manually extracted statements suggesting Government actions from a handful of Article IV documents.
- The Model applied the two-stage pattern identification algorithm to generate a set of linguistic markers indicating these suggestions.
- The model applied the evidence-based pattern-matching algorithm to extract suggested statements from the entire repository of Article IV documents

We analysed this data array to generate all the insights presented in this paper.

5.2. Model Calibration and Validation

The Article IV documents are generally stored as PDF documents. We used Apache PDFBox to export text blocks to serve as source as source documents for the application of pattern development and application.

The pattern file for pragmatic marker evidences using the two-stage model was developed and validated using ten separate Article IV documents listed below:

2017 Uruguay	2003 Uruguay
2007 Myanmar	2016 Columbia
2012 Chile	2016 Indonesia
2006 Argentina	2016 Mexico
2007 Laos	2014 Angola

Analysts in the team first identified sentences contained in these documents that they interpreted as suggested actions. The analysts were asked to use their subjective inference to extract suggested actions from the documents, and this exercise demonstrated that even among humans, detecting latent intent is not straightforward. First, analysts became cognizant of the indirect and sometimes vague manner in which the IMF suggests actions for government officials. For example, analysts identified expressions such as “the staff underscored the importance of holding fiscal deficit steady” and “maintaining close monitoring will help address” as indicators of suggested actions. In this case, then, following more rigorous linguistic patterns and grammatical specifications for suggested actions would vastly understate the actions that IMF prefers that Government Authorities undertake. Furthermore, there was significant heterogeneity among analysts as to what constitutes a valid suggested action. In the end, we conducted a survey asking a range of people to opine on a subset of “borderline case” statements, and on that basis created guidelines for analysts to use.

The two-stage pattern extraction module identified 151 distinct markers with 119 positive markers and 32 negative markers. These were grouped into 76 classes.

We then selected another set of seven documents to compare how well the machine extracted target statements of suggested actions as compared to the analyst trainers.

The test group consisted of the following seven documents:

2003 Bahamas	2007 Samoa	2011 Japan	2005 Mexico
2007 Bolivia	2015 Somalia	2003 Lithuania	

We used Precision, Recall and F1 metrics to measure the efficiency of target detection. These measures are defined below:

- *Precision* measures the ratio of statements detected by the process that are in fact target statements
- *Recall* measures the ratio of all possible target statements that are in fact detected by the process
- *F1* measure is a hyperbolic mean of Precision and Recall and is measured as:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

The analysts and the machine independently processed the seven test documents to extract possible suggested actions. The results for each document are shown in Tables 1 (a) through (f). Table 1(g) shows a summary of all documents.

Table 1. Summary of Test Results

1(a): Bahamas 2003 (60 suggestions in the document)

Metric	Machine	Analyst
Number of suggestions identified	51	30
Precision	0.98	1.00
Recall	0.83	0.50
F1 metric	0.90	0.67

1(b): Bolivia 2007 (36 suggestions)

Metric	Machine	Analyst
Number of suggestions identified	27	21
Precision	0.85	1.00
Recall	0.64	0.58
F1 metric	0.73	0.74

1(c): Japan 2011 (32 suggestions)

Metric	Machine	Analyst
Number of suggestions identified	22	18
Precision	0.95	1.00
Recall	0.66	0.67
F1 metric	0.78	0.72

1(d): Lithuania 2003 (69 suggestions)

Metric	Machine	Analyst
Number of suggestions identified	62	34
Precision	0.74	1.00
Recall	0.71	0.51
F1 metric	0.72	0.67

1(e):

Samoa 2007 (39 suggestions)

Metric	Machine	Analyst
Number of suggestions identified	23	26
Precision	0.96	1.00
Recall	0.56	0.67
F1 metric	0.71	0.80

1(f): Somalia 2015 (64 suggestions)

Metric	Machine	Analyst
Number of suggestions identified	49	40
Precision	0.94	1.00
Recall	0.72	0.63
F1 metric	0.81	0.77

1(g): Mexico 2005 (329 total suggestions)

Metric	Machine	Analyst
Number of suggestions identified	24	13
Precision	0.74	1.00
Recall	0.61	0.46
F1 metric	0.67	0.63

1(h): Aggregated (329 total suggestions)

Metric	Machine	Analyst
Number of suggestions identified	257	182
Precision	0.88	1.00
Recall	0.70	0.55
F1 metric	0.78	0.71

