2023 IEEE International Conference on Web Services (ICWS) | 979-8-3503-0485-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICWS60048.2023.00097

Detecting Intents of Fake News Using Uncertainty-Aware Deep Reinforcement Learning

Zhen Guo
Virginia Tech
Falls Church VA, USA
zguo@vt.edu

Qi Zhang
Virginia Tech
Falls Church VA, USA
qiz21@vt.edu

Qisheng Zhang
Virginia Tech
Falls Church VA, USA
qishengz19@vt.edu

Lance M. Kaplan
US Army Research Laboratory
Adelphi MD, USA
lance.m.kaplan.civ@army.mil

Audun Jøsang
University of Oslo
Oslo, Norway
audun.josang@mn.uio.no

Feng Chen
University of Texas at Dallas
Richardson TX, USA
feng.chen@utdallas.edu

Dong H. Jeong
University of the District of Columbia
Washington DC, USA
djeong@udc.edu

Jin-Hee Cho
Virginia Tech
Falls Church VA, USA
jicho@vt.edu

***Abstract*—Intent mining is critical for controlling the spread of false information across online social networks (OSNs). To this end, we develop deep reinforcement learning (DRL) agents guided by a delayed reward based on intent prediction using a classifier of long short-term memory (LSTM). Additionally, we incorporate an uncertainty-aware function that leverages subjective opinions derived from Subjective Logic (SL). Through evaluation using an annotated fake news tweet dataset, our results demonstrate that our intent classification framework surpasses competing methods in terms of intent accuracy. Our intent mining solutions using DRL algorithms can support effective and efficient intervention strategies for fake news spreading on OSNs.**

***Index Terms*—Intent mining, fake news, deep reinforcement learning, Long Short-Term Memory, online social network**

I. Introduction

Fake news in online social media platforms has become a pressing concern in today's information age. Misinformation spreads rapidly, leading to significant social, political, and economic implications. Understanding the dynamics behind the propagation of fake news is crucial for devising effective strategies to mitigate its adverse effects. While it is commonly assumed that users share fake news with malicious intent, recent social science studies [1] have shed light on the unintentional behaviors associated with fake news sharing. Users may unknowingly disseminate false information due to a lack of judgment capabilities or even with good intentions, such as entertaining friends or altruistically helping others (e.g., raising funds for a noble cause). These findings challenge the prevailing notion that all fake news propagators act out of malice, urging us to explore the intent behind the spread of fake news more comprehensively. We can tailor our mitigation efforts based on users' various intents and enhance the effectiveness. Thus, we propose an intent classification framework by deep reinforcement learning (DRL). By analyzing the textual content of verified fake news, we can extract embedding features and structure representations and optimize those representations from texts, which have been extensively studied for various text classification tasks [10]. However, existing DRL models for text classification suffer from a limitation in local response capabilities. These models typically assign equal weight to all steps of the Markov decision process, relying solely on the final state of the text encoding process to determine future or delayed rewards [4]. To address this limitation, we propose modifying existing DRL models with an uncertainty-aware immediate reward.

Our work aims to enhance the local response of DRL models by introducing a belief model called *Subjective Logic* [5]. This belief model estimates multidimensional uncertainty based on intent predictions at each local encoding step. By incorporating this uncertainty-aware local reward, we enable DRL agents to trust delayed rewards while considering local responses, thereby improving the overall performance of intent prediction. In general, this approach enables us to effectively analyze the intents associated with fake news propagation.

This work is partly supported by the Army Research Office under Grant W91NF-20-2-0140 and by NSF Grant 2107449, 2107450, and 2107451.

II. Intent Prediction Framework Design

A. Intent Classes and Data Annotations

From online social network (OSN) users' major intents of spreading fake news, by several social science studies [1, 6, 7], we identify several intent classes behind the spread of fake news on OSNs. These include "Information Sharing", where users unintentionally share fake news due to a lack of fact-checking or knowledge. "Socialization" involves sharing news for self-promotion and expanding social connections. The intent of "Political Campaign" focuses on creating false perceptions and manipulating public opinions using fake news. "Emotion Venting" refers to the propagation of fake news triggered by users' emotional states. Lastly, "Rumor Propagation" involves the dissemination of fake news linked to uncertain rumors. By understanding these intent classes, we gain insights into users' motivations and can develop targeted strategies to mitigate the spread of fake news effectively.

