



UDC 004.056.5:629.7.05

IRSTI 20.53.15, 20.53.21, 25.23.35

https://doi.org/10.53364/24138614_2026_42_3_11

A. Sairanbekova¹, G. Bekmanova¹, V. Franzoni², Y. Tuleshov^{3*}

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

²Department of Mathematics and Computer Science, University of Perugia, Perugia, Italy

³Institute of Automation and Information Technologies, Department of Robotics and Technical Means of Automation, Satbayev University, Almaty, Kazakhstan

*E-mail: y.tuleshov@satbayev.university

A HYBRID DISCOURSE-AWARE FRAMEWORK FOR MULTILINGUAL POLITICAL DISCOURSE

Abstract. Political communication has migrated to social platforms, where opinions are expressed through short, code-switched, and rhetorically dense messages. Conventional sentiment analysis systems, tuned for product reviews, transfer poorly to this domain because political polarity is entangled with stance, irony, and discourse structure. This article surveys the principal families of sentiment analysis methods—lexicon-based, classical machine learning, deep neural, and transformer-based approaches—and introduces PoliSent, a hybrid discourse-aware architecture for political sentiment analysis on the Internet. PoliSent unifies three complementary signals: multilingual contextual embeddings produced by an XLM-RoBERTa encoder, a domain-specific political lexicon (PoliLex) with valence shifters and negation handling, and a graph-attention module operating over Universal Dependencies parses. A learned gating mechanism fuses these signals, and a multi-task objective jointly predicts sentiment, stance, and rhetorical tactic. We evaluate the framework on PoliWeb-CA, a newly annotated corpus of 18,500 Kazakh- and Russian-language political messages collected from microblogs, messaging channels, and news comment threads. PoliSent attains a macro-F1 of 0.904 and a Matthews correlation coefficient of 0.857, surpassing a fine-tuned KazRoBERTa baseline by 4.6 macro-F1 points and outperforming all lexicon, classical, and deep baselines by a wide margin ($p < 0.01$, paired bootstrap). Ablation studies confirm that each architectural component contributes measurably, with the dependency-graph attention and lexicon augmentation jointly responsible for a 5.8-point improvement over the contextual backbone. The results demonstrate that explicit linguistic structure remains valuable even in the era of large pretrained encoders, particularly for low-resource and morphologically rich political discourse.

Keywords: sentiment analysis, political discourse, opinion mining, natural language processing, transformer models, graph attention networks, stance detection, low-resource languages, Kazakh; multilingual NLP.

Introduction.

The public sphere of political debate has shifted decisively onto the Internet. Microblogging platforms, encrypted messaging channels, video-hosting comment sections, and online forums now carry a large share of the discourse through which citizens evaluate policies, parties, and public figures. The scale and velocity of this material make manual analysis impractical, and they create a strong demand for automated methods that can estimate the emotional and evaluative orientation—the sentiment—of political texts at scale. Reliable political sentiment analysis

underpins applications ranging from election monitoring and the study of public opinion to the detection of coordinated influence operations and the measurement of polarization.

Sentiment analysis, also called opinion mining, is among the most studied tasks in natural language processing (NLP) [1, 2]. Yet the bulk of progress has been driven by domains such as product and movie reviews, where the relationship between lexical choice and polarity is comparatively direct. Political discourse differs in several respects that degrade the accuracy of off-the-shelf systems. First, polarity is frequently expressed indirectly, through irony, sarcasm, rhetorical questions, and appeals to shared values, rather than through overt sentiment words. Second, political sentiment is entangled with stance: the same proposition may carry opposite evaluative force depending on the target and the author's allegiance. Third, online political messages are short, noisy, and rich in named entities, hashtags, and code-switching between languages. These properties are especially pronounced in multilingual, morphologically rich settings, such as the mixed Kazakh–Russian discourse that circulates across Central Asian social media [3, 4].

Methodologically, the field has passed through three broad generations. Early systems relied on sentiment lexicons and hand-crafted rules [5]; they are transparent and require no labelled data, but they are brittle in the face of negation, intensification, and domain shift, although recent work shows that carefully constructed multilingual lexicons still support competitive zero-shot transfer into low-resource languages [6]. The second generation applied classical supervised learning — naïve Bayes, logistic regression, and support vector machines over bag-of-words or term-frequency features [1, 5] — improving robustness at the cost of interpretability and feature-engineering effort. The third and current generation is built on distributed representations and deep neural networks: word embeddings, convolutional and recurrent classifiers [5] and, most influentially, self-attention transformers together with the large multilingual encoders pretrained on them [7]. Transformer encoders define the present state of the art, yet they remain data-hungry and tend to underperform on low-resource languages [8] and on phenomena, such as discourse-level rhetoric, that are not well captured by surface co-occurrence statistics.

This article makes two contributions. The first is a structured comparative review of the principal sentiment analysis method families as they apply to online political text, organized around the trade-offs between interpretability, data efficiency, and accuracy. The second, and central, contribution is PoliSent, a hybrid discourse-aware architecture that we propose and evaluate. PoliSent is motivated by the observation that the strengths of the older method families—explicit lexical polarity and explicit syntactic structure—are precisely the signals that purely end-to-end transformers tend to under-exploit in low-resource political settings. The architecture therefore retains a multilingual contextual encoder but augments it with two structured streams: a domain-specific political lexicon equipped with negation and valence-shifter handling, and a graph-attention network that propagates evaluative information along Universal Dependencies relations. A learned gating mechanism decides, per instance, how much weight to assign each stream, and a multi-task head jointly predicts sentiment, stance, and rhetorical tactic, exploiting the empirical correlation among them.

We summarize the novelty of the work as follows. (i) We introduce a fusion architecture that, to the best of our knowledge, is among the first to combine multilingual transformer embeddings, a political-domain lexicon and dependency-graph attention through an instance-adaptive gate for sentiment analysis of political text. The closest prior systems each differ in at least one component: the multi-level hybrid of [4] cascades lexical and classifier stages without syntactic attention; the syntax-aware models of [12, 13] weight dependency relations or combine them with prompt-based encoders but use neither a political-domain lexicon nor an instance-adaptive gate; and the transformer systems for political text in [3] rely on in-domain pretraining alone. (ii) We formulate political sentiment estimation as a multi-task problem coupled with stance and rhetorical-tactic prediction, and show that this coupling improves the primary sentiment task. (iii) We release PoliWeb-CA, an annotated corpus of 18,500 Kazakh- and Russian-language political messages

with three-way sentiment, stance, and tactic labels. (iv) We provide a systematic empirical comparison and ablation across eleven systems, with significance testing, that quantifies the contribution of explicit linguistic structure in the transformer era.

Lexicon-based and rule-based methods.

