High Dimensional Latent Space Variational AutoEncoders for Fake News Detection

Saad Sadiq, Nicolas Wagner, Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL, USA
Emails: {s.sadiq, nxw178, shyu}@miami.edu

Daniel Feaster
Department of Public Health Sciences
University of Miami - Miller School of Medicine
Miami, FL, USA
dfeaster@med.miami.edu

***Abstract*—With the advent of social media and cell phones, news is now far more reaching and impactful than ever before. This comes with the exponential increase in fake news that blurs the lines of reality and holds the power to sway public opinion. To counter the impact of fake news, several research groups have developed novel algorithms that could fact check news as a human would do. Unfortunately, natural language processing (NLP) is a complicated task because of the underlying hidden meanings in human communication. In this paper, we propose a novel method that builds a latent representation of natural language to capture its underlying hidden meanings accurately and classify fake news. Our approach connects the high-level semantic concepts in the news content with their low-level deep representations so that the complex news text consisting of satire, sarcasm, and purposeful misleading content can be translated into quantifiable latent spaces. This allows us to achieve very high accuracy, surpassing the scores of all winners of the fake news challenge.**

***Keywords*-Fake news; latent representation; VAE (Variational AutoEncoder); LSTM (Long Short Term Memory);**

I. Introduction

In the recent years, “fake news” have been a buzzword used as a descriptor for publications or statements. These publications contain false or highly misrepresented facts used to manipulate people’s opinions or perceptions. These facts are quickly spread through social networks such as Facebook, Twitter, and Reddit, where users do not research the truthfulness of the articles [1], as seen in Figure 1.

Figure 1. Evolution of Fake News on Social Media

Since language and communication are fluid concepts, content and meaning of information can be conveyed and interpreted in a variety of ways depending on the reader/writer. Even words that have the same dictionary definition can carry different connotations depending on the context. As noted in [2], there are other challenges on how meaning is extracted from texts such as semantical syntax (grammar), morphology (pre/suffixes), and pragmatics (domain knowledge to interpret isolated pieces). As described in [3], sarcasm is a form of “indirect speech” that becomes difficult to identify and classify because it forms the basis on common idioms, play-on-words, and paradoxes. Sarcasm plays an important role in identifying fake news as it represents opinionated pieces of writing having higher potential of distorting or misrepresenting factual information. Thus, it is necessary to model a statistical method wherein rather than accounting for detail-sensitive rules, we enlist broader directives and let the sampling of clustered data points handle disambiguation [4].

The remaining of the paper is organized as follows. Section II covers the existing methods in fake news detection. The proposed framework and its underlying methods are presented in Section III, followed by the experiment and results in Section IV. Finally, Section V concludes the paper with discussions about future aspects.

II. Previous Work

In many real-world multimedia applications such as news, blogs, and social media, large quantities of unlabeled data are generated everyday, giving rise to the challenging semantic gap problem [5]–[12] which is to reduce the gap between high level semantic concepts and their low level features [13]–[17]. Despite rigorous research endeavors, this remains one of the most challenging problems in information sciences, particularly for text data. Moreover, the majority of the cases belong to only a few classes (i.e., the majority classes) and far fewer data instances belong to the minority classes. The minority classes, however, often represent the unusual and interesting events with high entropy values [18]–[20]. In content-based information retrieval applications, most classifiers are modeled by exploring data statistics. Hence, they may be biased towards the majority classes and show poor classification scores on the minority classes [21]–[23].