6. DISCUSSION

The results tabulated above show that the machine outperformed the analysts in terms of detecting suggested actions in most cases. Naturally, the precision displayed by humans is absolute, but the machine significantly outperformed analysts in recall and the overall F1 metric. These results were achieved with the threshold of 0.5 certainty level. At this level the algorithm assigns equal importance to precision and recall. As the threshold increases, precision improves at the expense of recall. On the other hand, if the threshold is reduced, recall is improved at the expense of precision. Thus the optimal threshold to be set will depend on the use case and the particular objectives of the project at hand. If the requirement is to be as comprehensive as possible, a lower threshold layered with a manual oversight or a domain-specific set of rules may be advisable. On the other hand, if the requirement is to detect unambiguous markers, a higher threshold will be advisable.

7. CONCLUSION

Extracting pragmatic markers embedded within documents is a complex task because, by definition, they are “hidden.” A 2-stage augmented intelligence model can successfully extract such markers from a set of documents that are developed using similar guidelines and ontology. This research project developed the system for extracting implied and suggested actions within a specific domain of economic analysis. Future research and model improvements will demonstrate its potential effectiveness in extracting pragmatic markers of various kinds (for example, threats, warnings, complaints, criticisms, accolades, etc.) across a range of domains and for a variety of useful purposes. It may also point to ways to help reduce the deception and manipulation of consumers of online content by helping platform curators filter out or at least highlight content prepared with intent to mislead or misrepresent facts [10][11]. These are areas of future research and model improvement.

REFERENCES

- [1] Frank Meng and Craig Morioka, “Automating the Generation of Lexical Patterns For Processing Free Text in Clinical Documents”, in Journal of American Medical Informatics Association, May 2015..
- [2] LL Berland, SG Silverman, RM Gore, WW Mayo-Smith, AJ Megibow and J. Yee, “Managing Incidental Findings on Abdominal CT”, in Journal of American Coll Radiology, 2010
- [3] Alex Marin, Roman Holenstein, Ruhi Sarikaya and Mari Ostendorf, “Learning Phrase Patterns for Text Classification Using a Knowledge Graph and Unlabeled Data”, Research Report by Microsoft Corporation and University of Washington, 2014.
- [4] Theresa Wilson, Janyce Wiebe and Paul Hoffman, “Recognizing Contextual Polarity in Phrase_Level Sentiment Analysis”, in Proceedings of HLTP-EMNLP, 2005.
- [5] Meliha Yetisgen-Yildiz, Martin L. Gunn, Fei Xia and Thomas H. Payne, “A Text Processing Pipeline to Extract Recommendations from Radiology Reports”, in Journal of Biomedical Informatics 46, January 2013
- [6] Y. Ko, “A Study of Term Weighting Schemes Using Class Information for Text Classification”, in Proceeding sof the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2012
- [7] Betty J. Birner, Introduction to Pragmatics, Wiley-Blackwell Publications, 2012.
- [8] Alex Marin, Bin Zhang and Mari Ostendorf, “Detecting Forum Authority Claims in Online Discussions”, in Proceedings of the Workshop on Language in Social Media, 2011.
- [9] George J. Klir and Mark J. Wierman, Uncertainty Based Information: Elements of Generalized Information Theory, Springer Werhlag Berlin Heidelberg GmbH, 1999.
- [10] Bruce Fraser, “Pragmatic Markers”, Pragmatics 6:2, International Pragmatics Association, 1992
- [11] Peter Furko, “Manipulative Uses of Pragmatic Markers in Political Discourse”, Palgrave Communications Article, 2016.
- [12] Jaspal Singh, “Review of Poststructuralist Discourse Analysis”, Journal of Pragmatics 88, 2016.
- [13] D. Crowe, M. La Pierre and M. Kebritchi, “Knwoledge Based Augmented Intelligence Technology”, Tech Trends Vol 61 Issue 5, September 2017.
- [14] J. Gao, M. Galley, and L. Li, “Neural Approaches to Conversational AI”, Foundations and Trends in Information Retrieval, Vol. 13, February 2019.

AUTHORS

Vijay Perincherry is data scientist involved in research related to the application of AI techniques in behavioural analysis, language and discourse processing

David White is a Senior Architect with Oteemo leading the development of cloud-based platforms to enable big data analytics

Staci Warden is the executive director of the Center for Financial Markets at the Milken Institute, where she leads initiatives on strengthening capital markets, access to capital, financial education and financial-markets solutions