To conduct data-centric intent mining, we require datasets that provide both fake news and intent class labels. As such,

979-8-3503-0485-5/23/$31.00 ©2023 IEEE
DOI 10.1109/ICWS60048.2023.00097

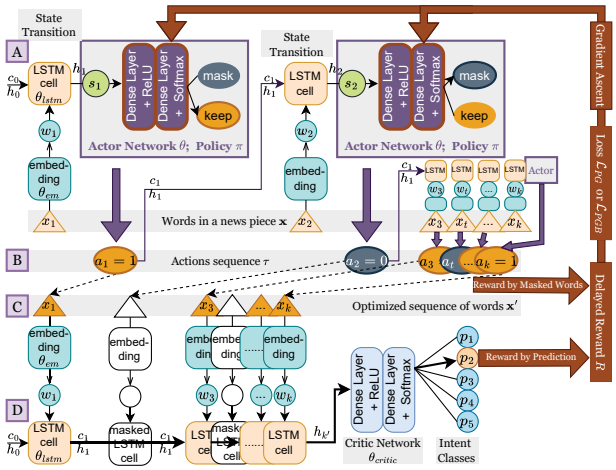

Fig. 1. A DRL episode to optimize structure representations of an annotated tweet $\mathbf{x}$ for the fine-tuned models in Section II-C (PG and PGB).

we manually annotate an existing fake news dataset [8] with three annotators, assigning the dominant intent class label (if two or more) as the golden intent label (denoted as $\mathbf{y}$) to each news piece (a tweet as $\mathbf{x}$) [2]. We ensure consistent sequence length by padding all tweets to a fixed length (k words).

B. LSTM Pre-Training Intent Mining Model

The proposed approach is built upon a pre-trained traditional LSTM serving as the backbone to obtain state information for the DRL agents. This LSTM is pre-trained with the known golden intent labels, minimizing a cross-entropy loss with the L2 regularization.

C. Intent Classification Fine-Tuning by REINFORCE

Fig. 1 illustrates this fine-tuned DRL intent classifier, where the LSTM intent classifier serves as the environment with invariable parameters in a set θ_I. Our news intent classification task benefits from DRL by finding the optimized input sequence $\mathbf{x}'$ in Fig. 1C. When noisy words are removed, a higher prediction of golden intent $\mathbf{y}$ is expected in the pre-trained LSTM, as $p(\mathbf{y}|\mathbf{x}', \theta_I) \geq p(\mathbf{y}|\mathbf{x}, \theta_I)$.

1) DRL Agent: For each LSTM recurrent embedding and encoding step, as $t \in [1, k]$, DRL decides if the current input word x_t is masked for intent prediction. **State**: An LSTM cell accepts x_t with the previous step's cell and hidden state vectors and generates the updated vectors c_t and h_t. The hidden vector h_t is a state s_t. **Actor**: A two-layer Neural Network (θ) takes in the state s_t and generate an output layer of two neurons through a softmax activation. A stochastic policy $\pi(a_t|s_t, \theta)$ is a distribution of two actions. **Action**: One action 'keep' allows x_t to stay in the optimized sequence $\mathbf{x}'$; while the other action 'mask' reduces the length of $\mathbf{x}'$ by a word. **Transition**: Given an action 'keep', the next step receives the LSTM outputs (c_t and h_t) and encodes x_{t+1} as s_{t+1}. However, the action 'mask' passes the LSTM outputs (c_{t-1} and h_{t-1}) from step $t-1$ to step $t+1$ to encode x_{t+1}. **Delayed Reward**: As the optimized sequence $\mathbf{x}'$ (k' words) is validated by the pre-trained LSTM, a delayed reward is generated after step k as a

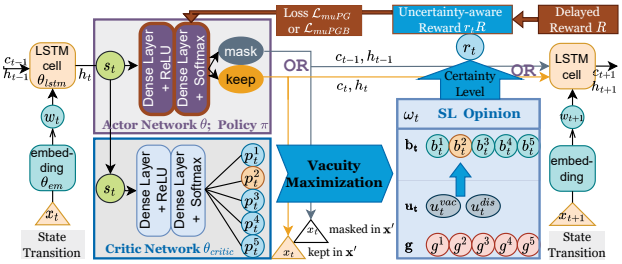

Fig. 2. The uncertainty-aware reward function $r_t R$ with a local Critic, vacuity maximization, and an SL opinion in Section II-D (muPG and muPGB).