The features of deceptive language and deception in online dating profiles were studied, respectively [24], [25]. Both studies found that there exist identifiable characteristics in deceptive human communications, which lays the foundation for linguistic features based fake news detection. For example, linguistic and visual characteristics, social user, post, and network were used to propagate the news, as well as style, objectivity, and stance for their analysis [26]. Their data-oriented approach required a large dataset, and assembling a reliable dataset of fake news is extremely labor intensive. A feature of fake news is that fake articles tend to have explosive popularity more often than real articles [1]. How different news spread through the internet based on their truthfulness were studied in [27], [28], as their propagation networks are not similar at the beginning of their life-cycle. The life cycle of news was studied more broadly in [27]; while [28] concentrated on the Facebook communities and connections, and how they shape echo chambers that amplify fake news. In [29], crowdsourcing techniques and how users interact with news were studied through logistic regression and boolean labeling, yielding 99% accuracy. Feature based statistical analysis of three datasets - Satirical, True, and Fake news was conducted and it was concluded that fake news is more similar to satire than real news in terms of their style and structure, as they use heuristics rather than facts for persuasion [30]. Conversely, [31] found that fake news have much more in common with real news containing satire. They also concluded that fake news disproportionately relies on emotion, rather than quality of information, to earn credibility.

The most popular fact checking websites such as Snopes.com, Politifact.com and FactCheck.org rely on human experts to check for truthfulness. However, this approach is slow and inefficient as articles have to be researched individually and most publications receive a mixed truthfulness rating.

III. Proposed Framework

The proposed method is based on the philosophy of modeling complex non-relational context in natural language such as sarcasm and satire using variational autoencoders (VAEs). The VAEs help to obtain disentangled embeddings of fake news in the form of high dimensional latent representations. To the best of our knowledge, this is the first attempt to use latent representations to classify fake news. The obtained results support our motivation that VAEs play a crucial role in clustering texts in the hidden space and generating a high score as compared to the other research teams that took part in the challenge. Figure 2 presents our proposed novel framework which has four major components, namely Bi-directional Long Short Term Memory (LSTM), Variational Autoencoders, Word Embeddings, and Synthetic feature creation. The final classification step is performed using Random Forests, as described in Section III-D.

As shown in Figure 2, a bi-directional LSTM end-to-end with a VAE using Keras 2.2 is first trained on Nvidia Titan Volta architecture (with 6 GPU clusters, 5120 CUDA Cores, 12GB Memory, and 640 Tensor Cores). Further details of the VAE architecture are provided in Section III-B. The resultant 300 dimensional latent space is augmented by 300 dimensional embeddings for each snippet using Google's news corpus. The 600 latent space variables are then used to create synthetic features (see more details in Section III-C). These 900 deep latent high dimensional features are then passed to train a random forest. The Out-of-Bag predictions are used for the final random forest based classifier.

A. The Datasets

The pre-trained word2vec model from the Google's News Corpus is used. Since the quality of the word vectors increases a lot with the training corpus size, 100 billion words from the corpus are used. There are 3 million words and phrases from the live stream of Google News. Each word is mapped to a 300 dimensional English word vector.

For the competition data, all training data and most of the test data were derived from the "Emergent" dataset (available at http://www.emergent.info/). The dataset contains about 2600 articles, clustered around 300 distinct claims. For each claim, there are between 5 and 20 articles in its cluster, each of which was hand-labeled as agree, disagree, or discuss relative to the central claim of that cluster. Further data was augmented by mixing and matching headlines and the story bodies from the 2600 articles, carrying over the appropriate labels from Craig's hand-labeling. After the mixing-and-matching process, there are about 75K headline/story body pairs, which are then split into a 50K training set and a completely disjoint 25K test set.

B. Variational AutoEncoders (VAEs)

In this paper, VAEs are used to learn a lower-dimensional feature representation from unlabeled training data. Take for example a case where the input data is $\mathbf{x}$ and we want to learn some feature vector $\mathbf{z}$. Then an encoder that uses a mapping function to map the input data to the feature $\mathbf{z}$ is utilized. The hidden space $\mathbf{z}$ is usually specified to be smaller than $\mathbf{x}$ to avoid trivial solutions and serves as a form of dimensionality reduction. Then, $\mathbf{z}$ represents the most important features in $\mathbf{x}$ that can capture meaningful factors of the variation in data.

