



دانشگاه تهران

دانشکدگان علوم و فناوری های میان رشته ای

گزارش پروژه پایانی درس روش های یادگیری ماشین

در پردازش زبان طبیعی

موضوع: تشخیص موضع گیری

امیر محمد کویش پور

تاریخ تحویل: ۱۴۰۱/۱۱/۱۰

مقدمه

تشخیص موضوع یکی از موضوعات در حال توسعه در عصر رسانه‌های دیجیتال است. با استفاده گسترده از شبکه‌ها در دنیای امروز، اخبار جعلی به صورت فزاینده‌ای رواج یافته است. اخبار جعلی به اطلاعات نادرست یا گمراه‌کننده گفته می‌شود که از طریق رسانه‌های در دسترس منتشر می‌شوند. این اطلاعات نادرست علاوه بر پیامدهای منفی می‌تواند بر درک عمومی از موضوعات مهم، تاثیر بگذارند. در پاسخ به این مسئله، تشخیص موضوع به عنوان ابزاری برای شناسایی اخبار جعلی و پشتیبانی از فرایند بررسی واقعیت، توسعه داده می‌شود. تکنیک‌های تشخیص موضوع از پردازش زبان طبیعی، یادگیری ماشین و تحلیل احساسات برای تعیین موضوع گوینده یا نویسنده در یک موضوع خاص استفاده می‌کنند.

با وجود علاقه روز افزون به تشخیص موضوع، پژوهش‌های بسیاری در زبان فارسی انجام نشده است. هدف از این پژوهش بررسی کاربرد تشخیص موضوع در زبان فارسی و نقش آن در فرایند حقیقت‌سنجی با استفاده از مبدل‌ها و شبکه‌های هم‌گشتی است. نتایج نشان‌دهنده عملکرد بهتر در مقایسه با دیگر روش‌ها در این زمینه است. همچنین فقدان پژوهش در زبان فارسی آن را به حوزه‌ای جذاب برای تحقیق تبدیل می‌کند و نتایج این پژوهش‌ها می‌تواند به صورت کاربردی برای حقیقت‌یابی، در خدمت روزنامه‌نگران و عموم مردم قرار بگیرد.

پژوهش به شرح زیر سامان‌دهی می‌شود: بخش ۲ تشخیص موضوع را تعریف می‌کند، بخش ۳ کارهای مرتبط را در زمینه بررسی می‌کند، بخش ۴ مجموعه داده‌های مورد استفاده در این تحقیق را معرفی می‌کند، بخش ۵ توسعه مدل و پیش پردازش را مورد بحث قرار می‌دهد، بخش ۶ نتایج آزمایش‌ها را ارائه می‌دهد و آنها را مورد بحث قرار می‌دهد و در نهایت بخش ۷ تحقیق را به پایان می‌رساند و کارهای آتی را پیشنهاد می‌کند. اهمیت این تحقیق در کمک به مجموعه رو به رشد ادبیات در زمینه تشخیص موضوع و کاربردهای آن در تشخیص اطلاعات نادرست نهفته است.

مفهوم موضوع‌گیری

تشخیص موضوع، به وظیفه تعیین رابطه بین عنوان داده شده و بدنه مقاله مربوطه اشاره دارد. هدف از تشخیص موضوع، طبقه‌بندی متن موضوع متن مقاله با توجه به عنوان به یکی از دسته‌های زیر است: موافق، مخالف، بحث، یا غیرمرتبط.

دسته "موافق" نشان می‌دهد که مقاله از عنوان پشتیبانی می‌کند. به عنوان مثال، عنوان "واکسن‌ها بی‌خطر هستند" و بدنه مقاله "مطالعات نشان داده است که واکسن‌ها در پیشگیری از بیماری‌ها ایمن و موثر هستند." به عنوان "موافق" طبقه بندی می‌شود.

مقوله "مخالف" نشان می‌دهد که مقاله با عنوان در تضاد است. به عنوان مثال، عنوان "واکسن‌ها بی‌خطر هستند" و بدنه مقاله "مطالعات اخیر نشان داده است که واکسن‌ها می‌توانند عوارض جانبی جدی ایجاد کنند." به عنوان "مخالف" طبقه بندی می‌شود.

مقوله "بحث" نشان می دهد که مقاله بحث یا مناظره ای را در مورد موضوع عنوان ارائه می دهد. به عنوان مثال، عنوان "واکسن ها بی خطر هستند" و بدنه مقاله "ایمنی واکسن ها موضوعی بحث برانگیز است و در سال های اخیر موضوع بحث های زیادی بوده است." به عنوان "بحث" طبقه بندی می شود.

دسته "غیر مرتبط" نشان می دهد که مقاله و عنوان هیچ ارتباطی با یکدیگر ندارند. به عنوان مثال، عنوان "واکسن ها بی خطر هستند" و متن مقاله "جدیدترین فیلم کارگردان استیون اسپیلبرگ قرار است هفته آینده اکران شود." به عنوان "غیر مرتبط" طبقه بندی می شود.