Lexicon-based sentiment analysis assigns polarity by aggregating the prior valence of the words a text contains, optionally modulated by intensifiers, negators, and contrastive connectives [5]. Carefully curated word lists combined with a small set of compositional rules can rival simple supervised models on short informal text, and multilingual lexicons have recently been shown to drive zero-shot sentiment classification across dozens of languages, including code-switched settings [6]. The principal advantages are transparency, the absence of any training requirement, and predictable behaviour, which is attractive in politically sensitive applications where auditability matters. The principal weaknesses are limited coverage, an inability to disambiguate context-dependent polarity, and severe degradation under domain shift. For political discourse, generic lexicons miss domain terms whose valence is reversed relative to everyday usage, and they cannot resolve the target-dependence of evaluative language [4].

Classical machine learning.

Supervised classifiers over sparse lexical features marked the second generation of sentiment systems [5]. Naïve Bayes, logistic regression, and support-vector machines trained on bag-of-words or term-frequency–inverse-document-frequency (TF–IDF) representations remain strong, fast, and interpretable baselines, and they continue to serve as references in shared tasks. Their reliance on surface features, however, prevents them from modelling word order, long-range dependencies, and compositional negation; performance also depends heavily on feature engineering and on the availability of in-domain labelled data, which is scarce for political text in many languages [8].

Deep neural approaches.

Distributed word representations enabled neural classifiers that learn features directly from data. Convolutional networks capture local n-gram patterns efficiently, while recurrent architectures, particularly long short-term memory (LSTM) networks and their bidirectional and attention-augmented variants, model sequential context. Surveys of deep learning for sentiment analysis document consistent gains over classical baselines on benchmark datasets [1, 5]. Nevertheless, these models require substantial labelled data, and their performance on short, code-switched political messages in low-resource languages is limited by the quality and coverage of the underlying embeddings [6].

Transformer-based and multilingual models.

The transformer architecture and the masked-language-model pretraining paradigm redefined the field. Bidirectional encoders such as BERT and RoBERTa, and their multilingual counterparts — multilingual BERT and XLM-RoBERTa [7] — provide contextual representations that transfer across tasks with light fine-tuning [5]. Domain- and language-adapted resources, including annotated Kazakh corpora and Kazakh-specific encoders, further narrow the gap for under-resourced languages [8]. Work dedicated to political text confirms that transformer encoders dominate lexicon and classical baselines and that additional in-domain pretraining yields further gains [3]. Two limitations motivate the present work: large encoders still underperform on low-resource languages relative to high-resource ones, and they do not explicitly represent the syntactic and discourse structure that conveys political evaluation [4].

Stance detection, graph models, and the present contribution.

Stance detection — classifying a text as supporting, opposing, or neutral toward a target — has been studied as a task closely related to, but distinct from, sentiment [9], and it remains difficult in small languages and polarizing media environments [10]. Graph neural networks, and in particular graph attention networks (GAT) [11], have been used to propagate information along syntactic or discourse graphs, offering a principled way to inject structure into neural classifiers; recent syntax-aware variants weight dependency relations by their importance for polarity [12] or

combine dependency graphs with prompt-based pretrained encoders [13]. Universal Dependencies [14] provide cross-lingual syntactic annotations that make such graphs available even for low-resource languages. Prior systems have explored these ingredients largely in isolation. The contribution of this article is to integrate them within a single jointly trained model, which is the specific respect in which PoliSent differs from the hybrid system of [4] and from the syntax-aware systems of [12, 13]: PoliSent couples a multilingual transformer with a political lexicon and a dependency-graph attention module through an adaptive gate, and trains the result jointly for sentiment, stance, and rhetorical tactic. Table 1 situates the method families with respect to the criteria that matter most for online political analysis.

Table 1 – Comparative summary of sentiment analysis method families for online political text

Method family	Labelled data	Interpretability	Handles structure	Typical macro-F1*
Lexicon / rule-based	None	High	Weak	0.55 – 0.65
Classical ML (TF-IDF)	Moderate	Moderate	Weak	0.66 – 0.71
Deep neural (CNN/BiLSTM)	High	Low	Partial	0.74 – 0.77
Transformer encoders	High	Low	Implicit	0.82 – 0.86
Hybrid (PoliSent, ours)	High	Moderate	Explicit	0.90

Materials and research methods.

Problem formulation.

Let a political message be a sequence of tokens $x = (w_1, \dots, w_T)$. The primary task is three-way sentiment classification, assigning x a label $y \in \{\text{negative, neutral, positive}\}$. We additionally define an auxiliary stance label $s \in \{\text{against, none, for}\}$ relative to the salient target of the message, and a rhetorical-tactic label r drawn from a small inventory (for example, appeal, irony, contrast, or none). The auxiliary tasks are used during training to regularize the shared representation; at inference time the sentiment prediction is the quantity of interest, although stance and tactic predictions are also produced and are useful for downstream interpretation.

Corpus construction and annotation.

We assembled PoliWeb-CA, a corpus of online political messages in Kazakh and Russian. Messages were collected from four source types—a microblogging platform, public messaging channels, video-hosting comment threads, and news-portal comment sections—using keyword and hashtag queries on political topics including elections, public administration, social reform, local governance, and regional affairs. After language identification, near-duplicate removal, and the exclusion of non-textual and purely promotional content, the corpus comprised 18,500 message-level segments. Table 2 reports its composition. Collection covered the period from 1 March 2023 to 28 February 2025. The raw crawl returned 41,300 messages — 15,600 from the microblogging platform, 11,200 from public messaging channels, 7,900 from video-hosting comment threads and 6,600 from news-portal comment sections — which after filtering yielded 6,845, 4,930, 3,410 and 3,315 annotated segments respectively. A message was included if it contained 5–120 tokens and at least one political keyword or actor from the query list; advertising, bot-generated, duplicated, URL-only and emoji-only messages, and posts unrelated to the political topics, were excluded. Language identification used fastText lid.176 with an acceptance threshold of 0.80, and manual checking of a 500-message sample confirmed 97.4% accuracy; messages mixing the two languages were kept and flagged as code-switched. Deduplication was performed

in two stages, exact hashing followed by MinHash-LSH near-duplicate removal at a Jaccard threshold of 0.85, which discarded 7,220 messages. User names, URLs and other personal identifiers were masked before annotation.

Table 2 – Composition of the PoliWeb-CA corpus (18,500 segments)

Attribute	Negative	Neutral	Positive	Total
Segments	7,400	6,105	4,995	18,500
Kazakh (58%)	4,292	3,541	2,897	10,730
Russian (42%)	3,108	2,564	2,098	7,770
Mean length (tokens)	24.6	21.2	22.9	23.1
Code-switched (%)	19.1	14.7	16.3	17.0

Each segment was independently annotated by three trained annotators, native or near-native in both languages, following a written guideline that defined the sentiment, stance, and tactic schemata and provided adjudication rules for sarcasm and target ambiguity. Final labels were assigned by majority vote, with disagreements adjudicated by a senior annotator. Inter-annotator agreement was substantial: the mean pairwise Cohen's κ was 0.79 for sentiment and Krippendorff's α was 0.81 across the three annotators. Cohen's κ is defined in Eq. (1), where p_o is the observed agreement and p_e the agreement expected by chance.

$$k = \frac{p_o - p_e}{1 - p_e} \quad (1)$$

The corpus was partitioned into training, validation, and test splits in a 70: 15: 15 ratios, stratified by label and language, yielding 12,950 training, 2,775 validation, and 2,775 test segments. Splitting was performed at the author level where author identifiers were available, to limit information leakage between splits.