The latent feature representation $\mathbf{z}$ is utilized to reconstruct original data by decoding them to the same dimensionality as $\mathbf{x}$, thus the term autoencoder – encoding itself. The same type of network as the encoder is used so it ends up being symmetric. Since we are dealing with text data, the Bi-directional LSTM Networks are used before and after the

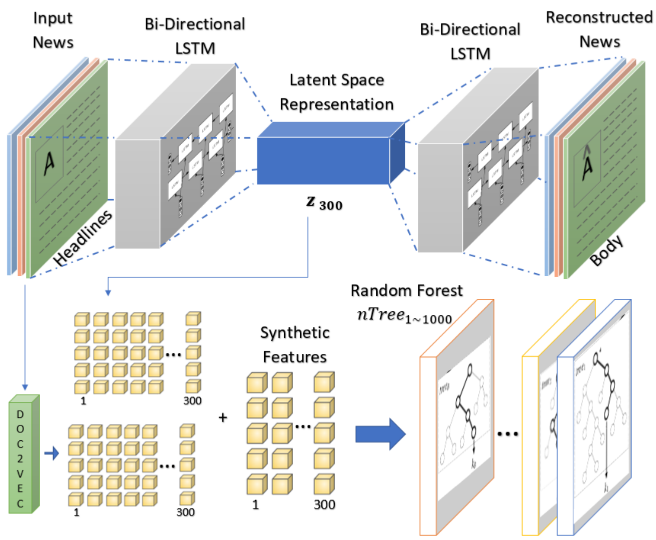

Figure 2. The overall architecture of the proposed high dimensional latent space framework

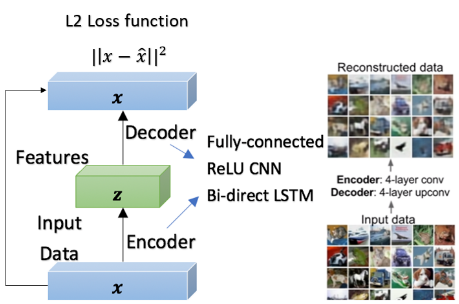

Figure 3. VAE training process such that the features can be used to reconstruct the original data

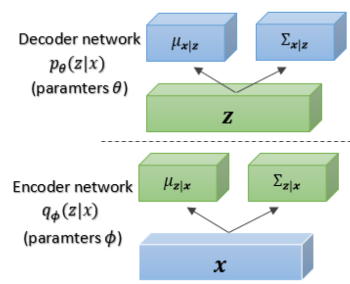

Figure 4. The probabilistic components: Mean and diagonal covariance of $\mathbf{z}|\mathbf{x}$ and $\mathbf{x}|\mathbf{z}$ of the Gaussian encoder and decoder

VAE as shown in Figure 3. In order to reconstruct the input data, the L2 loss function is adopted to make sure that the difference between the reconstructed data is very similar to the original data.

The data likelihood $p(\mathbf{x})$ is defined as taking the expectation over all possible values of $\mathbf{z}$, which is continuous, and the expression with the latent $\mathbf{z}$ can be obtained.

$$p_\theta(\mathbf{x}) = \int p_\theta(\mathbf{z}) p_\theta(\mathbf{x}|\mathbf{z}) dz \tag{1}$$

However, we are unable to take the gradient and maximize this likelihood because the integral is intractable. Here $p_\theta(\mathbf{z})$ is a simple Gaussian prior, $p_\theta(\mathbf{x}|\mathbf{z})$ is a decoder neural network, and given $\mathbf{z}$, $p(\mathbf{x}|\mathbf{z})$ can be obtained by solving the neural network. In addition, since it is intractable to compute $p(\mathbf{x}|\mathbf{z})$ for every possible value of $\mathbf{z}$, the data likelihood is intractable. Similarly, the posterior density also becomes intractable due to the intractable data likelihood $p_\theta(\mathbf{x})$ in the denominator.

$$p_\theta(\mathbf{z}|\mathbf{x}) = \frac{p_\theta(\mathbf{x}|\mathbf{z}) p_\theta(\mathbf{z})}{p_\theta(\mathbf{x})} \tag{2}$$