در نتیجه، تشخیص موضع یک کار طبقه بندی است که هدف آن تعیین رابطه بین عنوان و متن مقاله است. از دسته بندی های «موافقم»، «مخالفم»، «بحث و بحث» و «نامرتب» برای دسته بندی موضع متن مقاله با توجه به عنوان استفاده می شود.

مجموعه دادگان

ساختار دادگان در زمینه تشخیص موضع به صورت موضع مقاله-ادعا¹ و موضع سرفصل-ادعا² است. موضع مقاله ادعا شامل مقایسه یک مقاله با یک ادعای خاص است و مشخص می کند که آیا مقاله از ادعا حمایت می کند، رد می کند یا نسبت به آن بی طرف است. از سوی دیگر، موضع عنوان-ادعا شامل تعیین موضع یک عنوان در رابطه با یک ادعا است.

مجموعه دادگان موضع مقاله-ادعا معمولاً از جفت مقاله و ادعا به همراه برچسب موضع اختصاص داده شده به هر جفت تشکیل شده است. این برچسب ها می توانند «حمایت»، «انکار» یا «خنثی» باشند که نشان دهنده موضع مقاله در قبال ادعا هستند. از سوی دیگر، مجموعه دادگان موضع سرفصل-ادعا از جفت عنوان و ادعا به همراه برچسب موضع اختصاص داده شده به هر جفت تشکیل شده است. در این حالت از عنوان برای نشان دادن محتوای مقاله استفاده می شود و برچسب موضع بیانگر موضع تیتیر نسبت به ادعا است. این نوع داده معمولاً برای شناسایی موضع سرفصل ها در مقالات خبری آنلاین استفاده می شود. هم دادگان موضع مقاله-ادعا و هم دادگان موضع سرفصل-ادعا نقش مهمی در زمینه تشخیص موضع دارند و به طور گسترده در توسعه و ارزیابی الگوریتم های تشخیص موضع استفاده می شوند. از مجموعه دادگان گفته شده به پیشرفت توسعه روش های موجود در تشخیص اخبار جعلی و بهبود فرآیندهای حقیقت یابی بهره گرفت.

¹ Article Claim Stance

² Headline Claim Stance

کارهای پیشین

همانطور که در بخش‌های پیشین بیان می‌شود، در تشخیص موضوع‌گیری به تازگی پژوهش‌هایی انجام می‌شود و در زبان فارسی تعداد این پژوهش‌ها بسیار کمتر است. از این رو به بررسی پژوهش‌های پیشین در زبان انگلیسی و سپس در زبان فارسی می‌پردازیم.

در پژوهش [۱]، مدلی ترکیبی مبتنی LSTM و CNN به تشخیص اخبار جعلی بر روی مجموعه‌داده‌گان FNC-1 [۲] به کار گرفته می‌شود، همچنین در رویکرد پیشنهادی خود از متد کاهش ابعاد PCA استفاده می‌کنند.

در پژوهش [۳]، با استفاده جاسازی مدل BERT و بدست آوردن موضع‌گیری یک ادعا نسبت به یک متن خبری به بهبود تسک تشخیص اخبار جعلی می‌پردازند، بدین منظور در مدل نهایی تشخیص اخبار جعلی از مدل‌های CNN، Bi-LSTM و LSTM، ANN استفاده می‌کنند.

همچنین در زبان فارسی دو پژوهش مطرح می‌شود. در پژوهش [۴] مجموعه‌داده‌گان و مدلی مبتنی بر شبکه‌های هم‌گسستی ارائه می‌شود. ساختار مدل به کار رفته در این پژوهش stacked LSTM است. سپس در پژوهش [۵] با تقویت مدل پیشین و استفاده از مدل‌های زبانی بافتاری نتایج بهبود داده می‌شود. تغییرات داده شده به صورت استفاده از مدل زبانی ParsBERT و افزایش داده‌گان به روش تکنیک افزایش آسان داده‌گان^۳ (EDA) به تشخیص موضع‌گیری می‌پردازند. همچنین برای تنظیم ابرپرامترها (نرخ یادگیری، اندازه دسته و تعداد اپیاک) از الگوریتم‌های ASHA [۶] با معیار کمینه‌کردن خطا بهره گرفته می‌شود.

ساخت مدل

در معماری شبکه، ما از ParsBERT استفاده می‌کنیم. جهت استخراج ویژگی از BERT از دو LSTM استفاده می‌کنیم. LSTM ابتدایی برای استخراج ویژگی‌های اظهارنظر^۴ و LSTM بعدی، جهت استخراج ویژگی‌های متن مقاله^۵ در نظر گرفته می‌شود. جهت یادگیری بهتر بافت LSTM‌های خود دو جهته در نظر می‌گیریم. رند یادگیری با Fine-Tuning مدل از پیش آموزش داده شده BERT پایه بر مجموعه‌داده‌گان انجام می‌شود. در پایان خروجی لایه‌های LSTM به هم متصل می‌شوند و پس از گذشت چندین لایه خطی خروجی نهایی حاصل می‌شود.