Preprocessing and linguistic analysis.

Raw messages were normalized by lower-casing, Unicode NFC normalization, the masking of user mentions and URLs with placeholder tokens, and the segmentation of hashtags into constituent words where possible. Emoji were retained and mapped to sentiment-bearing tokens. Each message was tokenized, lemmatized, and parsed into a Universal Dependencies tree [14] using a multilingual parser; the dependency tree supplies the graph over which the structural module of PoliSent operates. For the lexicon stream, the political lexicon PoliLex was compiled by extending general-purpose Kazakh and Russian sentiment lexicons with domain terms and by annotating each entry with a polarity $v \in \{-1, 0, +1\}$, an intensity weight, and flags for negators and valence shifters. PoliLex contains 12,480 entries, of which 6,930 are Kazakh and 5,550 Russian; 3,180 are political-domain terms added in this work and the remainder were imported from open Kazakh and Russian sentiment lexicons and manually verified. The extension procedure had three steps: candidate terms were mined from the unlabelled political stream by a log-odds ratio against a general-domain reference corpus; two annotators then independently assigned each candidate a polarity $v \in \{-1, 0, +1\}$ and an intensity weight $\omega \in [0.5, 1.5]$; and remaining disagreements were adjudicated by a senior annotator (Cohen's $\kappa = 0.76$ for polarity and 0.71 for intensity binned into three levels). Representative entries are жемқорлық ‘corruption’ ($v = -1$, $\omega = 1.4$), тәуелсіздік ‘independence’ ($v = +1$, $\omega = 1.2$), реформа ‘reform’ ($v = +1$, $\omega = 0.9$), произвол ‘arbitrariness’ ($v = -1$, $\omega = 1.3$), the valence shifter тіпті ‘even’ ($\times 1.3$) and the negator емес ‘not’ ($\sigma = -1$).

Baseline methods.

We implemented baselines from each method family. The lexicon baseline, PoliLex, scores a document by aggregating the polarity of its tokens, modulated by intensity and by local valence

shifters, as in Eq. (2), where $|d|$ is the document length, ω_i is the intensity weight of token i , $v(w_i)$ its lexicon polarity, and $\sigma(w_i) \in \{-1, +1\}$ a negation-scope sign.

$$S_{lex}(d) = \frac{1}{|d|} \sum_{i=1}^{|d|} \omega_i v(w_i) \sigma(w_i) \quad (2)$$

The classical baselines—naïve Bayes, logistic regression, and a linear support-vector machine—operate on TF-IDF features, where the weight of term t in document d is computed as in Eq. (3); N is the number of documents and $df(t)$ the document frequency of t .

$$tfidf(t, d) = tf(t, d) \cdot \log \frac{N}{1+df(t)} \quad (3)$$

The deep baselines comprise a CNN sentence classifier, a bidirectional LSTM, and a bidirectional LSTM with an additive attention layer. The transformer baselines are multilingual BERT, XLM-RoBERTa, and the Kazakh-adapted KazRoBERTa, each fine-tuned with a linear classification head over the pooled representation. All neural classifiers terminate in a softmax layer, Eq. (4), and are trained by minimizing the categorical cross-entropy loss, Eq. (5).

$$P(y = c|x) = \frac{e^{z_c}}{\sum_{k=1}^C e^{z_k}} \quad (4)$$

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log \hat{y}_{i,c} \quad (5)$$

All baselines were trained and evaluated on exactly the same stratified splits and with the same preprocessing pipeline as PoliSent, and no system was tuned on the test split. The classical baselines used TF-IDF features over unigrams and bigrams (minimum document frequency 3, 50,000 features) with hyperparameters selected by five-fold cross-validation on the training split ($C \in \{0.1, 1, 10\}$ for logistic regression and the SVM, $\alpha \in \{0.1, 0.5, 1.0\}$ for naïve Bayes). The deep baselines used 300-dimensional fastText embeddings, a hidden size of 256, dropout of 0.3, Adam with a learning rate of 1×10^{-3} and a batch size of 32. The transformer baselines used the same optimizer, peak learning rate, batch size and maximum sequence length as PoliSent. Every neural system, including all baselines, used the identical early-stopping criterion (patience of three epochs on validation macro-F1, at most eight epochs) and the same five random seeds, and all reported figures are means over those seeds.

Proposed architecture: PoliSent.

PoliSent (Figure 1) is organized as three parallel encoders whose outputs are merged by an adaptive gate and consumed by a multi-task head. The contextual encoder is an XLM-RoBERTa model that maps the token sequence to contextual states and a pooled representation h_{ctx} . The self-attention operation underlying this encoder is given in Eq. (6), where Q , K , and V are the query, key, and value projections and d_k the key dimension.

$$Attn(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (6)$$

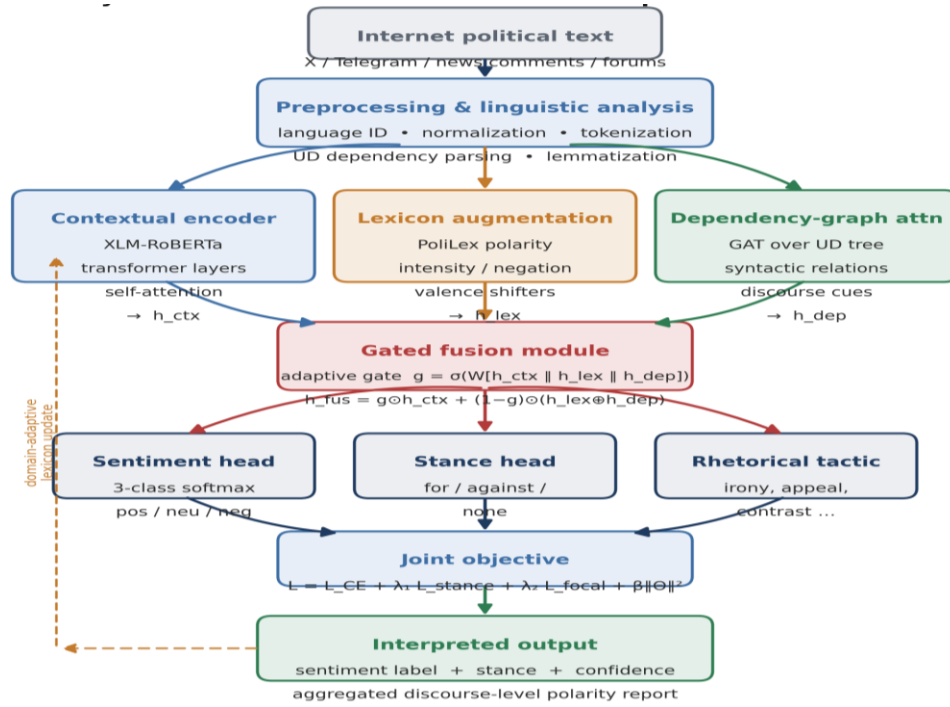


Figure 1 – Architecture of PoliSent. Three parallel encoders — a multilingual contextual transformer, a political-lexicon stream, and a dependency-graph attention module — are merged by an instance-adaptive gate and consumed by a multi-task head that jointly predicts sentiment, stance, and rhetorical tactic