The solution that will enable us to learn this model is that in addition to the decoder network modeling $p_\theta(\mathbf{x}|\mathbf{z})$, an additional encoder network $q_\phi(\mathbf{z})$ that approximates $p_\theta(\mathbf{x}|\mathbf{z})$ is defined. This allows us to derive a lower bound on the data likelihood that is tractable and can be optimized. Since we are modeling a probabilistic generation of data in variational autoencoders, the encoder and decoder networks are probabilistic. Our encoder network $q_\phi(\mathbf{z})$ with parameters ϕ is going to output a mean $\mu_{\mathbf{z}|\mathbf{x}}$ and diagonal covariance $\sum_{\mathbf{z}|\mathbf{x}}$. This will be the direct output of our encoder network. A similar method can be performed for the decoder network $p_\theta(\hat{\mathbf{x}}|\mathbf{z})$ which is going to start from $\mathbf{z}$ and outputs the mean $\mu_{\hat{\mathbf{x}}|\mathbf{z}}$ and diagonal covariance $\sum_{\hat{\mathbf{x}}|\mathbf{z}}$ as shown in Figure 4.

C. Synthetic Features

A collection of random forests are used with different combinations of the hyperparameters (*nodesize*, *mtry* and *nsplits*) to generate the synthetic features. These are the three key factors to optimize the maximum throughput from a random forest. The predicted values from these machines are used as synthetic features augmented to the original dataset. The final dataset also contains the original features (hidden vectors and doc2vec vectors) used in constructing the synthetic forest. Synthetic features are calculated using out-of-bag (OOB) data to avoid overfitting the training data. To guarantee that error rates and variable importance are regularized, same sized bootstrap draws are performed on all trees in the construction of the synthetic forest.

D. Random Forests

Random forests were chosen as the final classification stage of the framework because of their ability to handle very high dimensional spaces. Each tree in the random forest collection is grown non-deterministically with a two-stage method. In the first stage, randomization is induced in each tree by randomly selecting sub-sampled data (bootstrapping) from the original data. The second stage randomization is applied at the node level, where each node is split by randomly selecting a variable from the sub-sampled variables and only those variables are utilized to get the best possible split.

The observed data is assumed to be independently drawn from the joint distribution of ($\mathbf{X}$, Y) and comprises $n*(p+1)$ samples, namely $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)$. $\mathbf{X}$ is an n by p matrix indicating the total number of video frames (or samples) and their features Y, where $\mathbf{X}=[\mathbf{x}_1, ..., \mathbf{x}_n]^T$, $Y=[y_1, ..., y_n]^T$, $\mathbf{x}_i$ is the subsampled vector (of size 1 by p) from $\mathbf{X}$ for the i^{th} sample, Y indicates the vector of outcome variables (y_i, i=1 to n), and p is the total number of features (or dimensions).

The forest is built by growing the trees based on a random vector θ_k such that the tree predictor $h(\mathbf{x}, \theta_k)$ represents a predicted probability specified by the class, ranging from 0 to 1. Thus, the vector θ_k contains the predicted probabilities of the outcome variable Y. The final predictions are defined as the unweighted average over the collection of predictor trees as shown in Equation (3), where $h(\mathbf{x}; \theta_k), k = 1, ..., ntree$ are the collection of the tree predictors and $\mathbf{x}$ represents the observed input variable vector of length *mtry* with the associated i.i.d random vector θ_k.

$$\overline{h}(\mathbf{x}) = (1/ntree) \sum_{k=1}^{ntree} h(\mathbf{x}; \theta_k). \quad (3)$$

As $k \to \infty$, the Law of Large Numbers ensures

$$E_{\mathbf{X},Y}(Y - \overline{h}(\mathbf{X}))^2 \to E_{\mathbf{X},Y}(Y - E_\theta(\mathbf{X}; \theta))^2, \quad (4)$$

where θ represents the predicted probabilities of the outcome variable averaged over *ntree* trees. The convergence in Equation (4) implies that the random forests do not overfit.