جهت بهینه‌سازی مدل، از AdamW با نرخ یادگیری $2e-6$ استفاده می‌کنیم. جهت تنظیم مقدار نرخ یادگیری در زمان آموزش مدل از زمان‌بند CosineAnnealingLR استفاده می‌شود. برای آموزش آموزش مدل، از توقف زودهنگام جهت جلوگیری از بیش‌برازش مدل، بهره می‌گیریم. تعداد اپیاک‌های در نظر گرفته شده برای آموزش مدل، 16 اپیاک است.

^۳ EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks

^۴ Claim

^۵ Article Body

همچنین جهت جلوگیری از بیش‌برازش به کمک EDA به افزایش دادگان می‌پردازیم. روند کار انجام با افزایش دادگان آموزش با برچسب دادگانی کمتر، جهت افزایش تنوع دادگان و تعمیم‌پذیری بیشتر مدل مورد استفاده قرار می‌گیرد.

تابع هزینه مورد استفاده Supervised Contrastive است. از این تابع هزینه با هدف یادگیری بازنمایی استفاده می‌شود. به طوری که با نزدیک‌تر کردن دادگان مشابه به هم و دور کردن دادگان بدون شباهت از همدیگر، سعی در بهبود یادگیری بازنمایی می‌نماید.

مدل پایه ارائه شده به کمک زبان پایتون و کتابخانه‌ی Pytorch پیاده‌سازی می‌شود و کد آن به صورت است:

```
class BaseModel(nn.Module):

    def __init__(self):
        super(BaseModel, self).__init__()
        self.bert = AutoModel.from_pretrained(
            c_model, output_hidden_states=True)
        self.lstm_claim = nn.LSTM(self.bert.config.hidden_size,
                                   300, num_layers=2, bidirectional=True,
batch_first=True)
        self.lstm_body = nn.LSTM(self.bert.config.hidden_size,
                                   300, num_layers=2, bidirectional=True,
batch_first=True)

        self.fc1 = nn.Linear(1200, 512)
        self.act = nn.ReLU()
        self.drop = nn.Dropout(0.2)
        self.fc2 = nn.Linear(512, 4)

    def forward(self, input_ids, attention_mask, sep_idx):
        output = self.bert(input_ids=input_ids, attention_mask=attention_mask)
        _, (last_hidden_layer_claim, _) = self.lstm_claim(
            output[2][11][:, :][:, 0:sep_idx])
        _, (last_hidden_layer_body, _) = self.lstm_body(
            output[2][11][:, :][:, sep_idx:])

        pooled_output = torch.cat((last_hidden_layer_claim[2, :, :],
last_hidden_layer_claim[3, :, :],
                                   last_hidden_layer_body[2, :, :],
last_hidden_layer_body[3, :, :]), dim=1)
        out = self.act(self.fc1(pooled_output))
        out = self.drop(out)
        return self.fc2(out)
```

نتایج

پژوهش ارائه شده مبتنی بر پژوهش [۵] است، با این تفاوت که بر خروجی BERT، LSTM به کار می‌بریم و از تابع هزینه Supervised Contrastive جهت یادگیری مدل استفاده می‌شود.

جدول ۱: مقایسه نتایج صورت گرفته در زبان فارسی

مدل	مقاله - ادعا	
	صحت	امتیاز اف - یک
مدل پیشنهادی	۸۱	۷۶
Nasiri and Analoui [5]	۷۶,۳۳	۷۵,۶۴
Zarharan et al. [4]	۷۲	۷۱

نتیجه‌گیری و کارهای پیش رو

نتایج بدست آمده با دسترسی به منابع محاسباتی محدود می‌باشد و احتمال حاصل شدن نتایج بهتری با تغییر ابر پارامترها میسر است. همچنین با رشد استفاده از روش‌های یادگیری چندوظیفه‌ای و یادگیری گروهی در وظایف پردازش زبان طبیعی، بنظر می‌رسد که بتوان نتایج بهتری به کمک این روش‌ها بدست آورد. شایان ذکر است که بدست آوردن بازنمایی‌ها از مدل‌های زبانی چندزبانه، نتایج قابل توجه را نشان می‌دهند.

- [1] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access*, vol. 8, pp. 156695–156706, 2020.
- [2] *The fake news challenge: Exploring how artificial intelligence technologies could be leveraged to combat fake news.*
- [3] H. Karande, R. Walambe, V. Benjamin, K. Kotecha, and T. S. Raghu, "Stance detection with BERT embeddings for credibility analysis of information on social media," *PeerJ Comput. Sci.*, vol. 7, no. e467, p. e467, Apr. 2021.
- [4] M. Zarharan *et al.*, "Persian Stance Classification Data Set," in *Proceedings of the Conference for Truth and Trust Online 2019*, Oct. 2019. doi: 10.36370/tto.2019.30.
- [5] H. Nasiri and M. Analoui, "Persian stance detection with transfer learning and data augmentation," in *2022 27th International Computer Conference, Computer Society of Iran (CSICC)*, Tehran, Iran, Islamic Republic of, Feb. 2022. doi: 10.1109/csicc55295.2022.9780479.
- [6] L. Li *et al.*, "A system for massively parallel hyperparameter tuning," *arXiv [cs.LG]*, Oct. 13, 2018. [Online]. Available: <http://arxiv.org/abs/1810.05934>