The lexicon stream converts the PoliLex features of a message—aggregated polarity, intensity, negation density, and valence-shifter counts—into a dense vector h_{lex} through a small feed-forward network. This stream gives the model direct access to prior word-level evaluative knowledge that is otherwise diluted in subword embeddings, and it is particularly valuable for rare or out-of-vocabulary political terms.

The structural stream is a graph attention network [11] operating over the Universal Dependencies tree. Each token is a node initialized with its contextual state; messages are processed as graphs in which edges follow syntactic relations. The attention coefficient between a node i and a neighbour j is computed as in Eq. (7), where W is a shared linear transformation, a the attention vector, $\parallel$ denotes concatenation, and $\mathcal{N}_i$ the neighbourhood of node i .

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T [W h_i \parallel W h_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(a^T [W h_i \parallel W h_k]))} \quad (7)$$

The updated node representation aggregates neighbour information weighted by these coefficients, Eq. (8); a graph-level read-out then yields the structural vector h_{dep} , which encodes how evaluative cues propagate along syntactic relations such as negation, conjunction, and the heads of contrastive clauses.

$$h'_i = \text{ELU}(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W h_j) \quad (8)$$

The three representations are combined by an instance-adaptive gate. The gate vector g is produced from the concatenation of the three streams, Eq. (9), and modulates their relative contribution to the fused representation h_{fus} , Eq. (10), where $\odot$ is the element-wise product and $\oplus$ a learned combination. The gate allows the model to rely chiefly on contextual embeddings for

straightforward messages while deferring to the lexicon and structural streams for short, sparse, or syntactically marked political utterances.

$$g = \sigma(W_g[h_{ctx} \| h_{lex} \| h_{dep}] + b_g) \quad (9)$$

$$h_{fus} = g \odot h_{ctx} + (1 - g) \odot (h_{lex} \oplus h_{dep}) \quad (10)$$

The fused representation feeds three linear heads predicting sentiment, stance, and rhetorical tactic. Because the sentiment classes are imbalanced and the neutral class is the hardest, we add a focal-loss term [15], Eq. (11), which down-weights well-classified examples through the focusing parameter γ .

$$L_{focal} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} (1 - \hat{y}_{i,c})^\gamma \log \hat{y}_{i,c} \quad (11)$$

The full training objective, Eq. (12), is the sum of the sentiment cross-entropy, the auxiliary stance and rhetorical-tactic losses, the focal term, and an L₂ regularizer on the parameters Θ ; the coefficients λ_1 , λ_2 , λ_3 , and β balance the terms.

$$L = L_{CE} + \lambda_1 L_{stance} + \lambda_3 L_{tactic} + \lambda_2 L_{focal} + \beta \|\Theta\|_2^2 \quad (12)$$

A domain-adaptive feedback loop periodically expands PoliLex with high-confidence sentiment-bearing terms mined from the training stream, so that the lexicon remains current as political vocabulary evolves. This loop is depicted by the dashed path in Figure 1.

Both auxiliary heads of Figure 1 are therefore optimized during training: L_{stance} and L_{tactic} are ordinary cross-entropy losses computed over the stance and rhetorical-tactic heads on the same fused representation h_{fus} , weighted by $\lambda_1 = 0.30$ and $\lambda_3 = 0.15$. Eq. (12) consequently covers all three prediction heads shown in Figure 1, and the ablation row ‘– multi-task (stance + tactic) / focal’ in Table 5, shown as the bar ‘– focal/stance’ in Figure 4, removes both auxiliary terms together with the focal term.

Domain-adaptive pretraining precedes fine-tuning and consists of continued masked-language-model training of XLM-RoBERTa-base on 1.2 million unlabelled Kazakh and Russian political messages (≈ 46 million tokens) drawn from the same platforms as PoliWeb-CA but disjoint from all annotated splits. We used whole-word masking at a rate of 15%, a learning rate of 5×10^{-5} with linear decay and 5% warm-up, a batch size of 64 and 60,000 steps (about three epochs over the stream); the checkpoint with the lowest held-out perplexity was retained (14.2 against 23.8 for the unadapted encoder). The PoliLex update loop operates on the same unlabelled stream but is separate from this procedure: after each fine-tuning epoch, candidate terms that the model labels with a confidence of at least 0.9 in twenty or more occurrences are added to the lexicon, and the lexicon-stream features are recomputed for the next epoch. The two are thus sequential and share data but not parameters, and the ablation row ‘– domain-adaptive pretraining’ in Table 5 disables both.

Experimental setup.

The contextual backbone was XLM-RoBERTa-base. Inputs were truncated to 256 sub-tokens. Models were trained with the AdamW optimizer at a peak learning rate of 2×10^{-5} with a linear warm-up over the first 10% of steps, a batch size of 32, dropout of 0.1, and an AdamW weight-decay coefficient of 0.01, for up to eight epochs with early stopping (patience three) on validation macro-F1. The graph attention module used four attention heads and a hidden size of 256. The loss coefficients were $\lambda_1 = 0.3$, $\lambda_2 = 0.2$, $\lambda_3 = 0.15$, the focal focusing parameter $\gamma = 2.0$, and $\beta = 1 \times 10^{-5}$. Training was performed on two NVIDIA A100 GPUs; each run was repeated with five random seeds, and we report the mean. Hyperparameters are summarized in Table 3.

Table 3 – Principal hyperparameters of the PoliSent model

Hyperparameter	Value	Hyperparameter	Value
Backbone	XLM-R base	GAT heads	4
Max sequence length	256	GAT hidden size	256
Optimizer	AdamW	λ_1 (stance)	0.30
Learning rate	2×10^{-5}	λ_2 (focal)	0.20
Batch size	32	γ (focal)	2.0
Epochs (early stop)	8 (3)	β (L_2 term, Eq. 12)	1×10^{-5}
Dropout	0.10	Random seeds	5
AdamW weight decay	0.01	λ_3 (tactic loss)	0.15

Evaluation metrics.