IV. Experiment and Results

This paper builds the experiment on an existing fake news detection competition called the fake news challenge (FNC) [32] whose objective was to apply the state-of-the-art Artificial Intelligence (AI) and NLP methods to counter the fake news problems. Quantifying the veracity of a news article is a challenging and cumbersome task, even for the trained experts. Hence, following the competition helps us benchmark our results against an internationally renowned dataset. FNC models the fake news detection by introducing the 'Stance Detection' which they consider a building block in the machine learning fact-checking pipeline. Stance Detection consists of predicting a relative perspective of two text snippets related to each topic, claim, or issue. The task was to estimate the stance of a body text from a news article relative to a headline. Specifically, the body text may agree, disagree, discuss or be unrelated to the headline.

Table I
Confusion Matrix for Out-Of-Bag Samples

	Predicted			
Actual	Agree	Disagree	Discuss	Unrelated
Agree	5023	14	81	463
Disagree	24	1383	51	79
Discuss	45	73	12036	1219
Unrelated	518	859	4113	49404

There were fifty teams in total who took part in the first international fake news challenge (FNC-1). The methods were trained using data extracted from the "Emergent" dataset [1]. The participants were scored using the Codalab submission platform. The scoring system produced a raw score based on the differentially weighted scoring metric. The relative score was obtained by normalizing by the maximum possible score on the test set.

A. Results

Each was rewarded 0.25 points if it correctly estimates the correlation between a body/headline pair. The top three teams scored between 96.5% and 97% on the related/unrelated subtask. The proposed model, however, performed worse and was able to achieve only 90% score on this task. The main challenge was to predict if the body agreed with, disagreed with, or simply discussed the headline. This accounted for an extra 0.75 points for each correct labeling. The test/validation dataset contained 7064 pairs, out of the total 25414, that were to be classified as either agree, disagree, or discuss. On the 7064 validation samples for the 1-of-3 accuracy test, other teams scored 64.1%, 64.1% and 63.9%, respectively. The proposed model very accurately labeled agree, disagree, and discuss pairs with 89% score and increased the performance of the model dramatically. The confusion matrix is illustrated in Table I.

Further Model metrics are provided in Table II. The proposed model achieves very high precision and recall if trained only on the Agree/Disagree pairs or Agree/Disagree/Discuss pairs. Since the "unrelated" label is the majority class as compared to other classes, its influence

Table II
MODEL METRICS FOR THE PROPOSED FRAMEWORK

	Precision	Recall	Acc	F1	Macro Precision	Macro Recall	Macro F1	Micro Acc	Majority Class	Random Guess	Kappa Statistic
Agree/Disagree	97.8	93.85	97.17	95.11	97.8	93.83	95.65	97.11	65.26	0.33	94.15
Agree/Disagree/ Discuss	96.43	94.06	97.11	95.65	96.4	94.07	95.11	97.17	78.4	0.5	91.31
All Classes	79.85	89.99	89.99	83.91	79.85	89.99	83.91	89.99	72.81	0.25	78.32

Table III
COMPARISON WITH THE LEADING MODELS IN THE FAKE NEWS CHALLENGE

Team Name	**Score**	**Relative Score**
Talos Intelligence	9556.50	82.02%
TU Darmstadt	9550.75	81.97%
UCL	9521.50	81.72%
Chips Ahoy!	9345	80.21%
CLUlings	9289.50	79.73%
unconscious bias	9285	79.69%
Proposed	**10510**	**88.21%**

is detrimental to the scores. The proposed framework was compared to other state-of-the-art methods from the FNC. This is because stance detection is highly subjective and those models not trained on the FNC dataset could not be directly compared for their predictive accuracy.