prediction of the gold class in Fig. 1D. Additionally, a delayed reward encourages the action 'mask' to remove more noises. Considering λ as a weight of masking word reward, the total delayed reward is formulated as:

$$R = p(\mathbf{y}|\mathbf{x}', \theta_I) + \lambda(k - k')/k. \tag{1}$$

Policy Gradient Loss: This natural language processing (NLP) classification task has the property of missing immediate reward in each DRL step, resulting in a delayed reward and on-policy learning by policy gradient (PG). Under PG methods, the DRL agent's parameters in θ in Fig. 1A are trained by policy ascent, maximizing the future reward R. For each transition step t, the REINFORCE [9] gradients are from a *negative log loss* as:

$$\mathcal{L}_{PG} = -\sum_t R \log \pi(a_t|s_t, \theta). \tag{2}$$

2) Baseline Function: For each tweet, we run five mini-batch episodes to collect gradients. As a common strategy to reduce the variances of gradients $R \log \pi(a_t|s_t, \theta)$, we use a baseline b_t as the mean of R by a mini-batch and the loss is:

$$\mathcal{L}_{PGB} = -\sum_t (R - b_t) \log \pi(a_t|s_t, \theta). \tag{3}$$

D. Multidimensional Uncertainty-Aware Reward Function

1) Local Critic Network: A local Critic at each step t in Fig. 2 is from the pre-trained LSTM classifier, sharing the same input s_t as h_t with the Actor. Following the PG's fundamental on-policy training assumptions, this local Critic indirectly impacts DRL policy updates (θ) by a reward value.

2) Subjective Logic Opinion from the Critic's Intent Distribution: Local intent probabilities $\pi(y|s_t, \theta_I)$ derived from the Critic can be regarded as a multinomial opinion in Subjective Logic (SL) [5]. This conversion leverages a vacuity maximization [5] due to the zero vacuity assumed from local intent probabilities $\pi(y|s_t, \theta_I)$. Through vacuity maximization, an SL opinion has uncertainty masses, such as *vacuity* (u_t^{vac}), generally caused by insufficient evidence for classification, and *dissonance* (u_t^{diss}), due to supporting evidence of each class. *Dissonance* can evaluate the balance of each class, where high dissonance (close to 1.0) means the belief masses follow a uniform distribution towards each defined class.

3) Local Certainty Level: Owing to the local Critic and its transformed SL opinion, we can obtain uncertainty $\mathbf{u_t} = [u_t^{vac}, u_t^{diss}]$ at each DRL local step. The overall uncertainty level from $\mathbf{u_t}$ is determined sequentially by comparing vacuity

TABLE I
MULTI-CLASS INTENT TESTING ACCURACY

Model	Accuracy	Length	Accuracy by Each Class
LSTM	0.817	16.96	[0.676, 0.934, 0.864, 1.0, 0.667]
PG	0.894	9.7	[0.882, 0.983, 0.818, 0.882, 0.667]
PGB	0.911	9.422	[0.926, 0.984, 0.864, 0.824, 0.667]
muPG	0.917	9.394	[0.912, 0.984, 0.818, 0.941, 0.750]
muPGB	0.917	9.056	[0.926, 1.0, 0.818, 0.824, 0.75]

and dissonance to a threshold η because the dissonance is more effective in decision-making under high vacuity [3]. There are three steps to assess a local certainty/uncertainty level r_t by:

- If step t's vacuity is low with $u_t^{vac} < \eta$, the uncertainty level from $\mathbf{u_t}$ is **low**. The Critic indicates a high certainty level toward the Actor by $r_t = \eta + 1 - u_t^{vac} > 1.0$.
- If step t's vacuity is high and dissonance is low, denoted by $u_t^{vac} \geq \eta$ and $u_t^{diss} < \eta$, the level of $\mathbf{u_t}$ is also **low**. Therefore, the Critic indicates a high certainty level of the state s_t by $r_t = \eta + 1 - u_t^{diss} > 1.0$.
- When both are higher than η, as $u_t^{vac} \geq \eta$ and $u_t^{diss} \geq \eta$, the belief masses are uniform of each class by a **high** $\mathbf{u_t}$. So it fails a satisfying level of certainty of s_t. Then, the Critic provides low certainty by $r_t = 1.0$ to keep the original future reward as $r_t R = R$ at each local step.