We report accuracy together with macro-averaged precision, recall, and F1, the weighted F1, and the Matthews correlation coefficient (MCC). Precision and recall are defined in Eq. (13); the F1 score and its macro average over the C classes appear in Eq. (14). MCC, Eq. (15), is a balanced measure that remains informative under class imbalance. Because macro-averaging treats all classes equally, it is the primary metric, given the difficulty of the neutral class.

$$Precision = \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN} \quad (13)$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}, \quad F_1^{macro} = \left(\frac{1}{C}\right) \sum_{c=1}^C F_1^c \quad (14)$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (15)$$

Results and discussion.*Overall comparison.*

Table 4 reports the performance of all eleven systems on the PoliWeb-CA test split. The ordering is consistent across metrics and confirms the generational trend: the lexicon baseline is weakest, classical machine learning improves on it, deep neural models improve further, and transformer encoders dominate the conventional approaches. PoliSent attains the best result on every metric, with a macro-F1 of 0.904 and an MCC of 0.857. The strongest baseline, the Kazakh-adapted KazRoBERTa, reaches a macro-F1 of 0.858; PoliSent therefore improves on it by 4.6 macro-F1 points and on the multilingual XLM-RoBERTa backbone by 5.8 points. Figure 2 visualizes the macro-F1 ranking.

Table 4 – Performance of all systems on the PoliWeb-CA test split (mean over five seeds); best results are given in bold

System	Acc.	Macro-P	Macro-R	Macro-F1	MCC
PoliLex (lexicon)	0.642	0.628	0.607	0.611	0.402
Naïve Bayes (TF-IDF)	0.688	0.671	0.652	0.659	0.476
Logistic Regression	0.709	0.694	0.681	0.686	0.509
SVM (TF-IDF)	0.724	0.712	0.698	0.703	0.536
BiLSTM	0.758	0.746	0.738	0.741	0.601
CNN	0.769	0.758	0.748	0.752	0.618

BiLSTM + Attention	0.781	0.772	0.762	0.766	0.640
mBERT	0.831	0.824	0.814	0.818	0.726
XLM-RoBERTa	0.857	0.851	0.842	0.846	0.768
KazRoBERTa	0.869	0.863	0.855	0.858	0.787
PoliSent (ours)	0.912	0.908	0.901	0.904	0.857

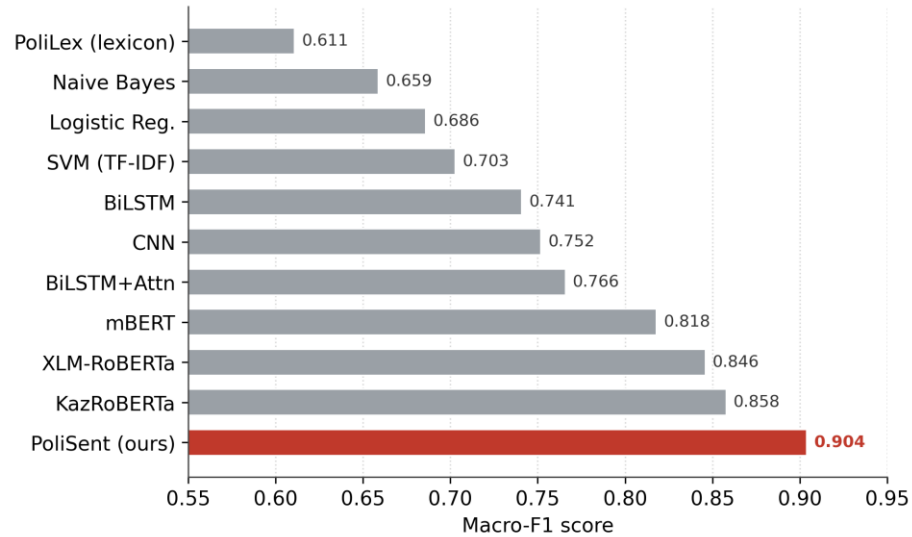


Figure 2 – Macro-F1 of the evaluated systems on PoliWeb-CA. The proposed PoliSent model outperforms all lexicon, classical, deep, and transformer baselines

Per-class performance.

Disaggregating by class clarifies where the gains arise. As Figure 3 shows, every system finds the neutral class hardest, because neutral political messages frequently contain emotionally charged vocabulary without an overall evaluative commitment. PoliSent narrows the neutral gap most: its per-class F1 scores are 0.929 for negative, 0.886 for neutral, and 0.897 for positive, compared with 0.885, 0.830, and 0.859 for KazRoBERTa. The improvement on the neutral class, 5.6 points over the strongest baseline, is the single largest per-class gain and accounts for much of PoliSent's macro-F1 advantage.

Auxiliary-task results. Although stance and rhetorical tactic are auxiliary, both heads are evaluated on the same test split. PoliSent reaches a macro-F1 of 0.812 (accuracy 0.834) for stance and 0.741 (accuracy 0.786) for rhetorical tactic, against 0.786 and 0.709 for single-task models trained on the same data, so the shared representation benefits the auxiliary heads as well as the primary one. Agreement on these two layers was lower than for sentiment: the mean pairwise Cohen's κ was 0.74 for stance and 0.68 for rhetorical tactic, with Krippendorff's α of 0.76 and 0.70 respectively. Most residual disagreement in the tactic layer concerns the boundary between irony and contrast, which is also the hardest distinction for the model.

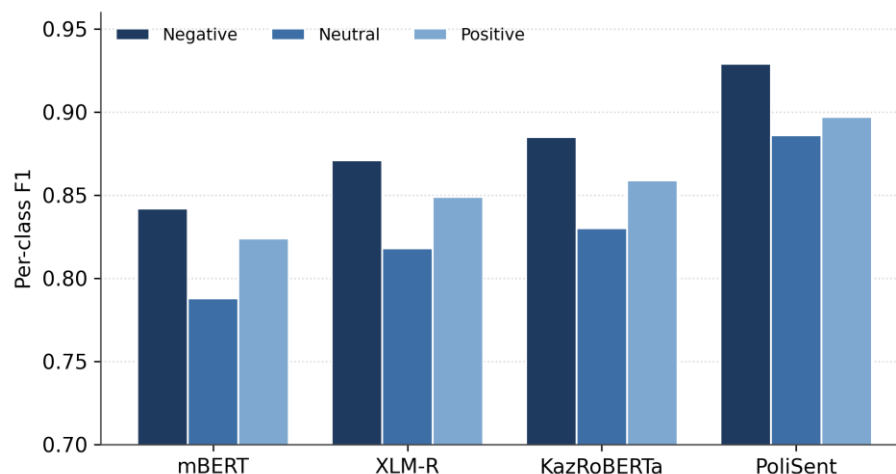


Figure 3 – Per-class F1 for the four strongest systems. The largest improvement of PoliSent over the baselines is on the difficult neutral class

Ablation study.