The competition result was evaluated on the benchmarks provided by FNC and displayed in Table III. Here, the proposed model outperforms other top contenders in the challenge. It was observed that other teams did not perform well on the hand labeled 266 samples appended at the end of the test data, which were all agree/disagree/discuss pairs. The 266 additional pairs were samples taken from different publications than the original source and labeled by different people. The top teams did well on the related/unrelated task, but only 40% of the time. This was higher than a coin flip, but the scores were less as compared to the Emergent-derived test pairs where they achieved nearly 64%.

When tested a simply "majority rule" average of the highest scoring teams (i.e., take the answer from the top team on a particular test example unless the 2nd and 3rd teams had the same answer as each other on that example which differs from the top team's answer, in which case take their answer), it was able to dramatically improve on any of the top three scores. This simple "ensemble of experts" approach got a score of 9702 (83.28%), compared with the scores for the top three teams individually of 9556.5 (82.02%), 9550.8 (81.97%), and 9521.5 (81.72%). The performance comparison demonstrates that our proposed high dimensional latent space model achieves high classification accuracy by capturing the hidden meaning behind each news article via several latent space representations and the augmented high dimensional dataset.

V. CONCLUSION

Machine understanding of textual information can be a very challenging problem to solve. In this paper, a novel framework that classifies fake news with very high accuracy using the data from the internationally renowned competition (called the fake news challenge) is proposed. Our framework detects fake news using variational inferencing on high dimensional latent spaces and synthetic features that represent hidden relationships in text data. The experimental results have shown that our proposed framework is able to surpass all the top contenders. There is still room for improvement and our future work is to develop a multi-modal method that can also incorporate images with the news content.

REFERENCES

[1] C. Silverman, "This analysis shows how viral fake election news stories outperformed real news on facebook," *BuzzFeed News*, vol. 16, 2016.

[2] G. G. Chowdhury, "Natural language processing," *Annual Review of Information Science and Technology*, vol. 37, no. 1, pp. 51–89, 2003.

[3] O. Tsur, D. Davidov, and A. Rappoport, "ICWSM - A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews." in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, 2010, pp. 162–169.

[4] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 544–551, 2011.

[5] S.-C. Chen, M.-L. Shyu, C. Zhang, and R. L. Kashyap, "Identifying overlapped objects for video indexing and modeling in multimedia database systems," *International Journal on Artificial Intelligence Tools*, vol. 10, no. 4, pp. 715–734, 2001.

[6] S.-C. Chen, S. Rubin, M.-L. Shyu, and C. Zhang, "A dynamic user concept pattern learning framework for content-based image retrieval," *IEEE Transactions on Systems, Man, and Cybernetics: Part C*, vol. 36, pp. 489–495, November 2006.

[7] Q. Zhu, L. Lin, M.-L. Shyu, and D. Liu, "Utilizing context information to enhance content-based image classification," *International Journal of Multimedia Data Engineering and Management*, vol. 2, no. 3, pp. 34–51, 2011.

[8] L. Lin, C. Chen, M.-L. Shyu, and S.-C. Chen, "Weighted subspace filtering and ranking algorithms for video concept retrieval," *IEEE Multimedia*, vol. 18, no. 3, pp. 32–43, March 2011.

[9] Q. Zhu and M.-L. Shyu, "Sparse linear integration of content and context modalities for semantic concept retrieval," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 2, pp. 152–160, June 2015.

[10] Y. Yan, Y. Liu, M.-L. Shyu, and M. Chen, "Utilizing concept correlations for effective imbalanced data classification," in *Proceedings of the IEEE 15th International Conference on Information Reuse and Integration*, Aug 2014, pp. 561–568.

[11] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, "Feature selection using correlation and reliability based scoring metric for video semantic detection," in *Proceedings of the Fourth IEEE International Conference on Semantic Computing*, 2010, pp. 462–469.

[12] L. Lin and M.-L. Shyu, "Weighted association rule mining for video semantic detection," *International Journal of Multimedia Data Engineering and Management*, vol. 1, no. 1, pp. 37–54, 2010.

