ODNI-OUSD(I) Xamine Challenge: Machine Verification of Collected Information

***Abstract*— A four phase system to verify the claims contained within an analytic product is described. Areas of concern regarding implementation of the design are covered, as is a timeline during which national intelligence activities utilizing the system, when compared to foreign powers, could be expected to remain asymmetric.**

INTRODUCTION

The Office of the Director of National Intelligence, and the Office of the Under Secretary of Defense for Intelligence, sought ideas and descriptions of a viable technical approach for enabling the automated inspection and validation of uncertain information prior to the dissemination of machine-generated intelligence products.

A four phase system to accomplish this is presented:

(I) Extracting: facts and claims in a given product are identified.
(II) Ranking: the identified claims and facts are assigned a priority and ranked; in addition to being a requirement for accurate scoring, any system capable of scaling to a useful level must have the ability to prioritize verification tasks.
(III) Verifying: the claims are verified in a modular and generic fashion; this facilitates scaling of the system as sources are added and removed (e.g. databases, newspaper articles, social media content, and so on).
(IV) Scoring: a weighted score is calculated and assigned to the product in question.

Finally, comments are made on when other nations could be expected to develop equivalent capabilties.

I. EXTRACTING

A standard artificial neural network, trained in a supervised fashion, should yield satisfactory performance. [2] Presidential election debates provide readily available labeled examples, as well as well researched and widely accepted ground truths (PolitiFact, NPR, and so forth). [3] The trained network is then employed to identify and extract facts from analytic products.

As an alternative, IBM provides a cloud API for these types of tasks. It is assumed, however, that their analytics should not include the contents or existence of API calls.

II. RANKING

Features used for ranking include bias, [5] sentiment, [4] assertiveness, and subjectivity. [1]

III. VERIFYING

Any text source can be used to verify claims; to ensure interoperability and proper functioning, all sources must have an associated data structure. It should characterize the following:

- A description of the source.
- Prestige of the source, i.e. how trustworthy it is, and in what disciplines we consider it a gold standard.
- How long – if at all – completion of a validation job should be postponed when the source is unavailable.
- Allowed forms of access.
- Record keeping data provided, so that detailed logs are possible when utilizing this source.

Standard XML is ideal for this purpose, since it is both human readable and nearly universal.

By sticking to textual sources, we can make use of text message records, emails, bank transactions, social media posts, blogs, news articles, Wikipedia, or even WikiLeaks.

The size of graphic media (pictures, video, and so forth) may pose a significant problem for the system at scale; if this functionality is required, a great deal of algorithmic optimization, as well as investigation in to current industry methods, would be necessary.

IV. SCORING

A weighted final score is provided in a similar fashion to TF-IDF (term frequency-inverse document frequency), as follows:

$$tf - idf_{t,d} = (1 + \log tf_{t,d}) \cdot \log \frac{N}{df_t}$$

Because of the detailed record keeping that is performed throughout the verification procedure, a more detailed description of claims and sources which refute or support said claim can be provided. The scoring formula can also be modified as necessary to suit the needs of users.

EFFICIENCY

Without ASICs (application-specific integrated circuits), the speed of this system could be expected to rival a typical search for files on a laptop computer. With grid (distributed) computing environments or the design and purchase of ASIC units, speeds that match the nation's concurrent Google queries should be possible.

TIMELINE PREDICTIONS

An absolute necessity to implement this system is a language-specific corpus of high quality and hand labeled examples; this is a necessity because it is the basis of much of the training of the artificial neural networks. Arabic currently has no such corpus (the source of this statement is ClaimRank [3]). Using machine translations of English training data is possible, but would introduce an unpredictable source of error; the quality of the system's features would degrade, but the severity of that degradation is extremely difficult to quantify.

It should be assumed that any state level aggressor could make use of any publicly available verification sources by using automated translation. Whether sources that are not public are available or not would vary by nation and region.

In areas with heavy censorship (e.g. China), maintaining existing infrastructure while still allowing unfettered access the way an American implementation would have may prove an insurmountable dilemma. It is this author's understanding that countries generally purchase hardware level units to censor networks within their borders, which would only compound the issue of interoperability. Generally speaking, the less free the country's internet is, the more difficulty said country would have in implementing a system of equivalent functionality.

Ultimately, any foreign power could be expected to have equivalent functionality – with the caveats mentioned above – within one to two years of beginning the process.

REFERENCES

[1] GENCHEVA, P., NAKOV, P., MÀRQUEZ, L., BARRÓN-CEDEÑO, A., AND KOYCHEV, I. A context-aware approach for detecting worth-checking claims in political debates. 267–276.

[2] HASSAN, N., LI, C., AND TREMAYNE, M. Detecting check-worthy factual claims in presidential debates. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (New York, NY, USA, 2015), CIKM '15, ACM, pp. 1835–1838. http://ranger.uta.edu/~cli/pubs/2015/claimbuster-cikm15-hassan.pdf.

[3] JARADAT, I., GENCHEVA, P., BARRÓN-CEDEÑO, A., MÀRQUEZ, L., AND NAKOV, P. Claimrank: Detecting check-worthy claims in arabic and english. *CoRR abs/1804.07587* (2018). http://arxiv.org/abs/1804.07587.

[4] LIU, B., HU, M., AND CHENG, J. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web* (2005), ACM, pp. 342–351.

[5] RECASENS, D. On the articulatory classification of (alveolo) palatal consonants. *Journal of the International Phonetic Association 43*, 1 (2013), 1–22.