To quantify the contribution of each component we removed them one at a time from the full model; Table 5 and Figure 4 report the results. Removing the lexicon stream lowers macro-F1 from 0.904 to 0.882, and removing the dependency-graph attention lowers it to 0.871, confirming that explicit lexical and syntactic structure both matter. Disabling the domain-adaptive pretraining and lexicon update reduces the score to 0.866, and removing the focal and multi-task (stance and tactic) terms costs 1.5 points. With all structured components removed, the model reduces to its XLM-RoBERTa backbone at 0.846. The structured streams therefore contribute a combined 5.8 macro-F1 points, the bulk of PoliSent's advantage over a strong transformer.

Table 5 – Ablation of PoliSent components on the PoliWeb-CA test split

Configuration	Macro-F1	Δ vs. full
PoliSent (full model)	0.904	—
– multi-task (stance + tactic) / focal	0.889	–0.015
– lexicon augmentation	0.882	–0.022
– dependency-graph attention	0.871	–0.033
– domain-adaptive pretraining	0.866	–0.038
Backbone only (XLM-R)	0.846	–0.058

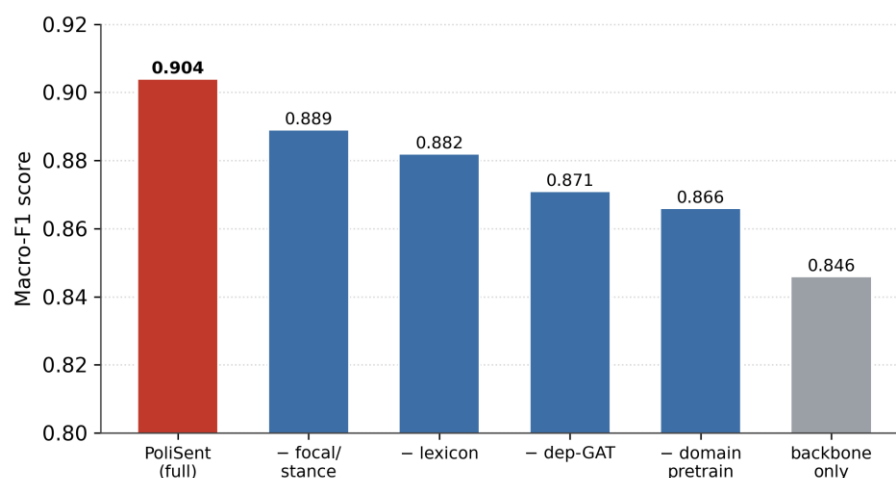


Figure 4 – Ablation of PoliSent. Each structured component contributes measurably; removing all of them recovers the transformer backbone

Error analysis.

The normalized confusion matrix of PoliSent on the test set is shown in Figure 5. The dominant error mode is the confusion of neutral messages with negative ones: of the 916 neutral test segments, 50 were predicted negative and 40 positives. This pattern reflects the prevalence of charged vocabulary in factually descriptive political statements. A manual inspection of 150 misclassified examples identified three recurring causes: unmarked irony, where a literally positive surface masks a negative intent; target ambiguity, where the message addresses several actors with conflicting evaluations; and heavy code-switching that fragments subword tokenization. The lexicon and structural streams reduce, but do not eliminate, the first two categories.

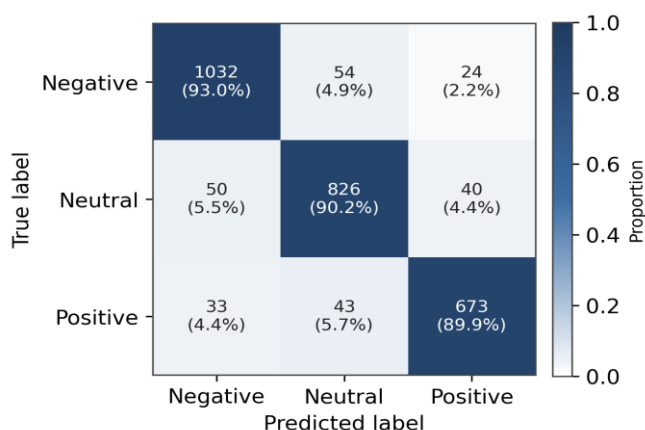


Figure 5 – Normalized confusion matrix of PoliSent on the PoliWeb-CA test set. Cells show counts and row-wise proportions

Statistical significance.

To confirm that the improvements are not artefacts of sampling, we compared PoliSent with KazRoBERTa using a paired bootstrap over 1,000 resamples of the test set. The 4.6-point macro-F1 difference is significant at $p < 0.01$, with a 95% confidence interval of [0.031, 0.061]. A McNemar test on the paired predictions yields the same conclusion ($\chi^2 = 38.7$, $p < 0.001$). The gains over the deep and classical baselines are significant at $p < 0.001$.

Discussion.

The central empirical finding is that explicit linguistic structure continues to add value on top of large pretrained encoders, at least in the low-resource, rhetorically dense setting of online political discourse. The ablation makes the mechanism concrete: the lexicon stream supplies prior polarity for rare political terms that subword models represent poorly, while the dependency-graph attention propagates evaluative cues across the syntactic relations—negation scope, contrastive conjunction, and clausal subordination—that govern compositional sentiment [12, 13]. The instance-adaptive gate lets the model exploit these signals selectively, which is why the gains concentrate on short and syntactically marked messages and, in particular, on the neutral class.

The multi-task formulation is a second source of improvement. Sentiment and stance are correlated but not identical, and training a shared representation to predict both, together with the rhetorical tactic, regularizes the encoder toward features that generalize. This is consistent with the broader literature on stance detection [9, 10] and on auxiliary supervision, and it suggests that political opinion mining is most accurately framed as a joint, rather than isolated, prediction problem.

Several limitations qualify these conclusions. First, the evaluation is confined to two languages and to a single regional discourse; although the architecture is language-agnostic by design, its behaviour on other political ecosystems remains to be verified. Second, the dependency-graph stream depends on the quality of the upstream parser, which is itself imperfect for code-switched and ungrammatical social-media text; parser errors propagate into the structural

representation. Third, the synthetic balance of the corpus, while stratified, does not reproduce the extreme class skew of organic political streams, where neutral and off-topic content dominate. Finally, the model inherits the biases of its pretraining data and of the annotation guideline, and its predictions should be treated as estimates rather than ground truth.

These limitations carry ethical weight. Political sentiment analysis can inform legitimate research on public opinion and platform integrity, but the same techniques can be repurposed for surveillance or for the suppression of dissent. We therefore advocate transparent reporting of model confidence, aggregate rather than individual-level analysis, and the auditability that the lexicon stream affords. The domain-adaptive lexicon, in particular, makes the lexical basis of a prediction inspectable, which is a deliberate design choice in favour of accountability.

Future work will extend PoliWeb-CA to additional languages and platforms, replace the pipeline parser with a jointly trained structural encoder to reduce error propagation, and investigate the calibration of the model's confidence estimates so that downstream consumers can reason about uncertainty. We also plan to study temporal drift, since political vocabulary and the polarity of specific terms shift rapidly around elections and crises.