4) Updated Loss: By replacing a uniform delayed reward R with the uncertainty-aware reward $r_t R$ in Eqs. (2) and (3), we have the updated losses of PG and PGB. The new loss for muPGB, combining the new reward and baseline, is:

$$\mathcal{L}_{muPGB} = -\sum_t (r_t R - b_t) \log \pi(a_t|s_t, \theta). \quad (4)$$

III. PRELIMINARY EXPERIMENTAL RESULTS

A. Experiment Setup

Dataset: We annotate a published Twitter fake news dataset LIAR 2015 [8] for our proposed intent classes. By annotating 835 fake news tweets, described in Section II-A, the distribution of each intent class is $[0.423, 0.275, 0.135, 0.086, 0.081]$, given the orders of intent classes in Section II-A.

Schemes: Comparing to the pre-trained LSTM, we test fine-tuned DRL classifiers by REINFORCE (PG), adding baseline (PGB), adding uncertainty (muPG), and both (muPGB).

Metrics: Average total delayed reward, multi-class intent classification accuracy, and length of optimized words in x'.

B. Intent Classification Accuracy

Table I compares the accuracy and length of fake news between LSTM and four DRL models. The weight is $\lambda = 0.4$ in Eq. (1), and the uncertainty threshold is $\eta = 0.3$. DRL models bring higher accuracies than LSTM, but they only increase the accuracy for intent class 1 'Information Sharing' and class 2 'Political Campaign'. PGB shows a larger accuracy level and masked words over PG. The uncertainty-aware reward function increases the accuracy from PG's 89.4% to muPG's 91.7% and from PGB's 91.1% to muPGB's 91.7%.

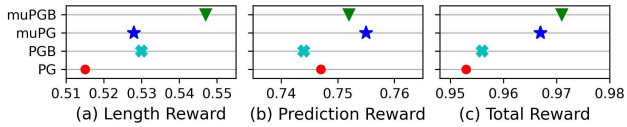

Fig. 3. The average total delayed reward from DRL testing.

C. Total Delayed Reward

Fig. 3 illustrates the total delayed reward from four DRL models, along with its components: prediction and length rewards. Our uncertainty-aware reward function achieves a higher total reward in both muPG and muPGB. The Actor in testing always chooses the optimal action based on policy $\pi(a_t|s_t, \theta)$. By adding a baseline in PGB and an uncertainty-aware function in muPG and muPGB, we observe fewer words in the optimized sequence. muPGB shows the least number of optimized words at 9.056, reducing PG's 9.7 by 6.6%.

IV. CONCLUSION & FUTURE WORK

From this study, we obtained the **key findings**: (1) The uncertainty metrics, vacuity and dissonance from an SL opinion, led to a higher intent accuracy and less noisy words in DRL models. (2) The gradient variance reduction and local certainty levels improved the delayed reward in the PG-based DRL classifiers. Our **future work** will include more sensitivity analysis and DRL variants. In addition, we will test the validity of our proposed reward in more fake news datasets.

REFERENCES

[1] O. D. Apuke and B. Omar, "Fake news and covid-19: Modelling the predictors of fake news sharing among social media users," *Telematics and Informatics*, vol. 56, p. 101475, 2021.

[2] Z. Guo, "Understanding and combating online social deception," Ph.D. dissertation, Virginia Tech, 2023.

[3] Z. Guo, Z. Wan, Q. Zhang, X. Zhao, F. Chen, J.-H. Cho, Q. Zhang, L. M. Kaplan, D. H. Jeong, and A. Jøsang, "A survey on uncertainty reasoning and quantification for decision making: Belief theory meets deep learning," *arXiv:2206.05675*, 2022.

[4] Z. Guo, Q. Zhang, X. An, Q. Zhang, A. Jøsang, L. M. Kaplan, F. Chen, D. H. Jeong, and J.-H. Cho, "Uncertainty-aware reward-based deep reinforcement learning for intent analysis of social media information," in *AAAI Workshop UDM '23*, 2023.

[5] A. Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer, 2016.

[6] H. Purohit and R. Pandey, "Intent mining for the good, bad, and ugly use of social web: Concepts, methods, and challenges," in *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, 2019, pp. 3–18.

[7] Y.-C. Shen, C. T. Lee, L.-Y. Pan, and C.-Y. Lee, "Why people spread rumors on social media: Developing and validating a multi-attribute model of online rumor dissemination," *Online Information Review*, 2021.

[8] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in *Proceedings of the 55th ACL (Volume 2: Short Papers)*, 2017, pp. 422–426.

[9] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, no. 3, pp. 229–256, 1992.

[10] D. Yogatama, P. Blunsom, C. Dyer, E. Grefenstette, and W. Ling, "Learning to compose words into sentences with reinforcement learning," in *5th ICLR*, 2017.