[13] M.-L. Shyu, K. Sarinnapakorn, I. Kuruppu-Appuhamilage, S.-C. Chen, L. Chang, and T. Goldring, "Handling nominal features in anomaly intrusion detection problems," in *Proceedings of the 15th International Workshop on Research Issues in Data Engineering: Stream Data Mining and Applications*, 2005, pp. 55–62.

[14] Y. Yan, M. Chen, S. Sadiq, and M.-L. Shyu, "Efficient imbalanced multimedia concept retrieval by deep learning on spark clusters," *International Journal of Multimedia Data Engineering and Management*, vol. 8, no. 1, pp. 1–20, 2017.

[15] S. Sadiq, Y. Yan, M.-L. Shyu, S.-C. Chen, and H. Ishwaran, "Enhancing multimedia imbalanced concept detection using vimp in random forests," in *Proceedings of the 2016 IEEE 17th International Conference on Information Reuse and Integration*, July 2016, pp. 601–608.

[16] S. Sadiq, Y. Tao, Y. Yan, and M.-L. Shyu, "Mining anomalies in medicare big data using patient rule induction method," in *Proceedings of the 2017 IEEE Third International Conference on Multimedia Big Data*, 2017, pp. 185–192.

[17] S.-C. Chen, S. Sista, M.-L. Shyu, and R. Kashyap, "Augmented transition networks as video browsing models for multimedia databases and multimedia information systems," in *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*, 1999, pp. 175–182.

[18] S. Sadiq, Y. Yan, A. Taylor, M.-L. Shyu, S.-C. Chen, and D. Feaster, "Aafa: Associative affinity factor analysis for bot detection and stance classification in twitter," in *Proceedings of the 2017 IEEE International Conference on Information Reuse and Integration*, 2017, pp. 356–365.

[19] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys*, vol. 51, no. 5, Article No. 92, 2018.

[20] S. Sadiq, M. Zmieva, M.-L. Shyu, and S.-C. Chen, "Reduced residual nets (red-nets): Low powered adversarial outlier detectors," in *Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration*. IEEE, 2018, pp. 436–443.

[21] M.-L. Shyu, T. Quirino, Z. Xie, S.-C. Chen, and L. Chang, "Network intrusion detection through adaptive sub-eigenspace modeling in multiagent systems," *ACM Trans. Auton. Adapt. Syst.*, vol. 2, no. 3, Sep. 2007.

[22] M.-L. Shyu, S.-C. Chen, M. Chen, and C. Zhang, "A unified framework for image database clustering and content-based retrieval," in *Proceedings of the 2nd ACM International Workshop on Multimedia Databases*, New York, NY, USA, 2004, pp. 19–27.

[23] L.-C. Chen, J.-W. Hsieh, Y. Yan, and D.-Y. Chen, "Vehicle make and model recognition using sparse representation and symmetrical SURFs," *Pattern Recognition*, vol. 48, no. 6, pp. 1979 – 1998, 2015.

[24] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 309–319.

[25] C. L. Toma and J. T. Hancock, "Reading between the lines: linguistic cues to deception in online dating profiles," in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 2010, pp. 5–8.

[26] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[27] Z. Zhao, J. Zhao, Y. Sano, O. Levy, H. Takayasu, M. Takayasu, D. Li, and S. Havlin, "Fake news propagate differently from real news even at early stages of spreading," *arXiv preprint arXiv:1803.03443*, 2018.

[28] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," *arXiv preprint arXiv:1708.07104*, 2017.

[29] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some like it hoax: Automated fake news detection in social networks," *arXiv preprint arXiv:1704.07506*, 2017.

[30] B. D. Horne and S. Adali, "This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," *arXiv preprint arXiv:1703.09398*, 2017.

[31] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," *arXiv preprint arXiv:1702.05638*, 2017.

[32] D. Pomerleau and D. Rao, "Fake news challenge stage 1 (fnc-i): Stance detection," http://www.fakenewschallenge.org/, 2017, accessed October 27th, 2018.