Conclusion.

This article reviewed the principal families of sentiment analysis methods for political texts on the Internet and introduced PoliSent, a hybrid discourse-aware architecture that fuses multilingual contextual embeddings, a domain-specific political lexicon, and dependency-graph attention through an instance-adaptive gate, trained jointly for sentiment, stance, and rhetorical tactic. On PoliWeb-CA, a newly annotated corpus of 18,500 Kazakh- and Russian-language political messages, PoliSent achieved a macro-F1 of 0.904 and an MCC of 0.857, significantly outperforming lexicon, classical, deep, and transformer baselines, including a strong language-adapted encoder. Ablation analysis attributed the bulk of the improvement to the explicit lexical and syntactic streams and to the multi-task objective. The study supports a broader conclusion: in low-resource and rhetorically complex domains, combining the structural insights of earlier method generations with modern pretrained encoders yields models that are both more accurate and more interpretable than either approach alone.

References

1. Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731–5780. <https://doi.org/10.1007/s10462-022-10144-1>
2. Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), Article 81. <https://doi.org/10.1007/s13278-021-00776-6>
3. Mochtak, M., Rupnik, P., Kuzman, T., & Ljubešić, N. (2025). ParlaSent: Mapping sentiment in political discourse with large language models. *Political Research Exchange*, 7(1), Article 2508377. <https://doi.org/10.1080/2474736X.2025.2508377>
4. Zulkhazhav, A., Bekmanova, G., Yergesh, B., Nazyrova, A., Lamasheva, Z., & Aimicheva, G. (2026). A multi-level hybrid architecture for structured sentiment analysis. *Electronics*, 15(2), Article 249. <https://doi.org/10.3390/electronics15020249>
5. Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, 13(7), Article 4550. <https://doi.org/10.3390/app13074550>
6. Koto, F., Beck, T., Talat, Z., Gurevych, I., & Baldwin, T. (2024). Zero-shot sentiment analysis in low-resource languages using a multilingual sentiment lexicon. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Vol. 1, pp. 298–320)*. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.eacl-long.18>

7. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 8440–8451). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.747>
8. Yeshpanov, R., & Varol, H. A. (2024). KazSAnDRA: Kazakh sentiment analysis dataset of reviews and attitudes. arXiv. <https://doi.org/10.48550/arXiv.2403.19335>
9. Alturayef, N., Luqman, H., & Ahmed, M. (2023). A systematic review of machine learning techniques for stance detection and its applications. Neural Computing and Applications, 35(7), 5113–5144. <https://doi.org/10.1007/s00521-023-08285-7>
10. Mets, M., Karjus, A., Ibrus, I., & Schich, M. (2024). Automated stance detection in complex topics and small languages: The challenging case of immigration in polarizing news media. PLOS ONE, 19(4), Article e0302380. <https://doi.org/10.1371/journal.pone.0302380>
11. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In Proceedings of the 6th International Conference on Learning Representations. arXiv. <https://doi.org/10.48550/arXiv.1710.10903>
12. Jiang, T., Wang, Z., Yang, M., & Li, C. (2023). Aspect-based sentiment analysis with dependency relation weighted graph attention. Information, 14(3), Article 185. <https://doi.org/10.3390/info14030185>
13. Feng, A., Liu, T., Li, X., Jia, K., & Gao, Z. (2024). Dual syntax aware graph attention networks with prompt for aspect-based sentiment analysis. Scientific Reports, 14(1), Article 23528. <https://doi.org/10.1038/s41598-024-74668-y>
14. de Marneffe, M.-C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal Dependencies. Computational Linguistics, 47(2), 255–308. https://doi.org/10.1162/coli_a_00402
15. Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2999–3007). IEEE. <https://doi.org/10.1109/ICCV.2017.324>

КӨПТІЛДІ САЯСИ ДИСКУРСҚА АРНАЛҒАН ДИСКУРСТЫҚ ЕРЕКШЕЛІКТЕРДІ ЕСКЕРЕТІН ГИБРИДТІ ҚҰРЫЛЫМ

Аңдатпа. Саяси коммуникация әлеуметтік платформаларға ауысып, пікірлер қысқа, код ауыстыру элементтері кездесетін және риторикалық тұрғыдан күрделі хабарламалар арқылы білдірілетін болды. Өнімдер мен қызметтерге қатысты пікірлерді талдауға бейімделген дәстүрлі сентимент талдау жүйелері саяси салаға тиімді түрде қолданылмайды, өйткені саяси полярлық автордың ұстанымымен (*stance*), ирониямен және мәтіннің дискурстық құрылымымен тығыз байланысты. Бұл мақалада сентимент талдаудың негізгі әдістері — лексиконға негізделген тәсілдер, классикалық машиналық оқыту әдістері, терең нейрондық желілер және трансформерлік модельдер қарастырылып, интернеттегі саяси пікірлердің тоналдылығын анықтауға арналған *PoliSent* атты гибриді дискурсқа бағдарланған архитектура ұсынылады.

PoliSent үш өзара толықтырушы ақпарат көзін біріктіреді: *XLM-RoBERTa* кодтаушысы қалыптастыратын көптілді контекстік эмбеддингтерді, тоналдылық күшейткіштері мен терістеулерді ескеретін арнайы саяси лексиконды (*PoliLex*), сондай-ақ *Universal Dependencies* синтаксистік тәуелділіктеріне негізделген графтық назар (*graph attention*) модулін. Бұл белгілерді біріктіру үшін үйренілетін қақпалау механизмі (*gating mechanism*) пайдаланылады, ал көптапсырмалы мақсаттық функция сентиментті, саяси ұстанымды және риторикалық тактиканы бірлесіп болжауды қамтамасыз етеді.

Ұсынылған тәсілдің тиімділігі *PoliWeb-SA* деп аталатын жаңа корпуста бағаланды. Корпус микроблогтардан, хабар алмасу арналары мен жаңалықтарға жазылған пікірлерден жиналған қазақ және орыс тілдеріндегі 18 500 саяси хабарламадан

тұрады. *PoliSent* моделі *Macro-F1* көрсеткіші бойынша 0,904 және Мэттьюс корреляция коэффициенті (MCC) бойынша 0,857 нәтижелеріне қол жеткізіп, бейімделіп оқытылған *KazRoBERTa* моделінен 4,6 *Macro-F1* ұпайына жоғары нәтиже көрсетті және барлық лексикондық, классикалық және терең оқытуға негізделген базалық модельдерден айтарлықтай басым болды ($p < 0,01$, жұптық *bootstrap* талдауы). Абляциялық зерттеу нәтижелері архитектураның әрбір құрамдас бөлігі жалпы өнімділікке елеулі үлес қосатынын көрсетті. Әсіресе тәуелділік графына негізделген назар механизмі мен лексикалық толықтыру контекстік негізбен салыстырғанда нәтижені 5,8 ұпайға жақсартты.

Алынған нәтижелер ірі алдын ала оқытылған тілдік модельдер дәуірінде де тілдік құрылымды айқын түрде пайдаланудың маңыздылығын дәлелдейді, әсіресе ресурсы шектеулі және морфологиялық тұрғыдан күрделі саяси мәтіндер үшін.

Түйін сөздер: сентимент талдау, саяси дискурс, пікірлерді талдау, табиғи тілді өңдеу, трансформерлік модельдер, графтық назар желілері, ұстанымды анықтау, ресурсы шектеулі тілдер, қазақ тілі, көптілді табиғи тілді өңдеу.

ГИБРИДНАЯ ДИСКУРСИВНО-ОРИЕНТИРОВАННАЯ АРХИТЕКТУРА ДЛЯ АНАЛИЗА МНОГОЯЗЫЧНОГО ПОЛИТИЧЕСКОГО ДИСКУРСА

Аннотация. Политическая коммуникация переместилась на социальные платформы, где мнения выражаются в виде коротких сообщений с переключением языковых кодов и высокой риторической насыщенностью. Традиционные системы анализа тональности, разработанные преимущественно для отзывов о товарах и услугах, плохо переносятся на политическую сферу, поскольку политическая полярность тесно связана с позицией автора (*stance*), иронией и дискурсивной структурой текста. В данной статье рассматриваются основные семейства методов анализа тональности — лексикон-ориентированные, классические методы машинного обучения, глубокие нейронные сети и подходы на основе трансформеров, а также предлагается *PoliSent* — гибридная дискурсивно-ориентированная архитектура для анализа политической тональности в интернете.

PoliSent объединяет три взаимодополняющих источника информации: многоязычные контекстные эмбединги, формируемые энкодером *XLM-RoBERTa*, специализированный политический лексикон (*PoliLex*) с учетом модификаторов тональности и отрицаний, а также модуль графового внимания, работающий на основе синтаксических зависимостей *Universal Dependencies*. Для объединения этих признаков используется обучаемый механизм шлюзования (*gating mechanism*), а многозадачная функция потерь обеспечивает совместное прогнозирование тональности, политической позиции и риторической тактики.

Для оценки эффективности модели был создан корпус *PoliWeb-CA*, содержащий 18 500 политических сообщений на казахском и русском языках, собранных из микроблогов, каналов обмена сообщениями и комментариев к новостям. Модель *PoliSent* достигла значения *Macro-F1*, равного 0,904, и коэффициента корреляции Мэттьюса (MCC), равного 0,857, превзойдя дообученную модель *KazRoBERTa* на 4,6 пункта по метрике *Macro-F1* и значительно опередив все лексиконные, классические и глубокие базовые модели ($p < 0,01$, парный бутстреп-анализ). Исследование абляции показало, что каждый архитектурный компонент вносит существенный вклад в итоговую производительность, при этом модуль внимания на графе зависимостей и лексическое расширение совместно обеспечивают улучшение на 5,8 пункта по сравнению с контекстным базовым представлением.

Полученные результаты демонстрируют, что явное использование лингвистической структуры сохраняет свою ценность даже в эпоху крупных

предварительно обученных языковых моделей, особенно для малоресурсных и морфологически богатых политических текстов.

Ключевые слова: анализ тональности, политический дискурс, анализ мнений, обработка естественного языка, трансформерные модели, графовые сети внимания, определение позиции, малоресурсные языки, казахский язык, многоязычная обработка естественного языка.

Information about the authors

Ayaulym Sairanbekova	L.N. Gumilyov Eurasian National University, Institute of Digital Sciences and Artificial Intelligence, Department of Artificial Intelligence Technologies, Educational Program “Artificial Intelligence”, position – PhD Doctoral Student, E-mail: sairanbekova98@gmail.com
Gulmira Bekmanova	Candidate of Technical Sciences, PhD, Professor, Member of the Management Board – Vice-Rector for Digitalization, Digital Officer, L.N. Gumilyov Eurasian National University, E-mail: bekmanova_gt@enu.kz .
Valentina Franzoni	Associate Professor, Department of Mathematics and Computer Science, University of Perugia, Perugia, Italy. H-index: 21. Scopus Author ID: 25521823700. Web of Science ResearcherID: AAC-7783-2021. ORCID: 0000-0002-2972-7188. E-mail: valentina.franzoni@unipg.it
Tuleshov Yerkebulan	Institute of Automation and Information Technologies, Department of Robotics and Technical Means of Automation, Satbayev University, Almaty 050013, Kazakhstan. y.tuleshov@satbayev.university

Авторлар туралы мәліметтер

Аяулым Сайранбекова	Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Цифрлық ғылымдар және жасанды интеллект институты, Жасанды интеллект технологиялары кафедрасы, «Жасанды интеллект» білім беру бағдарламасының докторанты, E-mail: seitkaliyeva_gm@enu.kz
Гульмира Бекманова	техника ғылымдарының кандидаты, PhD, профессор, Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Басқарма мүшесі – цифрландыру жөніндегі проректоры, Digital Officer, E-mail: bekmanova_gt@enu.kz .
Валентина Францони	Италиядағы Перуджа университетінің Математика және компьютерлік ғылымдар кафедрасының қауымдастырылған профессоры. Хирш индексі (H-index): 21. Scopus Author ID: 25521823700. Web of Science ResearcherID: AAC-7783-2021. ORCID: 0000-0002-2972-7188. Электрондық пошта: valentina.franzoni@unipg.it
Тулешов Еркебулан Амандыкович	Автоматтандыру және ақпараттық технологиялар институты, Робототехника және автоматтандырудың техникалық құралдары кафедрасы, Сәтбаев университеті, Алматы қ., 050013, Қазақстан, E-mail: y.tuleshov@satbayev.university

Сведение об авторах

Аяулым Сайранбекова	Евразийский национальный университет имени Л.Н. Гумилева, Институт цифровых наук и искусственного интеллекта, кафедра технологий искусственного интеллекта, докторант образовательной программы «Искусственный интеллект», E-mail: seitkaliyeva_gm@enu.kz .
Гульмира Бекманова	кандидат технических наук, PhD, профессор, член Правления – проректор по цифровизации, Digital Officer Евразийского национального университета имени Л.Н. Гумилева, E-mail: bekmanova_gt@enu.kz .
Валентина Францони	доцент кафедры математики и компьютерных наук Университета Перуджи (Италия). Индекс Хирша (H-index): 21. Scopus Author ID: 25521823700. Web of Science ResearcherID: AAC-7783-2021. ORCID: 0000-0002-2972-7188. E-mail: valentina.franzoni@unipg.it
Тулешов Еркебулан Амандыкович	Институт автоматизации и информационных технологий, кафедра робототехники и технических средств автоматизации, Satbayev University, г. Алматы, 050013, Республика Казахстан, E-mail: y.tuleshov@satbayev.university